To see, or not to see?
Psychological Perspectives on the Use of Transparency Disclosures on Microtargeted Political Advertisements

Von der Fakultät für Informatik
der Universität Duisburg-Essen

zur Erlangung des akademischen Grades

Doktor der Philosophie (Dr. phil.)

genehmigte kumulative Dissertation

von

Martin-Pieter Jansen
aus
Nijmegen, die Niederlande

1. Gutachterin: Prof. Dr. Nicole C. Krämer
2. Gutachter: Prof. Dr. Oliver B. Büttner

Tag der mündlichen Prüfung: 31.01.2024

Acknowledgments

Without help from others, I would not have been able to complete this dissertation the way I did. First, I would like to express my gratitude to Prof. Dr. Nicole Krämer for providing me with the opportunity to join her chair and fulfill my goal of starting and completing a doctorate. From the beginning, you provided me with the opportunity to investigate microtargeting from the foundation of persuasive communication, which was my main interest. Your supervision was suitable and there when I needed it. However, if I wanted to explore new topics or even different universities, you let me, which, I believe, made me a better researcher overall. Thank you for your guidance and opportunity to develop. I would also like to thank Prof. Dr. Oliver Büttner. Although we did not interact that often throughout the process, I appreciate your enthusiasm for evaluating my work.

Furthermore, I thank the entire chair of Social Psychology for welcoming me in strange times via Zoom, and welcoming me again when the world started to open up! Special thanks to everyone who I had the pleasure of sharing an office with: Bianca, Hannah, and Marcel. Thank you for the interactions, discussions, and laughter we shared.

I am grateful to the people I got to work with directly, my amazing co-authors, Dr. Yannic Meier and Dr. Iris van Ooijen. I firmly believe that we have conducted awesome research. Further, I would like to thank Dr. Barbara Müller for the constant interactions that in hindsight can definitely be labeled as mentoring, your impact on all of this is larger than you might think.

All my friends who have been there from the start, thank you. A special thanks to Duncan, who supported me from the start, and will certainly be there in the future as my biggest criticaster and best friend.

Finally, my biggest supporters in the galaxy, my mother Renée, *dank je wel voor alles*, I dedicate this work to you. My father Rob, if only you could have read this. My partner Liz: About four years ago, you asked me when it was going to feel like it was enough, and when I would be satisfied; I think you already knew the answer back then. Thank you for every-freaking-thing: You make me a better person. Finally, the being in the world that never judged me regarding my work but did so regarding everything else as is her feline nature, thanks, Kobe!

Annotation of the papers contained in the cumulus

Article 1

Jansen, M.-P., & Krämer, N.C. (2023). Empty Transparency? The Effects on Credibility and Trustworthiness of Targeting Disclosure Labels for Micro-Targeted Political Advertisements. *Weizenbaum Journal of the Digital Society*, *3*(1), https://doi.org/10.34669/WI.WJDS/3.1.5

Article 2

Jansen, M.-P., Meier, Y., & Krämer, N.C. (In revision). Time for transparent targeting: an investigation of targeting disclosures, coping mechanisms, credibility, and political attitude. *Behaviour and Information Technology.*

Article 3

Jansen, M.-P., & Krämer, N.C. (2023). Balancing perceptions of targeting: An investigation of political microtargeting transparency through a calculus approach. *PLOS ONE, 18*(12), e0295329. https://doi.org/10.1371/journal.pone.0295329

Article 4

Jansen, M.-P., & Van Ooijen, I. (In revision). For your eyes only? An eye-tracking experiment investigating microtargeting transparency, visual attention, and critical processing. *Journal of Media Psychology.*

Abstract

While in the past, most political communication occurred via television commercials and approaching people in the streets, political parties recently followed their target groups to where they often reside: online, on social media specifically. On these platforms, users and their behaviors leave traces. This information, combined with information that users freely share, such as demographics or location data, is stored by platforms. Platforms enable political advertisers to use these data and segment very specific target groups. Subsequently, these target groups can be targeted in detail and shown specific advertisements. These political advertisements are developed to resonate optimally within target groups. This specific method of narrowly personalizing political advertisements based on user data is called *political microtargeting*. While risks regarding the utilization of this technique exist, solutions to counter these potential negative effects also exist. Disclosure labels are able to inform receivers of messages that they are targeted directly on the advertisement itself. However, the implementation of disclosures on advertisements has certain consequences that have not yet been thoroughly investigated in the context of microtargeting. Thus, this dissertation investigated the impact of so-called targeting disclosures on political advertisements by scrutinizing receivers' recognition processes and reactions to the message if it contained a disclosure. Moreover, the consequences for users' perceptions of the source and message were examined. Finally, the privacy aspect of microtargeting was scrutinized by determining which indicators lead to users' intentions to protect their data online.

First, Study I compared a post containing a targeting disclosure to a regular advertisement and an organic post on Facebook. The findings of this study showed no relationship between different disclosures, users' level of targeting recognition, and their perceptions. However, most participants did not recall the disclosure, which may have influenced these findings.

Furthermore, Study II compared an advertisement containing an interactive targeting disclosure to one containing a sponsorship disclosure on Facebook and investigated whether this led to better recognition. Moreover, the study examined resistance as a way of coping with these messages, and scrutinized whether this led to differences in users' perceptions and attitudes. The findings showed no meaningful differences in recognition between the two disclosures. However, it was found that if users are aware of the targeting occurring, they cognitively resist the message, which relates to less favorable perceptions of the source and attitudes towards the politician.

Subsequently, Study III compared a targeting disclosure based on Instagram's misinformation disclosures used during the COVID-19 pandemic to the platform's currently used disclosure. Further, this study investigated users' privacy concerns and benefit perceptions and whether these led to intentions to protect one's data online. The findings show that the targeting disclosure increased users' recognition of the advertisement being targeted at them, which led to more benefit perceptions but not privacy concerns. However, users' privacy concerns were related to their intention to protect their online data.

Finally, Study IV compared two disclosures based on the misinformation disclosures that Instagram used in the COVID-19 pandemic and investigated whether the targeting disclosure led to more visual attention to the disclosure, and, in turn, to more critical processing, but also visual attention to the advertisement. The findings showed no meaningful effect of the disclosure on visual attention to the disclosure but did show that attention to the disclosure led to more critical processing of the message and more visual attention to the advertisement itself.

By means of these four studies, this dissertation shows that disclosures are able to inform receivers more effectively about microtargeting if the prominence is improved by using more attention-attracting factors such as red text and including an icon. In addition, the findings show that transparency measures can be effective in increasing receivers' knowledge and recognition of the persuasive messages they receive, which, in turn, can lead to more cognitive resistance to those messages. Moreover, recognition can lead to less favorable perceptions of the sender of the message and their credibility. Finally, the recognition of microtargeting does not lead to users minding microtargeted advertisements, but they tend to engage in more behavior to protect their online data if they have higher levels of privacy concerns.

The findings of this dissertation also imply that, while persuasion knowledge can still be used to investigate how receivers recognize and potentially cope with persuasive attempts, it is important to extend this model by adding the aspect of the technique to both the perspective of the sender and the receiver of the message. Moreover, it is demonstrated that if receivers are aware of the targeted nature of an advertisement, they can cope with it by resisting the persuasive attempt, and that this, in turn, can lead to less favorable attitudes towards the sender of the message or the person that is advertised. In doing so, this dissertation contributed to the understanding of transparency measures in the fields of persuasive and political communication.

Zusammenfassung

Während in der Vergangenheit ein Großteil der politischen Kommunikation über Fernsehwerbung und das Ansprechen von Menschen auf der Straße erfolgte, sind politische Parteien ihren Zielgruppen in letzter Zeit dorthin gefolgt, wo sie häufiger zu finden sind: im Internet, insbesondere in die sozialen Medien. Auf diesen Plattformen hinterlassen die Nutzer:innen und ihr Verhalten Spuren. Diese Informationen werden von den Plattformen zusammen mit Informationen, die die Nutzer:innen freiwillig weitergeben, wie demografische Daten oder Standortdaten, gespeichert. Die Plattformen ermöglichen es politischen Werbetreibenden, diese Daten zu nutzen und sehr spezifische Zielgruppen zu segmentieren. Anschließend können diese Zielgruppen detailliert angesprochen und mit individualisierter Werbung versorgt werden. Diese politische Werbung wird so gestaltet, dass sie bei den Zielgruppen optimal angenommen werden. Diese Methode der spezifischen Personalisierung politischer Werbung auf der Grundlage der Daten von Nutzer:innen wird als *politisches Microtargeting* bezeichnet. Es bestehen zwar Risiken bei der Anwendung dieser Technik, aber es gibt auch Lösungen, um den potenziell negativen Auswirkungen entgegenzuwirken. Mit Hilfe von Warnhinweisen können die Empfänger:innen der Nachrichten direkt darüber informiert werden, dass es sich um eine für sie personalisierte Anzeige handelt. Die Verwendung von Warnhinweisen in der politischen Werbung hat jedoch bestimmte Folgen, die im Zusammenhang mit dem Microtargeting noch nicht eingehend untersucht wurden. Diese Dissertation untersuchte daher die Auswirkungen so genannter *Targeting Disclosures* auf politische Werbung, im Hinblick auf die Erkennungsprozesse und Reaktionen der Empfänger:innen auf die Botschaft, wenn diese einen Hinweis enthielt. Zudem wurden die Auswirkungen auf die Wahrnehmung der Nachricht und derer, die sie sendeten, analysiert. Abschließend wurde der Datenschutzaspekt näher beleuchtet, indem ermittelt wurde, welche Faktoren dazu führen, dass Nutzer:innen beabsichtigen ihre Daten online zu schützen.

In der ersten Studie wurden eine Social-Media-Anzeige mit Targeting-Hinweis, eine Anzeige ohne Hinweis sowie ein unbezahltes Posting im Hinblick auf ihre persuasive Wirkung untersucht. Die Rezipient:innen beurteilten dazu ihre Wahrnehmung der Nachricht als auch die der Quelle, die sich zwischen den drei verschiedenen Bedingungen jedoch nicht signifikant unterschieden. Dies könnte dadurch bedingt worden sein, dass sich die meisten Teilnehmer:innen nicht an den Hinweis, erinnern konnten.

Darauf aufbauend, wurde in Studie II eine Anzeige mit einem interaktiven Targeting-Disclosure eingesetzt und einem auf Facebook üblichen Warnhinweis verglichen. Ziel war es

zu untersuchen, ob dies zu einer besseren Erinnerung an das Disclosure führen würde. Darüber hinaus untersuchte die Studie Resistenz als eine Möglichkeit, mit diesen Botschaften umzugehen, und ob dies zu Unterschieden in der Wahrnehmung und Einstellung der Nutzer:innen führte. Die Ergebnisse zeigten keine signifikanten Unterschiede in der Anerkennung zwischen den beiden Hinweisen. Es wurde jedoch festgestellt, dass Nutzer:innen, die sich der gezielten Ansprache bewusst sind, sich kognitiv gegen die Botschaft wehren, was mit einer weniger positiven Wahrnehmung der Quelle und einer weniger positiven Einstellung gegenüber dem Politiker zusammenhängt.

Anschließend wurde in Studie III ein Targeting Disclosure, das auf den von Instagram während der COVID-19-Pandemie verwendeten Fehlinformationswarnungen basierte, mit den derzeit von der Plattform verwendeten Hinweisen verglichen. Darüber hinaus untersuchte diese Studie die Bedenken der Nutzer:innen in Bezug auf die Privatsphäre und die Wahrnehmung des Nutzens sowie die Frage, ob dies zu der Absicht führt, die eigenen Daten online zu schützen. Die Ergebnisse zeigten, dass die Nutzer:innen durch die Targeting Disclosure stärker erkannten, dass die Werbung auf sie ausgerichtet war, was zu einer stärkeren Nutzenwahrnehmung führte. Während Datenschutzbedenken nicht durch die Hinweise beeinflusst wurden, hingen diese Bedenken der Nutzer:innen jedochmit der Absicht, ihre Online-Daten zu schützen zusammen.

Schließlich wurden in Studie IV zwei Social-Media-Anzeigen verglichen, die auf Warnhinweisen basierten, die Instagram im Rahmen der COVID-19-Pandemie verwendet hatte. Es wurde untersucht, ob Targeting Disclosures zu einer höheren visuellen Aufmerksamkeit auf den Warnhinweis führen, was wiederum mit einer kritischeren Verarbeitung einhergeht, gleichzeitig aber auch zu einer höheren visuellen Aufmerksamkeit für den Inhalt der Anzeige führen kann. Die Ergebnisse zeigten keine signifikante Auswirkung des Disclosures auf die visuelle Aufmerksamkeit für den Disclosure. Sie zeigten jedoch, dass die Aufmerksamkeit für den Disclosure zu einer kritischeren Verarbeitung der Nachricht und zu mehr visueller Aufmerksamkeit für die Werbung selbst führte.

Anhand dieser vier Studien zeigt diese Dissertation, dass Hinweise Empfänger:innen effektiv über Microtargeting informieren können, wenn die Prominenz durch die Verwendung von aufmerksamkeitsstärkenden Faktoren, wie beispielsweise rotem Text oder einem Symbol, verbessert wird. Darüber hinaus wurde herausgefunden, dass diese Transparenzmaßnahmen dazu beitragen können, dass die Empfänger:innen die persuasiven Botschaften, die sie erhalten, besser wahrnehmen und erkennen, was wiederum zu einer größeren kognitiven Widerstandsfähigkeit gegenüber der Botschaft führen kann. Darüber hinaus kann das Erkennen

des persuasiven Charakters einer Nachricht dazu führen, dass die Sender:innen der Botschaft und ihre Glaubwürdigkeit weniger positiv wahrgenommen werden. Schließlich führt das Erkennen von Microtargeting nicht dazu, dass sich die Nutzer:innen gegen Microtargeting-Werbung sträuben, jedoch neigen sie eher dazu ihrer Online-Daten zu schützen, wenn sie ein höheres Maß an Bedenken bezüglich ihrer Privatsphäre haben.

Die Erkenntnisse dieser Dissertation zeigen, dass die theoretischen Annahmen des *Persuasion Knowledge Models* immer noch hilfreich sind, um zu untersuchen, wie Empfänger:innen persuasive Appelle erkennen und möglicherweise damit umgehen. Darüber hinaus deuten die Ergebnisse der vorgelegten Dissertation darauf hin, dass es wichtig ist, dieses Modell zu erweitern, indem der Aspekt der verwendeten Technik sowohl aus der Perspektive der Sender:innen als auch der Empfänger:innen der Botschaft hinzugefügt wird. Außerdem konnte gezeigt werden, dass Empfänger:innen, die sich des zielgerichteten Charakters einer Anzeige bewusst sind, damit umgehen können, indem sie sich dem Persuasionsversuch widersetzen, was wiederum zu einer weniger positiven Einstellung gegenüber den Sender:innen der Botschaft oder der beworbenen Person führen kann. Damit leistet diese Dissertation einen Beitrag zum Verständnis von Transparenzmaßnahmen im Bereich der persuasiven und politischen Kommunikation.

Table of Contents

List of Figures

I Introduction

As large parts of users' lives occur online, the landscape of political communication has shifted with it. In recent years, political campaigns have moved from face-to-face contact to a focus on digital connections, interactions, and relationship development. Currently, one of the most common ways for political parties to reach potential voters is through social networking sites (SNSs) (Giasson et al., 2019). Users' online behavior leaves traces, especially on SNSs, where users interact with content, friends, brands, and political parties. These digital breadcrumbs often contain information that is valuable to political actors, such as consultancy firms and parties. The large volumes of information that platforms gather from usage and the information that users voluntarily share, such as demographics, likes, interest, and location data, enable political actors to segment small groups of users and narrowly target these groups with specific political advertisements (Gandy, 2000; Murray & Scime, 2010). These messages are developed to resonate with specific target audiences. The demarcation and targeting based on these data describes the concept of microtargeting, or in the case of politics, political microtargeting (PMT) (Kruikemeier et al., 2016; Zarouali, Dobber, et al., 2020; Zuiderveen Borgesius et al., 2018). Two prime examples of the use of PMT are the 2016 United States elections and the Brexit referendum in the United Kingdom. In both these examples, the political consultancy firm Cambridge Analytica allegedly gathered and used data from more than 50 million Facebook users to establish psychological profiles and target users with messages that would persuade them as strongly as possible (Cadwalladr, 2018). The firm allegedly contributed to Trump's presidential victory, and the Leave campaign's success in the Brexit referendum.

One of the major problems regarding PMT is the presumed black box, where neither laypersons, journalists, nor researchers know exactly what political consultancy firms do to microtarget users, and how and which data are used (Zuiderveen Borgesius et al., 2018). While this black box complicates the investigation of PMT, it also makes it practically impossible to estimate the effects of campaigns that parties run. While the lack of transparency and precise information about the used models makes it difficult to focus on the sender perspective of PMT, one solution is to focus on the receiver perspective. However, a major issue concerning the receiver perspective is that users often do not recognize that they receive persuasive messages that are specifically tailored and targeted at them (Binder et al., 2022; Kruikemeier et al., 2016). Disclosures are among the most transparent approaches to countering the potentially harmful effects of PMT from a receiver perspective (Binder et al., 2022; Kruikemeier et al., 2016).

Disclosures are labels often found in food packaging, alcoholic beverages, and cigarette packaging. In the field of advertising, earlier versions of disclosures have been used to inform users about the persuasive or sponsored nature of content (e.g., YouTube videos, Instagram, and Facebook posts). Regarding political advertising, Facebook advertising shows not only the fact that the message is an ad but also includes the party that paid for this ad, which potentially differs from the political party. In current developments, the EU's Digital Services Act (DSA) aims to have platforms implement disclosures that provide users with information on when and on whose behalf, content is displayed. Simultaneously, these disclosures should go beyond this and include background information about the individualized data used to target receivers and the parameters for those data points (European Commission, 2022). Although regulatory bodies aim to implement disclosures on microtargeted messages rapidly, research on targeting disclosures is scarce (e.g., Binder et al., 2022; Dobber, Kruikemeier, Votta, et al., 2023; Kruikemeier et al., 2016).

This dissertation investigates targeting disclosures through four empirical studies that build on insights from existing research. This is accomplished using four aspects embedded in the following overarching research question:

What is the impact of targeting disclosures on political microtargeted messages in terms of helping receivers recognize their targeted nature, influencing receivers' reactions to these messages, and shaping users' perceptions of the source and message, as well as their intention to protect their data?

The first aspect is the recognition of targeted political advertisements. This aspect builds on the persuasion knowledge model of Friestad and Wright (1994) and a recent adaptation in targeting knowledge (Binder et al., 2022). This theory provides insights into users' personal knowledge and beliefs about the motives and tactics of an advertisement and its source. Persuasion knowledge development depends on previous exposure to and experiences with persuasive messages. Moreover, knowledge of the topic and the sender of the message both influence a receiver's persuasion knowledge. Finally, this aspect is complemented by users' attention, which is a prerequisite for recognizing political advertisements (Jost et al., 2022).

The second aspect in this work is the receiver's reaction to politically targeted messages. Building on existing research with a communication outlook in critical processing, users process an advertisement more critically and skeptically when they recognize the

message as an advertisement (Boerman & Van Reijmersdal, 2016; Zarouali et al., 2018). Furthermore, this aspect builds on the concept of resistance. This concept describes how users' reactions to an advertisement can be to resist a persuasive message. This resistance can be cognitive (e.g., "I do not believe this") and affective (e.g., "I do not like this") (Knowles & Linn, 2004).

Through the third aspect, this work describes how receivers form new beliefs, attitudes, and perceptions after they become aware of the targeted nature of a political advertisement. This perceptive integrates receivers' adapted or newly developed beliefs, attitudes, and perceptions of the message and its sender. This aspect builds on existing work on attitudes towards the sender of a message and politician (Dobber et al., 2021). Moreover, potential fluctuations in the trustworthiness and credibility of the sender and message will be investigated (Appelman & Sundar, 2016).

Finally, the fourth aspect focuses on receivers' intention to engage in privacy protection behavior as a concept that describes the outcome of the calculus rationale described in the privacy calculus theory (Culnan & Armstrong, 1999). This rationale suggests that users outweigh their benefits and privacy concerns regarding a platform or technique to engage in different levels of privacy protection behavior.

In summary, the primary purpose of this dissertation is to examine the effects of disclosures as a measure of transparency regarding PMT. Through four experiments, this dissertation provides novel insights into the effectiveness of disclosures, and how receivers recognize targeted political advertisements through disclosures and their prior knowledge and experience. Subsequently, receivers' reactions regarding the processing of the message are investigated, and their adapted or newly developed beliefs, attitudes, and perceptions regarding the sender and message are scrutinized. Finally, the privacy aspect of microtargeting and users' perceptions of the benefits and their concerns of the technique are examined.

II Theoretical background

1. Microtargeting – current implementations

SNSs, such as Facebook and Instagram, and search engines, such as Google and Bing, offer considerable marketing potential. Through constant usage and online behavior, users of these platforms enable the platforms to create large databases that are perfect for brand promotion (Barbu, 2014). For senders of messages, these databases provide detailed information about what a user likes and dislikes, which messages they are more likely to interact with, and which messages have a higher potential to be persuasive or influential (Winter et al., 2021; Yan et al., 2009). Companies and other parties that use advertising can target specific messages at specific target groups with messages that are likely to become more effective, a practice known as *behavioral targeting* (Matz et al., 2017; Yan et al., 2009). At times when people are more online than ever, leaving breadcrumbs as they go from website to website and from page to page, behavioral targeting has become more efficient than ever. The breadcrumbs that people unconsciously leave include information about their personal lives and their (online) behavior, and sometimes this can even be used to predict personality traits based on previous interactions with websites and other online content (e.g., Facebook likes) (Matz et al., 2017; Yan et al., 2009; Zarouali, Dobber, et al., 2020). In addition, by applying intensive algorithms, senders can automatically cluster users who share attributes using machine learning to rapidly process these new data types (Papakyriakopoulos et al., 2018).

Within political communication, the overall goal is not to sell products but rather to sell the story and ideology of a politician or party to eventually receive votes. Using techniques borrowed from commercial companies and marketing agencies, political actors aim to persuade voters by spreading their stories and ideologies. While political communication targets all voters in general, the group that is most beneficial to target for political parties consists of voters who have not yet voted for the party or potential voters. This group usually consists of voters that are unsure of their partisanship, voters who are unsatisfied with previously preferred candidates or parties, and swing voters in general. The concept of dividing different types of potential voters by targeting them with certain messages is not new. Before the Internet, parties used canvassing strategies that relied on different fliers in different states or for different zip codes (Barbu, 2014; Gandy, 2000; Murray & Scime, 2010). Nevertheless, the amount of data gathered by platforms and advertisers makes it possible to target smaller groups, build look-a-like audiences, and even psychological profiles based on groups, such as the Big Five

personality traits (Matz et al., 2017; Zarouali, Dobber, et al., 2020; Zuiderveen Borgesius et al., 2018). While the goals of political parties may differ, their means of achieving them resemble approaches used by advertising agencies (Dobber et al., 2017). The use of precise targeting data and tailoring persuasive political messages, developed to resonate more effectively within specific target groups on SNSs and search engines, is referred to as political microtargeting (PMT) (Barbu, 2014; Endres & Kelly, 2018; Zuiderveen Borgesius et al., 2018).

Advertising strategies such as PMT are interesting for political parties because of their low costs and the potential to engage with as many potential voters as possible. To implement these strategies initially, few resources are necessary, which, combined with scalability, makes the approach all the more attractive to political parties (Dobber et al., 2017). PMT enables political advertisers to reach any individual or group with any message. The party that pays the most money for the placement of an advertisement "buys" the audience's attention. Disregarding the potential harm to democracy, this could ultimately lead to a situation in which only the wealthiest parties buy users' attention and thus get exposure (Bodó et al., 2017). PMT enables campaigns that previously retrieved knowledge about their targets from focus groups to know exactly which buttons to press to obtain the desired results, which in this case would be exposure and eventually votes (Gorton, 2016). These campaigns can then target groups that are more likely to vote in favor of their candidates out of the group of potential voters described earlier.

The use of PMT has the potential to benefit democracies and societies as a whole. The technique has the potential to, for instance, activate voters who are usually deemed to have a lower propensity to vote by reaching out to voters in personally relevant ways, with messages on topics that they care about (Zarouali, Dobber, et al., 2020). In addition, PMT can strengthen a person's general political interest (Matthes et al., 2022), as well as their general political knowledge (Holt et al., 2013). At the individual level, personalization of content leads to higher levels of attention, more accurate recall, and more positive evaluations of that content (Tam & Ho, 2006). Moreover, personal benefits might also lead to societal benefits when a reduction in time and cognitive effort in obtaining information, paired with higher content relevance, mobilizes individuals to vote (Kruikemeier et al., 2016). Finally, PMT can serve as an effective way to provide relevant information to voters on the issues they care about, which can lead to them being more informed and knowledgeable about those issues (Zarouali, Dobber, et al., 2020).

Although these effects of PMT could benefit societies, the overall consensus regarding the effects of the technique is negative. Multiple authors underline their concerns regarding

PMT: discouraging political participation by showing people topics they are less interested in (Bodó et al., 2017); decreasing scrutiny in the democratic process by excluding specific target groups (Jamieson, 2013); enlarging the gap in representation in governments by consciously not activating certain groups (Endres & Kelly, 2018); and manipulating voters by targeting them without their knowledge or consent (Zuiderveen Borgesius et al., 2018). Although these concerns might indirectly impact voters, they might not necessarily be users' concerns regarding PMT. In addition to societal effects, microtargeting affects individual citizens. On this individual level, users can perceive a personalized advertisement as a privacy risk and intrusive because the advertiser might 'know too much' (Segijn & Van Ooijen, 2022). This agrees with work showing that higher levels of personalization lead to higher perceived creepiness of an advertisement (De Keyzer, Van Noort, et al., 2022). In addition, users' recognition of online behavioral advertising (OBA), which can be seen as PMT's non-political predecessor, leads to more perceptions of privacy risks (Jain & Purohit, 2022). However, if users recognize a message as being politically microtargeted at them, they have been found to have more favorable party evaluations and lower intentions to engage in privacy. However, if they explicitly perceive the message as intended in a manipulative manner, the opposite is true, and users have less favorable party evaluations and a higher intention to engage in privacy behavior (Binder et al., 2022). Moreover, individuals might be more prone to believe in false information if they do not evaluate the message as persuasive or targeted by cue-taking, meaning that voters who already have an affiliation with a party or politician have less strict filters and uncritically accept the position of a party or actor (Bartels, 2002).

However, one of the biggest problems with microtargeting is that the technique appears to be a black box. Nobody knows precisely who is saying what to whom, where, and with what effect (Jamieson, 2013). This black box makes it challenging to research and practically impossible to estimate the effects of PMT (e.g., the models and algorithms used by Cambridge Analytica). Nevertheless, among the most important factors concerning PMT and its effects are users' perceptions and attitudes. While some studies have investigated users' attitudes towards PMT as a technique and their perceptions of the technique regarding society (Hirsch, Stubenvoll, et al., 2023; Matthes et al., 2022), more research on users' perceptions and attitudes regarding the source, message, and the party is needed. Thus, this dissertation investigates users' beliefs and attitudes regarding the sender and the message itself, while also investigating users' benefit perceptions, privacy concerns, and intention to engage in behavior that protects their online privacy. However, since the lack of transparency and information about the practices combined with the current lack of an existing regulatory body makes it hard to focus

on the sender side of PMT, another solution could lay in informing and/or shielding users by improving transparency regarding microtargeting, which is the focus of the current dissertation.

2. Transparency

While traditional advertisements are mostly the same for everyone and communicate to a broader audience, microtargeting tailors and targets specific segments of users based on data collection and analysis. This creates an information asymmetry where advertisers target users by means of their data, while the user is unaware of the basis of what data they are being targeted if they are aware that they are being targeted in the first place (Dobber et al., 2019). This imbalance is problematic for users because it limits their ability to evaluate political advertising, which has the potential to negatively affect their ability to make informed and autonomous decisions (Susser et al., 2018; Zuiderveen Borgesius et al., 2018). Providing users with transparent information alongside persuasive online political advertisements appears to be a solution to at least make users aware of the persuasiveness and the fact that they are being targeted (Dobber, Kruikemeier, Helberger, et al., 2023). Thus, the asymmetry should be minimized, which allows users to recognize the message as an advertisement, recognize that they might be targeted, and finally evaluate the advertisement and make informed autonomous decisions.

Existing research shows multiple countermeasures to minimize the potentially harmful effects of PMT, as described earlier. One starting point regarding transparency could be the development of *ad archives* or *ad libraries* that store all the advertisements that have run on a specific SNS, or more than one platform in the case of Meta's Facebook and Instagram that share an ad library. These libraries store all advertisements, including potential targeting measures, and make them publicly available (Dobber et al., 2019; Leerssen et al., 2019). However, other research remains critical regarding these tools that are inaccessible to some people (Dommett, 2021). While Meta has changed that and users are now able to access the platform without logging in, the larger critical point is the fact that Meta itself manages the mechanisms and nobody can track if the ads in the library are actually all the ads that run on Facebook and Instagram (Elswah & Howard, 2020). Moreover, Ben-David (2020) argues that Meta's sustained control over seemingly public data enables the company to keep citizens away from information until its content becomes the past. Other commentators emphasize that until the companies behind SNSs start regulating political advertisements that appear on their platforms, the best solution for voters is to rely on themselves, be cautious, and check their information diet, which the aforementioned tools might make possible (Ghosh, 2018).

However, this solution requires users to actively look up the advertisements they received to determine if they were microtargeted and if the message is specifically selected for them, which requires a lot of engagement for an advertisement where a user could also scroll past quickly.

2.1 Disclosures

One of the most promising transparency measures regarding informing users about the targeting practices occurring on SNSs are disclosures, more specifically targeting disclosures (Binder et al., 2022; Binford et al., 2021; Kruikemeier et al., 2016). Disclosures are labels that inform people about the content they are receiving. They explain that the content is, for instance, sponsored or that a vlog by an influencer is in collaboration with a brand, and thus, also sponsored content. However, the concept of disclosures goes back further. This concept can be linked to information labels, which are often used in the food industry. In this application, customers are informed about products being fair trade and about nutritional scores, which show a summary of the nutritional values of a product. Moreover, besides informing, the labels are also often used in a warning manner, for example, in the packaging of alcoholic beverages to show a standardized value (i.e., Alcohol by Volume, ABV; Ackermann, 2022). Some research has theoretically linked the effectiveness of disclosures to inoculation (McGuire, 1961). Inoculation theory is a biological analogy suggesting that exposure to a weakened virus stimulates people's defenses, making them resistant to infections. Similarly, individuals can also be protected from persuasion (Amazeen & Wojdynski, 2019, p. 228). Inoculation can be seen as an attempt to make people more resistant to persuasion before the persuasive attempt occurs (Dobber, Kruikemeier, Votta, et al., 2023; Roozenbeek et al., 2020). Recent research has shown that inoculation is still an effective strategy for increasing individuals' resistance to persuasion (Fransen et al., 2023).

One of the first times that companies started using disclosures regarding sponsored content was in advertorials. This form of native advertising refers to advertising within a publication that is presented in such a way that it resembles the publication's regularly published content (Wojdynski, 2019, p. 1). Since the first half of the twentieth century, the term advertorial has been used, which originated in reference to newspapers and magazine advertisements that were written in the style of news articles or editorials, and today, it continues to connote a print context. Nevertheless, the term native advertising refers to any paid promotional content that appears in a format that mimics adjacent non-sponsored content and is used not only to refer to advertisements that resemble news but also to advertisements that take the form of search engine results, vlogs, or social media content. Many experimental

studies on sponsored articles, publishers, and disclosures in a classic format (i.e., textual and usually in small font size) have found that receivers of sponsored content sometimes fail to identify the message as advertising (see Wojdynski, 2019). However, a disclosure could be the first step and a necessary condition for identifying content as advertising, especially concerning native advertising that resembles regular content (Boerman & Van Reijmersdal, 2016; Krouwer et al., 2017; Wojdynski & Evans, 2016).

Regarding social media, major platforms (e.g., Facebook, Instagram, and Twitter) use at least some form of disclosures to inform their users about the sponsored nature of certain content. Search engines, such as Google, also label sponsored search results, which are usually placed higher in their results section than unpaid search results (Jung & Heo, 2019; Van Reijmersdal et al., 2016). When advertisers on platforms such as SNSs undertake covert information-collecting strategies such as microtargeting, consumers may experience feelings of vulnerability. Information cues, such as disclosures, can be used to offset these feelings (Aguirre et al., 2015). In addition, disclosures can help users become aware of potential biases of the message's source and can integrate that information into their perceptions of that source and subsequently counter uncertainty regarding the source's intentions (Carr & Hayes, 2014). Moreover, previous studies have shown that more prominent disclosures are more effective (Beckert & Koch, 2022; Van Reijmersdal et al., 2012; Wojdynski & Evans, 2016). Prominence can be seen as a way in which disclosures stand out, in contrast to an individual's timeline. A higher prominence of disclosures means that a specific disclosure is easier to see, which leads to higher levels of recognition that an advertisement is indeed an advertisement (Amazeen & Wojdynski, 2020). In addition, prominence is found to be important because it helps receivers recognize sponsored content, especially when some disclosures are impartial (i.e., disclosures that vaguely state content is sponsored, or just contain an icon without explanations; Stubb & Colliander, 2019). Furthermore, a combination of text and a logo or symbol (compared to these aspects separately) has been found to lead to the most visual fixation (Boerman et al., 2015). Prior work also found no differences between the number of seconds receivers were exposed to a disclosure and their recall (Boerman et al., 2012). This underlines the importance of disclosure prominence, as it has been found to make a difference (Amazeen & Wojdynski, 2020; Stubb & Colliander, 2019). Furthermore, disclosures enable users to integrate the changed meaning of a message into their perceptions of both the source and the message itself (Friestad & Wright, 1994; Wojdynski & Evans, 2020). While prior research regarding disclosures mostly focuses on disclosures in the context of different communication channels, the current dissertation focuses on social media platforms, as most, if not all, microtargeting

occurs on those platforms. Furthermore, this work investigates the characteristics that contribute to prominence, such as font size and font color, and the location of the disclosure, which have been found to be important factors regarding users' recognition and, as a result, recall of the disclosures.

2.2 Targeting disclosures

It is not only research that recognizes the potential of disclosures. The European Union's Data Services Act (DSA; European Commission, 2022) states that disclosures regarding sponsored and targeted content on platforms should include information on when and on whose behalf the content is displayed. Instantaneously, the measures need to go beyond that and (if applicable) include background information about individualized data points used to target users and parameters for those data points (European Commission, 2022). While the way that the DSA aims to implement disclosures across all platforms is the first time that this would be initiated in Germany, other countries have tried to take this initiative. In Brazil, the "fake news bill", which includes (targeting) transparency disclosures, is being reviewed by Congress (Lubianco, 2022). In Europe, regulations were first proposed in Ireland (Online Advertising and Social Media (Transparency) Bill, 2021) and later adopted in France (Dobber, 2021; Lutte contre la manipulation de l'information, 2018).

While most prior research on disclosures occurred in the field of sponsored content, some studies have also implemented disclosures regarding microtargeted content on SNSs. In the case of PMT, disclosures contain more information than the fact that the message is persuasive or sponsored. While it is not yet fully obliged with the DSA, Facebook, for instance, already shows not only the party that is advertised but also the party—not always a political one—that pays for a specific advertisement. This party could be seen as an *ad buyer* since it is the party that buys the advertising space. An experimental study found that, regarding these disclosures, exposure to a personalized advertisement from a political party on Facebook activated persuasion knowledge and that this, in turn, decreased intentions to engage in electronic word-of-mouth; however, this was only the case for participants who recalled the disclosure (Kruikemeier et al., 2016). Moreover, experimental eye-tracking research showed that while users paid more visual attention to Facebook's political advertisement disclosure, the disclosure language was not effective in enhancing users' comprehension of who paid for the political ads (Binford et al., 2021). In addition, work on various transparency measures regarding political advertising in video form found that only basic transparency information shown concurrently with the start of a video advertisement had a significant direct effect on

users' attitudes towards the source of the advertisement (Dobber, Kruikemeier, Helberger, et al., 2023). They also found that only one basic transparency measure, shown before the start of a video, had a positive effect on users' attitudes towards the advertisement. Furthermore, other studies on Facebook targeting showed that disclosures explaining that users were targeted based on their demographic information, location, or preferences did not lead to significant differences in their targeting knowledge or perceptions of the manipulative intent of the sender (Binder et al., 2022). In addition, recent work found that labels in the form of traffic lights as indicators of the veracity of political ads on YouTube sometimes significantly decrease credibility perceptions (Dobber, Kruikemeier, Votta, et al., 2023). However, this was only the case for the red and orange traffic lights that were placed concurrently with the start of a political ad.

Altogether, research on targeting disclosures has shown promising results in some cases. In other studies, recall of disclosures appears to be a hurdle. Therefore, the current dissertation investigates the design characteristics that improve the recognition and recall of disclosures. In addition, this work will scrutinize whether, next to stating the political party and the ad buyer, stating specific parameters (such as age and gender) and users' data points used in those parameters—as obligated in the DSA—will actually help the recognition and recall of the disclosure. However, platforms must still comply with the regulations imposed by regulatory bodies and countries. Existing research shows the important factors regarding disclosures and their design. While different studies find results that do not always confirm the results found in other studies, the fact that regulations are upcoming could lead to more precise and uniform implementations of targeting disclosures that could even be constant regarding information and design across different platforms.

3. Recognition of the targeted attempt

Nowadays, Internet users are increasingly exposed to persuasive messages. However, users might not be as helpless as some may think. The first aspect of this dissertation focuses on users' recognition of targeted attempts. While transparency measures help users recognize persuasive attempts in both the physical and online worlds, users also develop beliefs and knowledge about persuasive messages that help them recognize these messages.

3.1 Persuasion knowledge

One of the major theories describing users' recognition of persuasive attempts is the *persuasion knowledge model* (PKM) by Friestad and Wright (1994). The PKM describes how

consumers develop personal knowledge about tactics that advertisers use to influence them and how this helps them achieve their own goals regarding the persuasive attempt (i.e., being influenced or not). The model first describes the *target* and refers to the people for whom a persuasion attempt is intended (e.g., consumers and voters). Second, the authors describe *agents* to represent whomever a target identifies as being responsible for designing, constructing, and sending a persuasion attempt (e.g., the company responsible for an advertising campaign or a political party). Third, a *persuasion attempt* is described as a target's perception of an agent's strategic behavior to present information designed to influence someone's beliefs, attitudes, decisions, or actions. From the perspective of a target, the directly observable part of an agent's behavior is defined as a *persuasion episode*. Finally, the model describes *persuasion coping behavior* as a term that describes what targets do in response to a persuasion attempt, namely, to contend or strive with the message (Friestad & Wright, 1994, p. 3). Moreover, the authors emphasize that persuasion knowledge is developmentally contingent. Individuals' persuasion knowledge develops throughout their lifespan. This means that, over time, the effects of certain actions by persuasive agents on people's attitudes and behavior will also change because individuals' persuasion knowledge shapes how they respond as persuasion targets, which could thus be seen as an iterative process. Earlier in this work, forewarning and its relation to disclosures has been described (see Section 2. Transparency). However, persuasion knowledge can also be seen as a version of forewarning that, instead of being externally supplied, is internally supplied, as it can be seen as a self-generated forewarning from what individuals believe about the situation or agent or from what they observe about the agent's behavior as the interaction occurs (Friestad & Wright, 1994, p. 18).

More recent work describes persuasion knowledge as a construct with two dimensions (Boerman et al., 2012; Rozendaal et al., 2011). The first dimension is the cognitive dimension, which the authors call *conceptual persuasion knowledge*. This dimension describes users' recognition of advertising, its source and audience, and their understanding of the advertising's persuasive intent, selling intent, and tactics (Rozendaal et al., 2011). The second dimension is *attitudinal persuasion knowledge*. This dimension includes attitudinal mechanisms that help users to cope with advertising. The attitudinal dimension includes critical attitudes, such as skepticism and disliking, applied to a specific persuasion attempt. In other words, the concept involves critical feelings regarding honesty, trustworthiness, and credibility (Boerman et al., 2012).

One of the main goals of disclosures is to help users recognize a persuasive attempt or at least inform them about the content being sponsored. Because persuasion knowledge is a

construct that investigates users' recognition of a persuasive attempt, it makes sense that research combining disclosures and persuasion knowledge has been conducted prior. Earlier work on persuasion knowledge and disclosures regarding television programs shows that a disclosure lasting three or six seconds (compared to no disclosure) has a positive effect on users' conceptual persuasion knowledge, but not directly on their attitudinal persuasion knowledge (Boerman et al., 2012). However, this study found that conceptual persuasion knowledge had a positive effect on affective persuasion knowledge. Moreover, this study finds a significant indirect effect of the disclosure on affective persuasion knowledge through the mediation of conceptual persuasion knowledge. Later research confirmed this result regarding disclosures on Facebook advertising (Boerman, Willemsen, et al., 2017). In line with this, research on native advertisements on news websites found that disclosure recognition led to higher levels of conceptual persuasion knowledge, whereas this was not true for attitudinal persuasion knowledge (Krouwer et al., 2017). In addition, work on sponsorship disclosures on television content shows that when a disclosure (compared to no disclosure) was shown prior to or concurrent with sponsored content, this increased users' recognition of the sponsored content, which the authors described as the first level of persuasion knowledge (Boerman et al., 2014). Besides, van Reijmersdal et al. (2016) showed in two studies that in the case of sponsored blogs, a disclosure (compared to no disclosure) leads to significantly higher levels of persuasion knowledge activation. In work on e-commerce, it was found that an impartial disclosure leads to lower persuasion knowledge activation compared to no disclosure and an explicit disclosure, which led to almost the same level of activation (Stubb & Colliander, 2019).

Overall, prior work confirms that persuasion knowledge helps receivers of messages cope with them by recognizing their persuasive and influential nature. In addition, the construct helps researchers investigate this recognition as a first step of coping with the overall message that senders aim at targets. While persuasion knowledge can be activated through the use of transparency measures, this dissertation aims to scrutinize whether this is different regarding targeting disclosures, as targeted political advertisements are potentially more covert and thus harder for users to recognize.

3.2 Political persuasion

While the studies mentioned above provide an overview of persuasion knowledge in different contexts, the construct is also appropriate for political advertising, especially microtargeting. While the processes behind the development and delivery of the advertisements—the actual microtargeting— differ from regular social media advertising, for

users of a platform, the advertisement could also be just another ad on a SNS. However, as described earlier (see Section 2.2 Targeting disclosures), targeting disclosures should not only inform users about the persuasive nature of an ad but also go beyond that and inform them about the targeting practices occurring on the platform. This makes it important to investigate persuasion knowledge in this context, as the results found in advertising research might not apply directly to the context of PMT. In addition, the effectiveness of the disclosures might also be different because the disclosures inform users more than they do in regular advertising, which can lead to less absorption of the message in the disclosure. Moreover, research shows that some users do not have the necessary persuasion knowledge regarding online behavioral advertising in general, which makes it impossible for them to activate it (Boerman, Kruikemeier, et al., 2017; Ham & Nelson, 2016; Nelson et al., 2021). Another study on political advertising on Twitter found that a tweet labeled as promoted led to higher levels of persuasion knowledge than a labeled tweet (Boerman & Kruikemeier, 2016). However, this study also showed that compared to tweets promoted by brands, promoted tweets from political parties activated lower levels of persuasion knowledge, disregarding the "promoted" label. Taken together, these studies show that the persuasion process in the political context is not always comparable to the results found in the field of persuasive communication.

In one of the first studies that combined disclosures and microtargeting, Kruikemeier et al. (2016) found no differences in participants' persuasion knowledge across three conditions: an organic Facebook post, an advertisement with a disclosure, and an ad with a larger explanation of microtargeting before exposure to the Facebook ad. However, the authors also ran their analyses with only the subsample that correctly recalled the labels and found that participants in the condition with the disclosure and the larger explanation had higher levels of persuasion knowledge compared to other participants who were shown an organic Facebook post. However, no differences in persuasion knowledge between the disclosure and explanation conditions were found. This means that when the disclosure was recalled, it was successful in informing users, while an extra explanation beforehand, which could be seen as the next level of forewarning compared to a disclosure, was not. Likewise, recent work on microtargeting found that if participants are exposed to transparency interventions that oblige with the DSA, they have higher levels of persuasion knowledge, whereas this was not the case for transparency measures obliging with the US Honest Ads Act (Dobber, Kruikemeier, Helberger, et al., 2023). Moreover, a recent eye-tracking study found that a disclosure shaped like a border around a political Facebook advertisement led to higher levels of ad recognition compared to the disclosure currently used by Facebook (Jost et al., 2022). Furthermore, another study found

that if targeted ads are perceived as fitted or congruent with one's beliefs (one of the objectives of microtargeting), this activates lower levels of perceived manipulative intent, which can be seen as a dimension of persuasion knowledge (Hirsch, Stubenvoll, et al., 2023).

Although research shows promising results regarding persuasion knowledge and political communication, other studies seem to find fewer effects. One reason for this is that disclosures are not always recognized by receivers of persuasive messages. While political advertisements could seem more covert in the perception of receivers, it is important to investigate whether disclosures could help them recognize these messages in order to cope with them and alter their perceptions to counter an imbalance, as described earlier (see Section 2.1 Disclosures).

3.3 Targeting knowledge

Persuasion knowledge, as described above, is a concept that helps us understand the knowledge that receivers of persuasive messages have regarding the nature of those messages and the intentions of the advertiser. However, this may not always be fully applicable in the context of political targeting (Kruikemeier et al., 2016). In a recent study, Binder et al. (2022) investigated perceived manipulative intent but also went beyond that measure and defined and investigated the concept of *targeting knowledge*. They defined this construct as individuals' beliefs about agents' use of their online data to tailor messages to recipients (p. 4). The rationale behind the adapted version of persuasion knowledge is that the original concept might not always be applicable in the context of political advertising, but also that users' knowledge about persuasion evolves as the tactics used evolve. Therefore, users are better able to shield themselves from unwanted persuasive attacks by adapting or updating their beliefs about persuasive attempts (Friestad & Wright, 1994). While the iterative process in itself is something that Friestad and Wright described in the paper in which they introduced the PKM (1994), new types of persuasive attempts have developed rapidly. As most research adapts the measurements for persuasion knowledge to a specific situation or study, targeting knowledge allows researchers to investigate the PMT with a standardized measurement instrument, which is beneficial in the fields of persuasive and political communication (Binder et al., 2022). The current work investigates this newly developed construct to not only scrutinize whether disclosures help receivers recognize the persuasive nature of a targeted message, but also internalize the targeting aspect and the fact that receivers' personal data are used to target these messages and make them potentially more persuasive.

3.4 Attention

While some individuals have developed substantial persuasion or targeting knowledge and are able to distinguish (targeted) advertisements from organic content, others lack this knowledge and can benefit more from information cues such as disclosures. One of the prerequisites for recognizing a targeted persuasive attempt by individuals who are unable to do so based on their own targeting knowledge is attention. However, participants' levels of recall regarding disclosures in persuasive communication research fluctuates. For instance, work on native advertisements in online news media shows a recall rate of 20% (Krouwer et al., 2017). Another study scrutinizing different disclosure location conditions and different types of wording used in disclosures showed a recall rate of 7% (Wojdynski & Evans, 2016). Recent research has found no differences in attention between participants who received disclosures and those who did not (Van Reijmersdal et al., 2021). However, another study showed that regarding online news, 48% of participants recognized advertising, especially when the disclosure was prominent and explicit (Amazeen & Wojdynski, 2020). In addition, a sponsorship disclosure helps participants better recognize third-party influence (i.e., sponsorship) in blog posts (Carr & Hayes, 2014). Besides, work on brand placement disclosures found that 53% of participants recalled the correct brand due to the disclosure (Janssen et al., 2016). The recall in these studies corresponds to work on microtargeting and disclosures. For instance, Binder et al. (2022) found that most participants in their study correctly recalled their disclosed information (demographic-targeted, location-based targeted, or preference-based targeted). Kruikemeier et al. (2016) found that 32% of their participants distinguished between a regular Facebook post and the same post as a personalized advertisement, even though the disclosure stated a difference. These studies find diverse results, which emphasizes the importance of further investigation of the effects of disclosures on attention, especially since, as mentioned above, results in the context of persuasive communication cannot always be transferred to the context of political persuasion or even microtargeting.

Another opportunity to investigate users' attention is through eye-tracking studies. In these studies, which mostly have an experimental nature in the case of persuasive communication, researchers have the opportunity to not only ask participants what they think they recall regarding the stimuli, but also measure their visual attention to specific areas of interest where an informational cue could be (e.g., a disclosure above a post). Eye-tracking research in the field of disclosures has been previously conducted. Research on influencer marketing videos found that disclosures before a video led to more visual attention to the

disclosure, which in turn led to significantly more disclosure recognition (Van Reijmersdal et al., 2020). Work on disclosures in television programs found that textual disclosure (compared to just a logo) led to significantly higher levels of recognition of advertisements. Moreover, other work shows that, in the middle (90%) or bottom (60%), located disclosures receive more attention than a disclosure located above (40%) a sponsored message (Wojdynski & Evans, 2016). Research that investigates visual attention concerning disclosures and political content is scarce at the moment but seems to be forthcoming. Regarding targeting disclosures in political Facebook posts, research has shown that only 30% of participants are able to correctly identify the sponsor (or the lack of a sponsor) (Binford et al., 2021). Moreover, work on digital political advertising shows that a sponsored disclosure that also stated who paid for the advertisement (41%) and a sponsored disclosure that was made more prominent, including a blue box around the whole advertisement (43%), led to slightly higher levels of recall compared to a sponsored disclosure (36%); however, these differences were not significant (Jost et al., 2022). Nevertheless, the authors found that the prominence of the disclosure was positively related to users' correct recall of a targeted ad. Overall, these results suggest that prominence is an important factor in improving the recall and effectiveness of disclosures.

While different aspects of the recall and recognition of persuasive and targeted advertisements have been investigated previously, no study has combined these factors and included a mixed-method approach regarding PMT. Therefore, this dissertation scrutinizes whether different disclosures on different social media platforms lead to differences in receivers' recall of the disclosure and, subsequently, the recognition of the content being sponsored and targeted. In addition to an experimental approach, an eye-tracking approach will provide not only self-reported data, but also behavioral information in the form of users' visual attention regarding targeting disclosures. This will help understand potential differences in users' self-reported recall and visual attention, as these two do not always overlap according to existing research (Binford et al., 2021; Jost et al., 2022). In addition, this will provide implications regarding the different types of disclosures, which are deemed important as new regulations are being developed, but no designs have yet been incorporated into those regulations.

4. Reactions to the targeted attempt

When receivers of messages develop beliefs and knowledge about persuasive attempts, they often recognize these attempts either with or without the help of a transparency measure. The second aspect of this dissertation focuses on user reactions to targeted attempts.

Recognition of targeted attempts can lead to different styles of processing and coping with messages, which can be deemed the initial reaction to the targeted attempt.

4.1 Critical processing

When users are made aware of the persuasive nature of a message, they can adopt psychological mechanisms to cope with it. When users are made aware of a persuasive attempt, they must figure out a way to effectively manage their response to that attempt (Friestad & Wright, 1994). The awareness of a message being a persuasive attempt can encourage a more systematic and biased processing of that message (Janssen et al., 2010). This, in turn, can lead users to adopt a more critical, skeptical, and effortful way of evaluating and processing messages (Boerman & Van Reijmersdal, 2016). *Critical processing* is defined as the adoption of an evaluative style of processing in which the content of a message is criticized (Boerman et al., 2014, p. 217). This effortful style of processing the message aligns with the systematic way of processing described in the *Elaboration Likelihood Model* (ELM; Petty & Cacioppo, 1986). The ELM states that if receivers have the motivation, ability, or cognitive capacity to process a message, they will do so in a more systematic way, compared to a peripheral way, if people either lack the motivation or ability to process the message at that time. The ability to process, for instance, means that if receivers are depleted because of solving tasks that demand a high cognitive load, this hinders their ability to critically process the message, resulting in less elaborate processing (Rozendaal et al., 2011). Earlier research has suggested that individuals' reactions to persuasion depend not only on their cognitive ability to critically evaluate the persuasive message, but also on their motivation to apply that ability towards the processing of that message (Kruglanski et al., 1993). However, systematic processing does not always mean that users are aware of the persuasiveness of a message and, thus, apply critical processing. Nevertheless, the *Processing of Commercialized Media Content* (PCMC; Buijzen et al., 2010), which is specifically tailored to young adults but is still applicable to the general user, states that there could be two levels of systematic processing. At the most elaborate level, *critical systematic processing* of a message involves an awareness of its persuasive nature, with the recipient actively applying relevant persuasion knowledge. At a less elaborate level, *noncritical systematic processing* can occur, which involves a high awareness of the message or the sender without awareness of the persuasive attempt. This shows that users' persuasion or targeting knowledge could play a role in their recognition of targeted advertisements, which in turn might activate the critical processing of a message.

Empirical investigations of critical processing show that it can be elicited through the use of sponsorship disclosures (compared to no disclosures) in the case of sponsored content on television (Boerman et al., 2014). Another study showed that if a message is intrusive, which can be the case for microtargeted ads, people are more skeptical of those messages (Ham et al., 2022). In the case of targeted political advertising, critical processing is also highly determined by whether the party that sends the message matches the receiver's party preferences (Binder et al., 2022). Furthermore, if the persuasive attempt is helpful for the reader, which might be the case for political messages, critical processing can be suppressed (Wilson & Sherrell, 1993). Moreover, in the case of adolescents, users do not engage in critical reflection of targeted advertising unless they are exposed to a cue that triggers their critical elaboration. This cue can be a disclosure that informs them not only about the targeted, but also the persuasive nature of the message. Without such cues, users are more susceptible to a message's persuasive effects (Zarouali et al., 2017, 2018). However, other studies do not find evidence of critical processing of a message in the case of personalized advertising (De Keyzer, Van Noort, et al., 2022), even when users are made aware of the persuasive nature of a message through a disclosure (Van Reijmersdal et al., 2013). In the case of video news releases, a tactic to send covert persuasive messages to receivers in the format of a news item, Wood et al. (2008) found no differences in users' critical attitudes towards the message. In summary, empirical research has yielded diverse findings on users' critical processing as a reaction to persuasive messages. Moreover, as mentioned earlier, it is important to investigate whether the findings regarding reactions to persuasive messages are transferable to a microtargeting context, as users might react differently or more skeptically, especially when they are made aware of the targeted nature.

In line with Boerman and van Reijmersdal (2016), this dissertation argues that the diversity of findings regarding users' critical processing of persuasive messages might indicate that they do not find every persuasive or sponsored message inappropriate (e.g., they do not mind sponsored content in movies or games, but do not appreciate sponsored content in television shows). The same rationale can be applied to the critical processing of politically targeted advertising, which has not yet been researched in the context of targeting disclosures. Thus, the current work investigates whether critical processing is activated as a psychological mechanism to cope with a political persuasion attempt, and whether this mechanism is elicited through the recognition of the message as being persuasive through the help of disclosures.

4.2 Resistance

In addition to critical processing, the existing literature states another important psychological mechanism that users can apply to cope with persuasive messages: *resistance*. According to Quinn and Wood (2004), resistance to persuasion is the process by which a person rejects influence and retains their own views. While the PKM (Friestad & Wright, 1994) proposes that people use their knowledge of a persuasive attempt to either be persuaded or resist persuasion, this might not always be the case. Reactance theory (Brehm & Brehm, 2014) states that people generally want to maintain their freedom of choice and do not want to be manipulated in the first place (Van Reijmersdal et al., 2016). Other work underlines that persuasion knowledge activation can lead to resistance since receivers might see a persuasive message as a perceived threat to their freedom (Dillard & Shen, 2005); in line with this, it is assumed that when receivers recognize persuasive attempts, they tend to resist them (Petty & Cacioppo, 1977; Quinn & Wood, 2004). Resistance, in this case, can be an increase in receivers' certainty of their initial attitudes to counter a persuasive attempt (Tormala & Petty, 2002). Additionally, advertising disclosures might alert consumers about persuasive attempts, thereby making them more vigilant and leading to higher levels of resistance (Quinn and Wood, 2004). Persuasion has been found to have different cognitive and affective dimensions (See, 2018), which makes receivers react differently to stimuli depending on which dimension is activated more (Di Plinio et al., 2022). Additionally, attitude change can be a function of one's favorable cognitive and affective reactions to a persuasive message simultaneously (Petty & Cacioppo, 1986; Rosselli et al., 1995). Although these two dimensions can be activated simultaneously, the consequences of this activation may differ. Cognitive responses, for instance, have been found to mediate the effect of exposure to a message on a persuasive outcome more than affective responses (Banerjee & Greene, 2012).

Receivers' resistance to a persuasive message can also be both cognitive ("I do not believe this!") and affective ("I do not like this"), even at the same time (Knowles & Linn, 2004, p. 7). Moreover, existing studies have argued that both affective and cognitive resistance can be activated because of persuasion knowledge activation (Fransen, Smit, et al., 2015; Fransen, Verlegh, et al., 2015). Cognitive responses to a persuasive message can be either positive or negative, depending on various factors and pre-existing thoughts, which makes cognitive resistance a negative response. Subsequently, these cognitions impact attitudes and other related responses such as behavior and perceptions. When receivers are motivated to resist a message, their responses are likely to be negative. Counterarguing is an effective and often-used strategy for resisting advertisements or persuasive and influential messages (Van Reijmersdal et al., 2016; Zuwerink Jacks & Cameron, 2003). Moreover, disclosures have been

found to increase thinking about advertising (which is something that could facilitate persuasion knowledge activation) and decrease positive or neutral cognitions about brands or advertisers, thereby activating more cognitive resistance (Boerman et al., 2012). In addition, prior research confirms that persuasion knowledge activation leads to higher levels of cognitive resistance, which, in turn, might lead to less favorable brand attitudes (Van Reijmersdal et al., 2016). Although different types of cognitions are important regarding receivers' responses to persuasive messages, receivers also tend to have affective responses to these messages. An affective response can be defined as a change in the receiver's mood or feelings after exposure to a message. Examples include receivers becoming annoyed or irritated by commercials on television or even upset because of an advertisement (Edell & Burke, 1987; Schwarz et al., 1991; Zuwerink Jacks & Cameron, 2003). A positive mood reduces individuals' motivation to systematically process content and contextual signals. Consistently, a negative mood induces more effortful, detail-oriented, and analytical processing, which can lead to skepticism towards a persuasive message (Boerman et al., 2012; Bohner et al., 1992). In line with cognitive resistance, affective resistance has also been found to be activated as a result of receivers' activation of persuasion knowledge due to a disclosure (Van Reijmersdal et al., 2016). However, existing research shows that disclosures can help resist persuasion by activating resistance (Beckert & Koch, 2022). However, this occurs only if the receiver has the cognitive ability to internalize the disclosure, whereas, in the case of cognitive depletion, disclosures do not lead to resistance (Janssen et al., 2016). When users are made aware of the targeted nature of a message, they may resist the message more because it can be perceived as intrusive or creepy, especially if users are made aware of the specific data and parameters used to target the message at them.

Although resistance and critical processing are often deemed separate mechanisms that receivers can apply, this may not always be the case. Receivers may first engage in critical processing and then resist messages. However, the existing literature shows that for these mechanisms to be activated sequentially, receivers need to have more cognitive capacity available compared to when only resistance is activated (Kruglanski et al., 1993). The suspicion of persuasive attempts or intentions to manipulate has been found to lead to resistance to persuasion, resulting in less favorable attitudes (Campbell, 1995; Campbell & Kirmani, 2000). A person's cognitive ability to critically judge new evidence is an important skill (Wright, 1975). Nevertheless, the cognitive effort it may take to translate this suspicion of a persuasive or manipulative message could be greater than simply recognizing a persuasive attempt (e.g., through persuasion knowledge). This would be a critical distinction as receivers are required

to allocate more cognitive resources to the suspicion of the attempt, which, on the one hand, might lead to more systematic processing of the message itself, but on the other hand, it might also be that they do not have enough cognitive resources left to critically process the message, let alone resist it by counterarguing with it. Moreover, it can be argued that if receivers apply persuasion knowledge to recognize persuasive attempts, they already start to critically process the message as they are aware of the source's intentions. Additionally, both advertising recognition and critical processing could be mediators that elicit resistance and reduce persuasion (Boerman et al., 2014). However, older research shows that, while the effects are small, pre-training in critical abilities may slightly increase people's resistance to persuasion (Infante & Grimmett, 1971; R. V. Wood et al., 1970). In addition, other research has found that the more critical acts are focused on arguments presented in a persuasive message, the more likely it is that the critical act promotes counterarguing, which will lead to individuals being resistant to persuasive messages (Burgoon et al., 1978). Thus, while critical processing and resistance can be viewed as two different reactions or coping mechanisms regarding persuasive messages, critical processing might also be an outcome of advertisement recognition, which in turn might lead to resistance.

Although research regarding regular persuasion shows promising results for disclosures and their relation to persuasion knowledge and, in turn, receivers' resistance strategies, no study has investigated the construct of resistance in the case of microtargeting. In a political context, the mechanisms of activating resistance, and the levels of resistance, might differ from regular persuasion. For instance, a targeted message may be perceived as more intrusive, leading to higher levels of affective resistance. Conversely, while political advertisements usually contain influential information on subjects that ask for receivers' cognitive elaboration, it could also be the case that they process the information more systematically and critically. Therefore, they might be more inclined to process the advertisement more systematically, potentially making it easier to resist the message because they are already thinking more elaborately. Thus, the current work investigates whether receivers recognize a persuasive attempt with the help of disclosures, which in turn relates to higher levels of critical processing of the persuasive attempt, as well as activation of different levels of affective and cognitive resistance.

5. Beliefs, attitudes and perceptions

If receivers of persuasive or targeted messages recognize that they are being targeted, they will implement different strategies to cope with these messages. This can lead to differing

beliefs and attitudes. Furthermore, users can develop differences in their perceptions of the sender of the message as well as the message itself. The third aspect of this dissertation discusses the changes in beliefs and attitudes.

5.1 Attitudes towards the sender

Most brand–consumer relationships are built over a longer period of time. Through repeated exposure to persuasive messages and other contact or 'touchpoints', brands try to influence consumers to buy their products or services. Through these touchpoints, consumers develop attitudes towards a brand and its products. Attitudes can be understood as object evaluation associations in an individual's memory (Fazio, 2001). In the field of political communication, the process between parties and (potential) voters resembles the brand–consumer relationship (Snyder & Ting, 2002). Through repeated exposure and contact, parties aim to persuade potential and existing voters to spread their ideologies and ultimately receive votes. In the case of potential voters, this means persuading them to vote for the party, while in the case of existing voters, this could mean bolstering their existing affiliation to ensure they cast a vote for that specific party. One way voters develop attitudes towards parties and politicians is through exposure to content on SNSs. Noticeably, the sender of the content (source) does not always have to be the person that is displayed (e.g., a post from a party that promotes a specific politician). Therefore, these attitudes do not always have to be the same (e.g., not every Republican is, was, or will be a fan of Donald Trump). Furthermore, the circumstances of the different touchpoints that parties and politicians have with voters can also influence attitudes.

In the context of this work, the most important circumstance is that users are made aware of the fact that they are being targeted by a political party or actor, which in turn can have consequences for their attitudes towards the party or actor. For instance, personalized advertising may be perceived as intrusive and a privacy risk (Segijn & Van Ooijen, 2022). Moreover, if users observe that they are being microtargeted, this has been found to decrease trust in democracy but increase political interest over time (Matthes et al., 2022). However, not all attitudinal effects of microtargeting are negative. Prior work has shown that the use of microtargeting reinforces party affiliations and makes voters less likely to deviate from their preferred party (Lavigne, 2021). Moreover, recent work on microtargeting has shown that matching messages with prior attitudes is an effective persuasive strategy (Decker & Krämer, 2023). Other work shows that targeting knowledge can be positively related to party evaluations, while perceived manipulative intent, operationalized as a dimension of persuasion

knowledge, can be negatively related to it (Hirsch, Binder, et al., 2023). Existing research on microtargeting has largely focused on the effects of the technique on users' attitudes; however, little is known about the effects that arise when platforms aim to inform their users that they are being targeted through a disclosure on a message. To further investigate this process, it is possible to build on the work done in non-political persuasive communication.

Regarding regular persuasion, informing users about the persuasiveness of a message through a disclosure has also been found to make receivers' attitudes about a brand less favorable (Wojdynski & Evans, 2016). Moreover, awareness of persuasive intentions has been shown to lead to less favorable perceptions of the message's communicator (Allyn & Festinger, 1961; Campbell & Kirmani, 2000). In the case of political communication, the sender of the message is not always the same as the politician who is advertised. Much social media content is posted in the name of the party, while the most prominent candidate of that party is usually depicted in the visual content. Therefore, users' attitude towards the politician is not always the same as their attitude towards the political party (Dobber et al., 2021). Earlier work has shown that viewers' recognition of an advertisement increases their brand memory (Boerman et al., 2015). In addition, Evans et al. (2019) found that sponsorship transparency can have a positive effect on receivers' attitudes towards the ad. Furthermore, prior work has shown that a privacy disclosure message can lead to greater trust in an online retailer (Pan & Zinkhan, 2006). Moreover, prior research has shown that if sponsored content contains a disclosure, this can lead to higher levels of persuasion knowledge, which in turn can lead to less favorable brand attitudes (Van Reijmersdal et al., 2016). Besides, critical processing of persuasive messages has been found to lead to less favorable brand attitudes (Boerman et al., 2014). However, studies investigating the effects of disclosures on targeted messages and users' attitudes towards the source of the message are scarce (Binder et al., 2022; Kruikemeier et al., 2016).

Thus, the current work investigates how users' attitudes differ if platforms or regulators make them aware of the targeting practices occurring through the implementation of disclosures. Overall, it is important to scrutinize the effects of disclosures on users' attitudes, as these attitudes might become less favorable if users are made aware of the targeting practices being employed, especially since this could be seen as intrusive, misleading, or even creepy. In turn, this could negatively affect users' attitudes towards political actors or parties in general, which could have strong implications for political parties, as the use of microtargeting along with regulated disclosures might lead to unintended effects.

5.2 Source credibility and trustworthiness

A crucial aspect of users' evaluation of a source is its perceived credibility. This plays a significant role in shaping individuals' attitudes and perceptions of various sources of information. However, users' perceptions of a source's credibility can also shape their attitudes (Ayeh et al., 2013; Kumkale et al., 2010). This reciprocal relationship makes it important to investigate both constructs separately from the perspective of PMT. *Source credibility* is a construct commonly used to indicate a communicator's positive characteristics that affect the receiver's acceptance of a message (Ohanian, 1990, p. 41). Perceived credibility influences how individuals utilize and process information from a specific source (Madsen, 2019). Existing literature shows multiple ways of describing source credibility as a multifaceted construct (Winter & Krämer, 2014). One of the most commonly used implementations of this concept is the work of Ohanian (1990), which combined eight earlier studies to develop and validate a scale to measure receivers' perceived credibility of celebrity endorsers. According to this work, the perceived credibility of a source comprises three dimensions: expertise, trustworthiness, and attractiveness. Another way to assess the credibility of sources is to distinguish between perceived competence and perceived trustworthiness (Winter & Krämer, 2014). In this dissertation, both the assessment of Winter and Krämer (2014) through two dimensions and the partially overlapping assessment of Ohanian (1990) are used to scrutinize receivers' perceptions of the credibility of the source or sender of the message.

Source credibility has been found to influence persuasion (Ohanian, 1990). This means that when receivers of messages perceive a source of a persuasive message as credible, it might lead to stronger persuasive effects of that message. Conversely, when receivers have less favorable credibility perceptions of that source, the persuasive effects might be weaker. For instance, information from high-credibility sources can lead to more positive attitude changes (Pornpitakpan, 2004; Wilson & Sherrell, 1993). In addition, higher levels of source credibility lead to more favorable attitudes than lower levels of source credibility (Tormala et al., 2006). Moreover, Krouwer et al. (2017) found that regarding brand content on news websites, if brand presence is high, the recognition of a disclosure leads to significantly lower credibility perceptions of the website itself. Moreover, resisting messages cognitively and affectively decreases the credibility of Instagram influencers (Beckert & Koch, 2022). In addition, source credibility is harmed if receivers perceive the source as biased or if the message is perceived as communicating for a purpose other than information (Hass, 2014). This might be possible if people are exposed to a political advertisement containing a disclosure that informs them about the targeting occurring.

Furthermore, credibility and trust are important when a subject of discussion—or a statement in a political ad—is something the receiver of that message has less knowledge about and when they lack access to information regarding the subject (Yoon et al., 2005). While politicians are often considered experts who specialize in a specific subject, their knowledge will be understood as being more valid if they are perceived as more credible and trustworthy. In the voting process, trust has been found to directly influence voters' choice of a political candidate (Hetherington, 1999) and their intention to vote in their favor (Housholder & LaMarre, 2014). This means that, if a potential voter considers a candidate to be more trustworthy and capable, they are more likely to vote for them. However, credibility is not only perceived regarding the source but also influenced by the channel of communication, which in turn may lead to different perceptions regarding the credibility of the source. While studies from the beginning of this century show that citizens found traditional television news more credible than digital channels concerning political information, recent work shows more diverse results (for an overview, see Flanagin & Metzger, 2014, p. 4).

While research on credibility and microtargeting is scarce, there is some evidence of this relationship, and most authors are cautious about their statements. However, there is also information from political communication to build upon. Research suggests that microtargeting tactics could result in decreased evaluations of source credibility (Brown-Devlin et al., 2022). Nevertheless, other work finds that the use of microtargeting in campaigns might lead to backfire effects if candidates are perceived as lacking credibility, whereas if candidates are perceived as highly credible, microtargeting appears to be effective (Pilditch & Madsen, 2021). Moreover, research argues that microtargeting could be a risk within the process of campaigning, as using messages that differ too much from the parties' ideology can lead to a loss of credibility. However, credibility is an important factor, especially in a multiparty system where more distinctiveness might be needed to prevail (König, 2020). Research on promoted tweets sent by politicians and brands finds that the recognition of advertising through persuasion knowledge leads to decreases in source trustworthiness and less favorable attitudes and causes higher skepticism towards a tweet when the message comes from a political party compared to a brand (Boerman & Kruikemeier, 2016). Other research investigated whether the disclosure of information about the targeting taking place influences the trustworthiness of the source, but found no results in their sample (Kruikemeier et al., 2016). In this study, the concept of trustworthiness is investigated separately and interpreted as a user's assessment of only the honesty of the communicator.

The rationale behind this is that if people are made aware of the targeting occurring on a platform through the disclosure, this may cause a difference in their perceived trustworthiness of the source of the post (Ohanian, 1990). Trustworthiness is a concept that focuses more on the reliability and honesty of a source than on its expertise and attractiveness (Sparks & Rapp, 2011). When people understand the persuasiveness of a message, the trustworthiness of the sender decreases as the persuasive nature of the message becomes known (Main et al., 2007). This is in line with other research that found that when people are confronted with personalized communication, they trust online companies slightly less (Bol et al., 2018). People often give more credibility to a message from a source they trust (Greer, 2003; Lavigne, 2021). The dimensions of credibility discussed earlier do not mean that people automatically support the party that they think is credible, especially in a multiparty system. Although credibility remains a valid indicator of potential voting behavior, it is important to recognize the complexity of voting as a process. Moreover, looking at these dimensions, it is questionable whether a user weighs the attractiveness of a politician as strong as their expertise or the trust they have in this politician. Therefore, in this dissertation, credibility and trustworthiness are investigated separately and through different dimensions (Kruikemeier et al., 2016; Ohanian, 1990; Winter & Krämer, 2014). Thus, the current work investigates how users' credibility and trust perceptions of the source of the message differ if they are made aware of an advertisement being targeted at them through the implementation of disclosures. While the information that messages are persuasive and targeted can lead to a change of meaning of the advertisement, this potentially lowers perceptions of credibility and trust towards the source of the message. In addition, receivers might resist the message more after being (made) aware of its persuasive and targeted nature, which potentially decreases perceptions of credibility even further. This, in turn, can be problematic for political parties, as credibility and trust have been found to influence potential voters' choices of political candidates.

5.3 Message credibility

In addition to users' perceptions of the source, credibility can be a multifaceted construct that is relevant at different levels (Metzger et al., 2003). It can refer to an entire medium, a source within a medium, or single messages (Sundar & Nass, 2001; Winter & Krämer, 2014). While the credibility of a message can overlap with elements of users' credibility perceptions of the source, Appelman and Sundar (2016) showed that message credibility is indeed a construct distinct from source credibility. Their work concluded that the three main proxies that reflect users' perceived credibility of a message are its accuracy,

authenticity, and believability. In the current dissertation, the definition of Appelman and Sundar (2016) is used to explain message credibility as "an individual's judgment of the veracity of the content of communication" (p. 5). Additionally, research notes that message credibility refers to a receiver's perception of the believability of a message (Metzger et al., 2003).

Research on political advertising confirms that message credibility and source credibility are distinguishable constructs (Dobber, Kruikemeier, Votta, et al., 2023). In that study, traffic light veracity labels were used as warning labels for misinformation. This study found that showing an orange traffic light concurrent with a video decreases source and message credibility, whereas a red traffic light warning concurrent with a video only decreases message credibility and not source credibility. However, research on disclosures in influencer marketing finds that sponsorship compensation justification increases message and source credibility as the sender is deemed more honest (Stubb et al., 2019; Stubb & Colliander, 2019). Moreover, receivers' attitudes towards a message are also important and do not always align with their perceptions of the source. If people primarily have negative thoughts in response to a message, then the effects of source credibility are reversed. This means that higher levels of source credibility lead to less favorable attitudes compared to lower source credibility (Tormala et al., 2006).

In the case of political targeting, it is important to scrutinize both the source and message credibility. As mentioned earlier, voters' political processes and attitudes are complex, and sometimes change over time (Hmielowski et al., 2020). This contrasts with the potential reactions that receivers of a targeted message might have due to a disclosure being placed on a message, as this might be a direct reaction compared to a reaction over a longer period. As the disclosure directly informs the receiver that the message is targeted, this might not only lead to differences in the credibility perceptions of the source, but also directly lead to receivers deeming a message as less credible. Thus, this dissertation investigates whether users' message credibility differs if they receive a disclosure label on a microtargeted message. With more regulations being implemented, there is no way not to label content with disclosures to improve transparency. This enlarges the relevance of investigating whether the use of these labels leads to different credibility perceptions and whether these perceptions differ regarding the source and message. For instance, it is possible that the source is deemed even more credible because of improved transparency but that receivers find the message less credible because they are being targeted. In addition, the awareness of being targeted could also lead to primarily negative thoughts, which could reverse the usual effects of source credibility.

6. Targeting and privacy

The fourth and final aspect of this dissertation describes the relation of microtargeting and what this means for users' privacy perceptions, as microtargeting is a technique that uses large amounts of data, mostly behavioral, but also self-disclosed. Nevertheless, most users are unaware that their data are being used, if they are aware that they are being targeted at all (Dobber et al., 2019). The use of data for advertising practices, such as microtargeting, can cross boundaries regarding users' privacy on SNSs. An example of this is work that shows that online behavioral advertising (OBA), which can be seen as the predecessor to PMT, is often perceived as a privacy risk and intrusive (Segijn & Van Ooijen, 2022). However, the same research points out that OBA also has its advantages, with personal relevance, improved advertising value, and economic benefits being the top three perceived benefits. Culnan and Armstrong's privacy calculus theory (1999) helps us better understand online privacy and users' perceptions of it. The privacy calculus theory states that individuals weigh privacy costs and benefits before disclosing personal information and are more likely to disclose this information if the benefits are at least balanced with, or greater than, the privacy costs (Dinev & Hart, 2006). Within the domain of social media, this can be regarded as a subjective experience of privacy, as delineated in Trepte's social media privacy model (2021). According to this framework, users' privacy perceptions are shaped by their usage goals, contextual boundary conditions, and privacy mechanisms. Subsequently, these privacy perceptions prompt diverse forms of privacy regulation behaviors, such as altering self-disclosure and control by restricting access to information (Trepte, 2021). Furthermore, empirical work by Dienlin and Metzger (2016) showed that an extended privacy calculus that includes users' self-withdrawal behaviors and privacy self-efficacy holds true. Notably, regarding self-disclosure, benefits outweigh privacy concerns, whereas in the case of self-withdrawal, privacy concerns outweigh both self-efficacy and benefits. Similarly, other work has shown that, in the context of mobile application usage, such as COVID-19 warning apps, both privacy concerns and perceived benefits are predictive of app usage (Meier, Meinert, et al., 2021).

Despite the existence of theoretical frameworks (Schäwel et al., 2021), investigations of aspects of privacy regarding microtargeting over time (Dobber et al., 2019; Stubenvoll et al., 2022), and targeting privacy investigations with an adolescent sample (Zarouali, Verdoodt, et al., 2020), an empirical investigation of the privacy calculus in a microtargeting context has not yet been conducted. However, the rationale grounded in the calculus could explain users' behavior regarding their online privacy in the context of microtargeting. For instance, users

might experience privacy concerns through the use of their personal and behavioral data, while also having concerns regarding their data protection (Binder et al., 2022; Dobber et al., 2019; Segijn & Van Ooijen, 2022). Simultaneously, users may experience higher usefulness by receiving political advertisements that match their personal preferences or interests (Barocas, 2012; Binder et al., 2022). These coexisting experiences and perceptions can engender shifts in attitudes towards PMT techniques and users' intentions to engage in privacy/data protection behavior. Engaging in more of this behavior could protect them against data usage by platforms and advertisers, while engaging less in this behavior might give platforms and advertisers more information and data to work with, potentially leading to more fitting targeted ads, and thus, higher usefulness. Hence, the current dissertation applies the perspective of the privacy calculus to the context of microtargeting, while analyzing the perceived benefits and privacy concerns that users might experience concerning PMT. Subsequently, it will be scrutinized whether these perceptions and concerns lead to differences in users' intentions to engage in privacy protection behavior. These different aspects of the privacy calculus theory in the context of microtargeting are discussed next.

6.1 Benefit perceptions

The benefits aspect of the rationale employed in the privacy calculus theory can be described as the advantages that users receive in exchange for their online data, both self-disclosed and gathered through behavior. In existing work, the benefits of PMT are primarily explained on a societal (or macro) level, and the consensus seems to be rather negative. However, microtargeting has also been found to provide benefits. For instance, the use of the technique can activate potential voters who are usually deemed to have a lower propensity to vote by reaching out to them with messages that are personally relevant, in ways that they find personally relevant as well (Zarouali, Dobber, et al., 2020). This, in turn, potentially strengthens general political interest among all voters (Matthes et al., 2022; Zuiderveen Borgesius et al., 2018). Nonetheless, these benefits are at a macro level and may not apply to users in the direct moment they receive a targeted message. At the individual level, personalization of content has been found to lead to higher levels of attention, more accurate recall of information, and more favorable evaluations of this content (Tam & Ho, 2006). In contrast, other work does not find that more personalization leads to higher levels of perceived relevance, but the authors note that this could be a result of the fact that users might not be aware of the amount of information that platforms gather and subsequently the processing of this information, which is done to tailor advertisements or messages to them (De Keyzer, Dens,

et al., 2022). Moreover, in a study on disclosures and personalization, Brinson and Eastin (2016) found that if consumers recognize an advertisement as being personalized because of a disclosure, they have more favorable attitudes towards the ad.

Furthermore, individual-level benefits might also lead to societal benefits. For instance, the reduction in time and cognitive effort required to obtain information and higher content relevance can mobilize people to vote (Kruikemeier et al., 2016). In addition, earlier work on OBA found that this technique can help narrow down alternative solutions to the most relevant and helpful information (Kerem & Ulla, 2018). Similarly, Zarouali et al. (2020) concluded that PMT can serve as an effective way of providing relevant information to citizens on issues they care about, which potentially leads to citizens being more informed and knowledgeable about these issues. Finally, personal relevance could lead to higher motivation to process information, which can subsequently lead to more central attitude changes, as described in the ELM (Petty & Cacioppo, 1986).

Thus, following the rationale of the privacy calculus theory, the current work investigates the benefit perceptions that users might have when they are made aware of the targeting taking place by means of disclosures. The recognition of political targeting may lead to direct benefit perceptions if people perceive the positive aspects of personalized advertising. Moreover, this makes it possible to investigate users' benefits, privacy concerns, and their subsequent intention to engage in privacy protection behavior according to the privacy calculus rationale.

6.2 Privacy concerns

While users may experience benefits from receiving personalized political advertisements, the overall consensus on microtargeting and its effects on society appears rather negative. As microtargeting demands personal information, this technique potentially conflicts with users' privacy concerns (Papakyriakopoulos et al., 2018). Multiple researchers have highlighted potential concerns regarding PMT: decreasing scrutiny in the democratic process by excluding certain target groups (Jamieson, 2013); discouraging political participation (Bodó et al., 2017); enlarging the gap between citizens and their representation within governments (Endres & Kelly, 2018); and manipulating voters by targeting them without their knowledge or consent (Zuiderveen Borgesius et al., 2018). However, these concerns are on a societal (or macro) level, meaning that they potentially harm individuals eventually but are not necessarily concerns for users at the direct moment they receive a targeted message.

On an individual level, users perceive certain privacy concerns if they are made aware of the personalization and microtargeting that occurs. Users may perceive an advertisement as a privacy risk and intrusive (Segijn & Van Ooijen, 2022). This is in line with work showing that higher levels of personalization lead to higher levels of perceived creepiness of advertisements (De Keyzer, Van Noort, et al., 2022). A recent study that validated a scale on perceived surveillance concerning personalization effects found that users experience creepiness, surveillance concerns, high perceptions of privacy risks, and privacy concerns (Segijn et al., 2022). Additionally, research has found that in the case of data-driven OBA, individuals who are aware of the persuasiveness of an advertisement experience higher levels of privacy risks, a construct that could be seen as a substitute for privacy concerns or at least the costs-side of the privacy calculus (Jain & Purohit, 2022). Moreover, research has found that privacy concerns correlate with perceptions of manipulative intentions (Hirsch, Stubenvoll, et al., 2023). Additionally, research has found that higher levels of privacy concerns lead to more cognitive avoidance over time (Stubenvoll et al., 2022). Finally, Dobber et al. (2019) found that privacy concerns regarding microtargeting lead to less favorable attitudes towards the technique itself, and that reversibly, favorable attitudes towards the technique lead to a decrease in privacy concerns.

Existing research indicates that targeting practices are associated with users' privacy concerns and risk perceptions. Therefore, this dissertation aims to examine whether these relationships vary when users are informed of the targeting practices that occur on platforms. The work then investigates the effects of recognizing targeting through the use of disclosures and its impact on users' privacy concerns and, subsequently, their intention to engage in privacy protection behavior in accordance with the privacy calculus rationale.

6.3 Privacy protection behavior

The rationale described in the privacy calculus theory states that users outweigh benefit perceptions and privacy risks or concerns to adapt the amount of information they self-disclose on social media. However, recent research has provided a different outcome for this rationale in the concept of privacy protection behavior (Meier, Schäwel, et al., 2021). Examples of privacy protection behavior range from altering the privacy settings on a social network or using software to disguise oneself (e.g., a VPN or the TOR browser) to deregistering from a platform altogether (Schäwel et al., 2021). In the case of PMT, the concept of intention to privacy protection behavior is a better fitting outcome measure for the privacy calculus approach, as users are often not aware of the information that is used to target them, which, in

turn, is not always self-disclosed. Earlier work shows that higher perceptions of privacy risks lead to less intentions of self-disclosure and a higher desire for privacy protection, which, in turn, leads to more intention to use a tool to protect oneself online (Meier, Schäwel, et al., 2021). Additionally, higher levels of privacy concerns have been found to lead to higher intentions to withdraw information from social networks and lower intentions to disclose information (Dienlin & Metzger, 2016). Besides, earlier work found that when people feel that their online privacy has been violated, they aim to implement more privacy protection measures (Büchi et al., 2017). However, these results may not always be directly applicable to the context of personalization and microtargeting.

Regarding advertising personalization, existing work suggests that the perceived costs side of the privacy calculus (or privacy concerns) outweighs users' benefits perceptions (De Keyzer, Van Noort, et al., 2022). In the context of microtargeting, higher perceptions of manipulative intent have been found to increase users' intentions to engage in privacy protection behaviors (Binder et al., 2022). Other research shows that privacy concerns indeed lead to higher intentions to engage in privacy protection behavior, but this does not lead to users actually applying ad-blockers to block targeted ads (Stubenvoll et al., 2022).

The privacy calculus theory provides a suitable framework for analyzing user perceptions related to privacy, specifically in the context of microtargeting. Microtargeting involves using individuals' data to direct messages to them, making privacy a highly relevant concept. This dissertation examines users' intentions to engage in privacy protection behavior as an outcome of the privacy calculus rationale, considering the impact of disclosures on users' recognition of targeted attempts. While some users may be motivated to take precautions to protect their data and identity upon becoming aware of targeting practices, others may perceive greater benefits from tailored and targeted advertising and consequently have reduced intentions to do so.

7. Research objectives

The previous chapters of this dissertation reviewed the current state of microtargeting transparency research and the role of disclosures and discussed the potential effects of the use of these labels on receivers' perceptions and attitudes. In the introduction to this dissertation, four aspects that can help answer the questions that arise from the literature review are described. However, before focusing on these aspects, it is important to address questions related to the disclosures themselves. It is important to investigate the different design characteristics that have proven to be effective in the context of microtargeting. Moreover, the

regulations provided in the DSA should be considered when scrutinizing the characteristics of disclosures. Therefore, all the studies in this dissertation investigate targeting disclosures through experimental designs. Study I compares the effects of a Facebook advertisement containing a targeting disclosure to one containing a sponsored disclosure or an organic post without a disclosure. Study II compares a salient, prominent, and interactive targeting disclosure to a static sponsored disclosure on a Facebook advertisement. Studies III and IV compare a sponsorship disclosure to a targeting disclosure on Instagram.

The first aspect in this dissertation investigates what the effects are when receivers of a targeted political advertisement recognize the targeted nature of the message with a disclosure. All studies in this dissertation scrutinize whether disclosures lead to differences in users' recognition and awareness of the targeted nature of advertisements. In Studies I and II, this is done by investigating persuasion knowledge (Friestad & Wright, 1994) in the context of microtargeting. In Study III, this is done by investigating targeting knowledge (Binder et al., 2022). Finally, Study IV uses behavioral eye-tracking data to investigate visual attention.

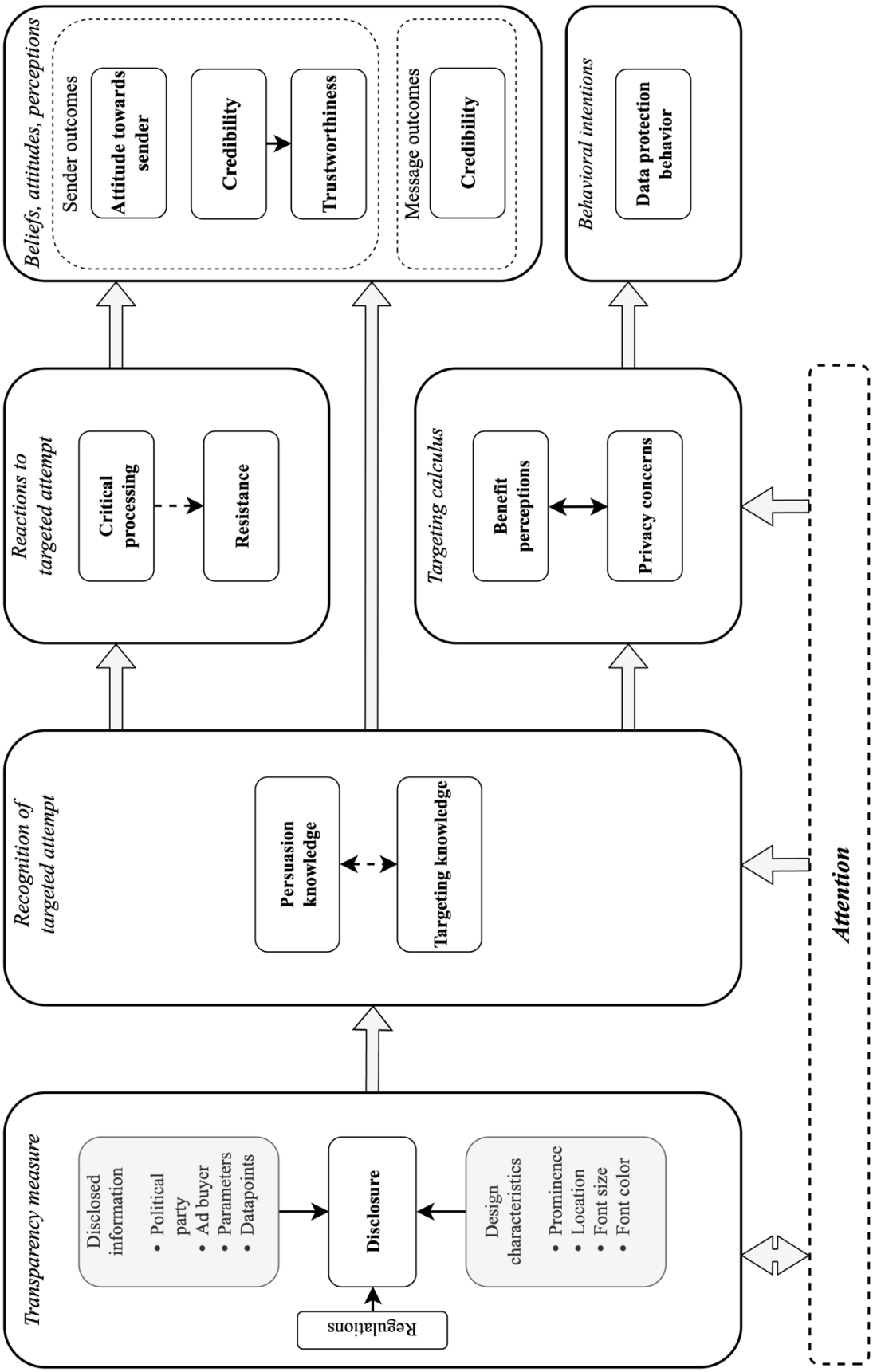
The second aspect of this dissertation investigates users' reactions to targeted political advertisements. This aspect scrutinizes whether awareness of the targeted nature of the advertisement, through disclosures and the activation of persuasion or targeting knowledge, leads to different ways of psychologically processing the message. In Study II, this is done by investigating whether users resist the message in a cognitive and affective way after recognizing the targeting that is used (Knowles & Linn, 2004). In Study IV, this is done by investigating whether users activate critical processing of the message after they recognize the targeted nature of the attempt (Boerman & Van Reijmersdal, 2016).

The third aspect of this dissertation investigates how receivers of targeted messages form different beliefs, attitudes, and perceptions regarding the message and its source after they become aware that the advertisement is a politically targeted message. In Study I, the focus is on source credibility, trustworthiness and message credibility (Appelman & Sundar, 2016; Kruikemeier et al., 2016). Concurrently, Studies I and II both focus on source credibility (Ohanian, 1990).

The fourth and final aspect of this dissertation investigates receivers' perceived benefits and privacy concerns regarding microtargeting. Following the rationale presented in the privacy calculus theory, this aspect investigates users' intention to engage in privacy protection behavior as an outcome of the weighing of benefits and privacy concerns regarding microtargeting as a technique (Culnan & Armstrong, 1999). Study III uses this rationale to describe the privacy aspect of microtargeting and the use of receiver data.

In summary, the four studies in this dissertation will contribute to the theoretical and practical understanding of using disclosures as a means of transparency regarding microtargeting, the psychological processes that users apply to recognize these advertisements, and subsequently cope with them while scrutinizing how this changes their perceptions regarding the source and message. Moreover, this work will investigate users' benefit and risk perceptions regarding PMT as a technique, and whether this leads to intentions to protect their data online. A visualization of these studies in sum is depicted in Figure 1.

Figure 1. Conceptual model of the four studies in this dissertation



III Summary of research papers contained in the cumulus

The following chapter summarizes the four research articles included in this dissertation. Each article combines different perspectives on receivers' responses to political advertising and targeting disclosures. This chapter provides an overview of the potential of disclosures to enhance transparency.

Article I: Empty Transparency? The effects on credibility and trustworthiness of targeting disclosure labels for micro-targeted political advertisements (Jansen & Krämer, 2023)

Drawing on the persuasion knowledge model and source credibility literature, this study aimed to identify how users respond to a targeted political advertisement if there is a disclosure stating that it is one. Participants were exposed to either a regular Facebook post without any disclosure, a regular political Facebook advertisement with the disclosure that Facebook uses on its platform ("Sponsored, paid for by …"), or a Facebook advertisement with a more salient targeting disclosure ("This content has been targeted at you based on your online behavior"). Subsequently, the study investigated what the different types of disclosure meant for users' credibility perceptions of the source and message and their perception of message trustworthiness. Moreover, this study investigated the potential mediating effect of persuasion knowledge on the effect of each disclosure on credibility and trustworthiness.

This study builds on earlier work by Kruikemeier et al. (2016), which showed that if individuals notice disclosure labels on a microtargeted Facebook post, they better understand that the post is an advertisement through the activation of persuasion knowledge. This mechanism helps users identify persuasion attempts based on previous exposure and experience and can be influenced by awareness of the advertisements' topic and prior knowledge of the sender of the message (Friestad & Wright, 1994; Jung & Heo, 2019). However, this study goes beyond the work of Kruikemeier et al. (2016) by including source and message credibility as dependent variables. Both constructs were important predictors of the voting process. Richmond and McCroskey (1975) found that the credibility of opinion leaders, to which individuals were exposed regarding upcoming elections, significantly influenced voters' candidate choices. Moreover, source credibility positively predicts intended political participation towards a candidate or party and users' perceptions of credibility regarding messages from that candidate (Housholder & LaMarre, 2014). In addition, trust directly influences voters' choices of political candidates (Hetherington, 1999).

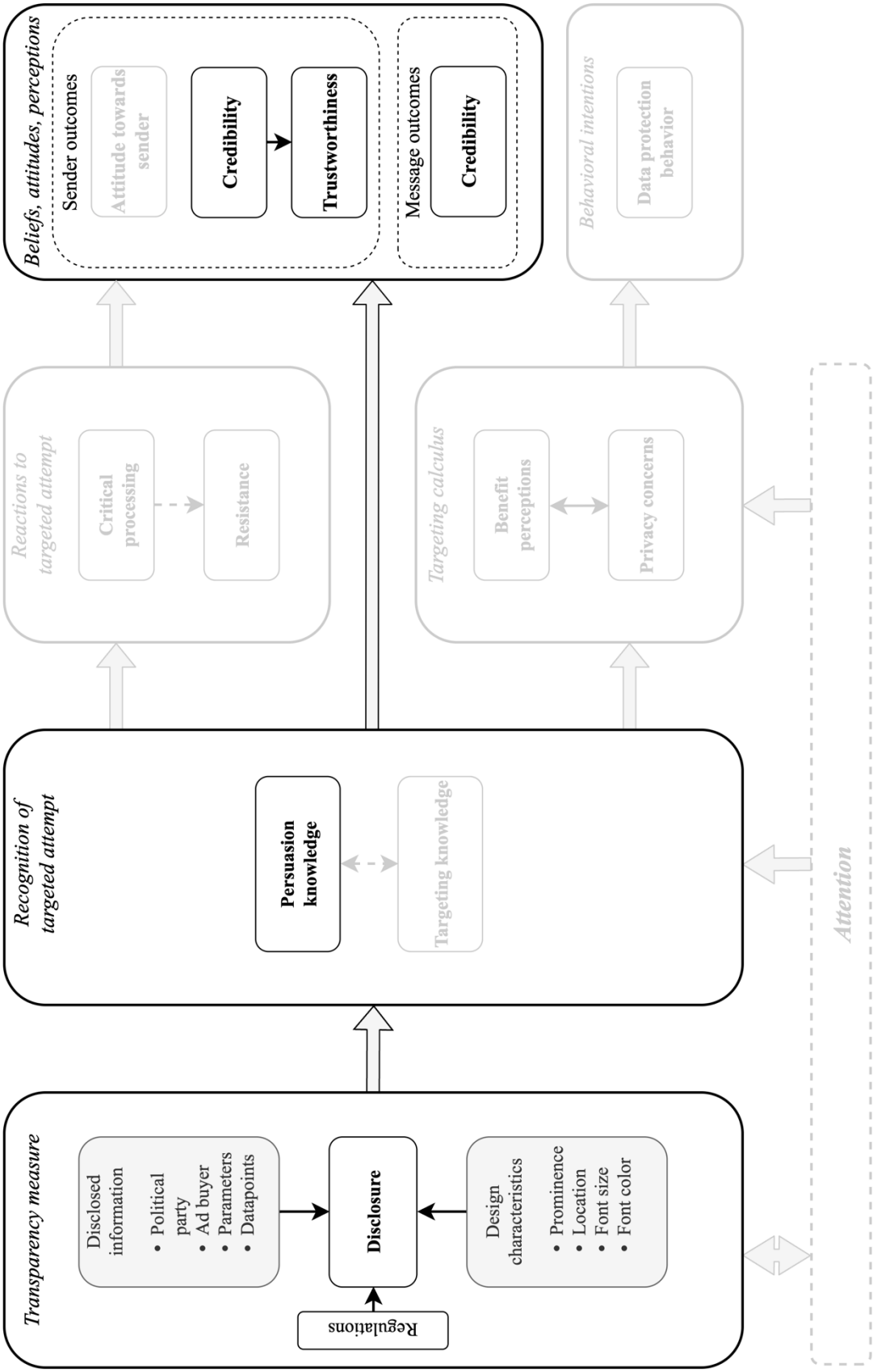
In this study, microtargeting was simulated by exposing participants to an advertisement that corresponded to their beliefs. To achieve this, participants were asked whether they agreed or disagreed with statements on climate change regulations. These statements were pre-tested with 100 participants who saw them in a battery of eight statements to determine which statements clearly described the view of someone who favors climate change regulations (the view of someone who wants to counter climate change) or against climate change regulations (the view of someone who does not believe in climate change and thinks the regulations do not work). Four of these pre-tested statements, most clearly pro and against, were used in the main study. This enabled us to show participants the stimulus material with a statement that was more in line with their own views.

In the online experiment, 227 adult German Facebook users were recruited and randomly assigned to one of the experimental conditions. The study used a one-factorial (control vs. sponsored disclosure vs. targeting disclosure) between-subjects design. The data in this study were analyzed using the PROCESS Macro, including bootstrapping with 1,000 samples (A. F. Hayes, 2018). Study I revealed no significant effects of manipulations on source and message credibility or source trustworthiness. Moreover, there were no significant effects of the manipulations on persuasion knowledge. Consequently, there were no mediating effects of persuasion knowledge on the effect of the manipulations on source and message credibility, or source trustworthiness. To further investigate the results of Study I, Bayesian analyses were performed. By adding Bayesian hypothesis testing, the probability of the observed data given the null hypothesis was compared with the probability of the observed data given the alternative hypothesis. The Bayes Factor (BF) is the ratio of these probabilities and can be interpreted as the weight of evidence in support of the null hypothesis versus the alternative hypothesis. Study I relied on Wagenmakers et al. (2011) for this. In line with the frequentist analyses that were carried out through the PROCESS Macro, the results showed substantial evidence in favor of the absence of all hypothesized effects.

The most consistent observation throughout Study I was that participants did not notice any disclosures in the experimental conditions. Although this was not the main objective of this study, this result aligns with other research on disclosures in the field of microtargeting (Kruikemeier et al., 2016) and regular advertising (Evans et al., 2017; Van Reijmersdal et al., 2021; Wojdynski & Evans, 2016). However, most previous studies have shown higher levels of disclosure recall than those observed in the current study. Although it was not the goal of this study to evaluate the effectiveness of Facebook's disclosure practices, this study shows that the company uses disclosures that are too subtle (Boerman & Kruikemeier, 2016;

Kruikemeier et al., 2016), and thus do not get noticed enough to be encoded in a manner that improves transparency (Binford et al., 2021). While other studies have demonstrated promising results concerning the recall of information embedded in disclosures, the most significant prerequisite is that disclosures need to be noticed first (Binford et al., 2021; Van Reijmersdal et al., 2016). The current study shows that while platforms try to keep everything within their corporate layout, currently used disclosures lack prominence and are not likely to be seen and recalled by users; thus, they do not contribute to improving transparency about microtargeting. Figure 2 depicts a schematic overview of this study.

Figure 2. Schematic overview of Study I



Article II: Time for transparent targeting: an investigation of targeting disclosures, coping mechanisms, credibility, and political attitude (Jansen et al., 2023)

While Study I focused on the effects of targeting disclosures on users' credibility and trust perceptions of the source and message, Study II aimed to build on this and entangle the processes that occur between users receiving the message and their perception development. Following Study I, this study investigated users' recognition of targeting disclosures through the persuasion knowledge model (Friestad & Wright, 1994), and aimed to investigate users' reactions to the persuasive attempt through the concept of resistance to the persuasive targeted message as a coping mechanism. Building on existing work in the field of persuasive communication, a distinction between cognitive and affective resistance was made to gain novel insights into the process of users receiving targeted political ads. Furthermore, in line with Study I, this study examined source credibility and another important concept in users' attitudes towards the advertised politician.

Prior research has shown that when an advertisement is recognized as being an advertisement, persuasion knowledge is activated, and, in turn, resistance could be activated as a coping mechanism that negatively affects users' attitudes towards the content (Youn & Kim, 2019). Moreover, existing studies show that when persuasion knowledge is activated as a coping reaction to a persuasive attempt, receivers use their existing knowledge to either be persuaded by the message or resist it (Sagarin et al., 2002; Tormala & Petty, 2004; Van Reijmersdal et al., 2016). Quin and Wood (2004) defined resistance as the process through which one rejects influence and retains one's views. One reason for resisting a message is that receivers feel that the persuasive message is a perceived threat to their freedom (Dillard & Shen, 2005). Users aim to counter this threat of freedom by activating resistance, increasing the certainty of their initial attitudes (Tormala & Petty, 2002).

People's attitudes contain distinctive cognitive and affective components, with valence dimensions ranging from negative to positive (Ostrom, 1969; Rosenberg et al., 1960). Persuasion can be considered an attempt to change existing attitudes towards a brand or, in the case of PMT, a political actor. Existing work shows that users also process persuasion through both cognitive and affective dimensions (See, 2018) and that their reaction to persuasive stimuli depends on which dimension is activated more (Di Plinio et al., 2022). Additionally, attitude change can be a function of one's favorable cognitive and affective reactions to persuasion at the same time (Petty & Cacioppo, 1986; Rosselli et al., 1995). Like an individual's attitudes,

resistance can also be both cognitive ("I do not believe this!") and affective ("I do not like this!") (Knowles & Linn, 2004, p. 7; Sagarin & Cialdini, 2004). Earlier studies on persuasive messages have shown that affective and cognitive resistance might be activated as a result of persuasion knowledge activation (Fransen, Smit, et al., 2015; Fransen, Verlegh, et al., 2015). Therefore, Study II investigated both routes simultaneously and scrutinized the potential differences in resistance and psychological effects on users. The two types of resistance were measured as *counter-arguing* for cognitive resistance and *negative affect* for affective resistance (cf. Van Reijmersdal et al., 2016)

While this work built on Study I through the inclusion of source credibility, it also went beyond that and included attitude towards the politician as a concept that investigated an additional step after credibility perceptions. Users' attitudes towards a source are influenced by their perceived credibility of that source (Kumkale et al., 2010). Moreover, existing work has found that in the case of disclosures and brand attitudes, both cognitive and affective resistance are negatively related to receivers' attitudes (Van Reijmersdal et al., 2016). Furthermore, reactance to a message has been found to negatively affect brand attitude (Beckert & Koch, 2022).

Although this study used the same microtargeting replication as Study I (see Article I), a newly developed type of targeting disclosure was used. Participants in the control condition were exposed to a political Facebook advertisement containing the platforms' current sponsored disclosure ("Sponsored, paid for by …"). The reason for this is that Facebook advertisements without a disclosure are nonexistent. This means that the message is deemed organic content and nobody paid to expand the audience of this message beyond the people who like the page that posted the message, whereas disclosures on Facebook are mandatory and regulated in the case of sponsored content. In the targeting disclosure condition, the manipulation consisted of a disclosure based on one of the experimental conditions in Study I. The disclosure stated that the participant was targeted with this message ("This message is targeted at you based on your age, gender, and online behavior"). However, interactive components were added to this disclosure. When hovering over the terms age, gender, and online behavior, which were red and underlined to imply interactive possibilities, participants saw a pop-up box that included information about these parameters. This information (in the case of age and gender) was based on the demographic information that participants provided after they agreed to the informed consent at the beginning of the experiment. Participants' registered age was converted into groups (e.g., a participant who answered their age was 35, fell into the 31-40 years group). The following message was given: *This message is targeted*

at people between the ages of [age group]. The gender parameter was a direct insert of the gender that the participant answered that they identified with the most (female, male, or diverse). The following message was provided: *This message is targeted at [gender] users*. For online behavior, such a possibility did not exist. Therefore, participants were shown the following message: *This message is targeted at you through your online behavior and interactions on Facebook*. The number of times that the participants hovered over the individual parameters was also measured in this study.

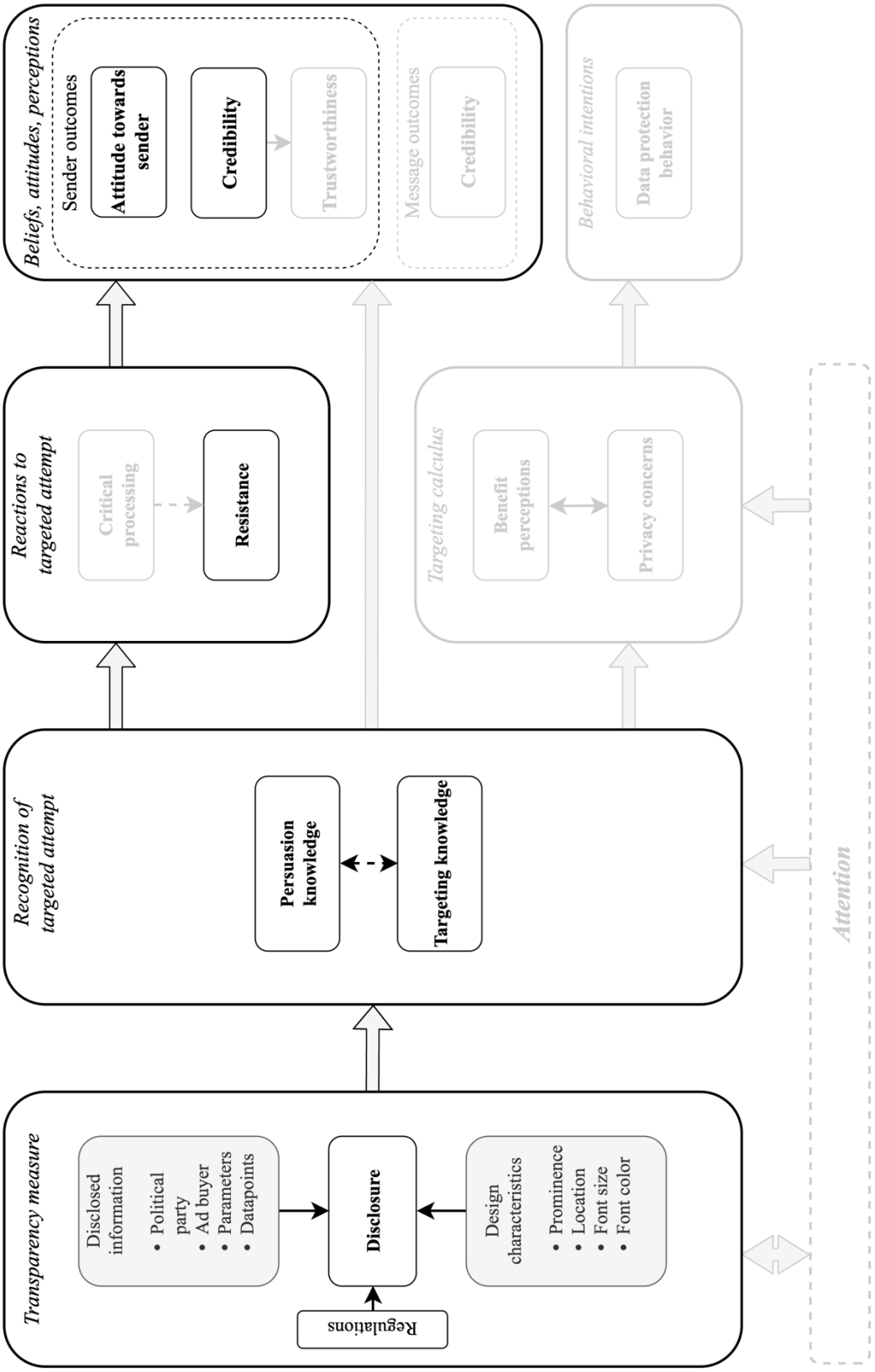
In the online experiment, 547 adult German Facebook users were recruited and randomly assigned to one of the experimental conditions. The study used a one-factorial (control vs. targeting disclosure) between-subjects design. Statistical analyses were performed using R (version 4.2.1; R Core Team, 2021) and jamovi (version 2.0.0.0; The jamovi project, 2022). Structural equation modeling with variable mean scores as observed variables was used to investigate differences in the groups and relationships between the constructs. The data showed a significant negative effect of the disclosure type on persuasion knowledge. However, this effect was below the set smallest effect size of interest (SESOI; Lakens et al., 2018), and was thus not interpretable in this study. The results showed a significant relationship between persuasion knowledge and cognitive resistance but not affective resistance. Subsequently, significant relationships were found between cognitive resistance, source credibility, and attitude towards the politician. Conversely, the results showed a significant relationship between persuasion knowledge and affective resistance; however, this effect was also below the SESOI threshold. Moreover, there was no relationship between affective resistance and source credibility or attitude towards the politician. Finally, the results showed a significant relationship between source credibility and attitude towards the politician.

As an exploratory analysis, this study investigated the recognition of disclosure by scrutinizing the differences in persuasion knowledge and the more recent construct targeting knowledge, which is an adaptation of persuasion knowledge adapted to the context of targeting. The results of these analyses showed that in the targeting condition, the mean for persuasion knowledge was lower than that in the control condition. In contrast, the mean for targeting knowledge was lower in the control condition but was also lower than the mean for persuasion knowledge overall. Another exploratory analysis in this study investigated the hovering that participants in the targeting conditions could perform. The results showed that within this condition, the possibility was used less than once per participant. Nevertheless, the people who used the possibility used it between two and three times per person.

In this study, not every participant recalled the correct disclosure. More precisely, only 17% of participants in the targeting condition recalled the correct disclosure, whereas 70% of the participants in the control condition recalled the correct disclosure. This confirms the results of Study I and other studies in which not all participants recalled the correct disclosure (Binder et al., 2022; Evans et al., 2017; Kruikemeier et al., 2016). While the sponsored disclosure in the control condition was designed to look the same as the one Facebook is currently using to improve ecological validity, it shows that for users, it is easier to recall disclosures with which they are more familiar. This means that a repeated exposure effect, increasing recognition and potential familiarity, may have occurred (Montoya et al., 2017). However, because a difference between the two conditions and targeting knowledge was found, the results still implied that disclosures could improve transparency regarding microtargeting; however, users might need to become aware of targeting processes as a whole before they can recognize these processes through a disclosure.

Moreover, the results of Study II showed that when users are aware that they are being persuaded, they are more likely to resist the message by counterarguing with it (Fransen, Smit, et al., 2015; Fransen, Verlegh, et al., 2015; Youn & Kim, 2019). However, according to the results, users may not affectively resist the message. One reason for this might be that users may not get angry because they are targeted. While counterarguing as a psychological reaction to a message is something users do multiple times per day, actually getting angry or annoyed might be a stronger reaction to an ad you receive on a platform where you consume content rather passively. In line with earlier research, the results imply that users who counterargue with a message perceive the source of that message to be less credible (Carr & Hayes, 2014; Deng et al., 2020; Wojdynski & Evans, 2016). Finally, the results showed that the perceived credibility of a source is positively related to users' attitude towards the politician. In agreement with other work, this implies that source credibility is an important construct to investigate, not only regarding voting behavior but also in the field of political advertising and communication (Funk, 1999; Madsen, 2019; Markus, 1982). Figure 3 depicts a schematic overview of this study.

Figure 3. Schematic overview of Study II



Article III: Balancing perceptions of targeting: An investigation of political microtargeting transparency through a calculus approach (Jansen & Krämer, 2023)

This study, in line with Studies I and II, investigates disclosures on SNSs. However, this study moved away from Facebook as a platform to Instagram. Although the platforms differ, the data points that can be used to target specific groups of users are the same. This is due to the fact that both platforms are owned by the same company, Meta, formerly known as Facebook. Moreover, little research has been conducted on microtargeting on Instagram, although this platform is growing at a faster rate than Facebook (Pew Research Center, 2021). Furthermore, the manner in which the platform is intended to be used is different. Most content on Facebook exists as a combination of text, images, or videos. On Instagram, the main focus of content is an image or video with a textual caption that is less prominent. Moreover, in line with Studies I and II, this study built on the persuasion knowledge framework but applied an adapted version of it in the concept of targeting knowledge, which was previously used in microtargeting research (Binder et al., 2022).

A considerable problem with PMT is that users are often unaware that they are being targeted, whereas existing transparency advances do not seem to suffice in informing users (Binder et al., 2022; Jansen & Krämer, 2023; Jost et al., 2022). However, increasing transparency may have consequences for users' privacy perceptions. Therefore, this study investigated targeting disclosures on Instagram using a calculus approach based on the *privacy calculus theory* by Culnan and Armstrong (1999). The privacy calculus theory states that individuals weigh privacy costs and benefits before they disclose personal information and are more likely to disclose this information if the benefits are at least balanced with, or greater than, the privacy costs (Dinev & Hart, 2006). Dienlin and Metzger (2016) showed that in the context of SNSs, an extended privacy calculus including users' self-withdrawal behaviors and privacy self-efficacy holds true. Other work has shown that, in the case of app usage concerning a COVID-19 warning app, both privacy concerns and perceived benefits predict less and more app usage (Meier, Meinert, et al., 2021).

Concerning microtargeting, Study III proposed that the calculated rational approach grounded in the privacy calculus could explain users' behavior regarding their online privacy perceptions. Users may experience privacy concerns because their personal and behavioral data are used while also having concerns regarding data protection (Binder et al., 2022; Dobber et al., 2019; Segijn & Van Ooijen, 2022). Conversely, users may experience higher levels of

usefulness by receiving political advertisements that match their personal preferences or interests (Barocas, 2012; Binder et al., 2022). These experiences and perceptions can coexist and lead to changes in attitudes towards microtargeting as a technique and users' intention to engage in privacy protection behavior. Engaging in this behavior could shield users against data usage by platforms, and advertisers engaging in less of this behavior might give platforms and advertisers more information and data to work with, potentially leading to better targeted ads and thus higher usefulness of those ads.

Although making users aware of the targeting practices occurring on platforms through disclosures might lead to changes in their perceptions regarding those practices, users' attitudes towards the platform potentially influence this as well. In an earlier study on Facebook, Debatin et al. (2009) found that while users did recognize the potential privacy issues of the platform, they simultaneously uploaded large amounts of personal information, and that this behavior may be explained by the high levels of gratification of using the platform. According to the mood congruency hypothesis, recipients' mood states may influence the associations generated during exposure to a message, leading to more positive elaboration of the content or more positive reactions to peripheral cues (Schwarz et al., 1991). In the domain of intrapersonal communication, the medium used to send a message has been found to positively affect a message's persuasiveness (Heim et al., 2002). Furthermore, regarding personalization and the privacy calculus, Hayes et al. (2021) found that the consumer–brand relationship has a positive moderating effect on the privacy calculus. In their study, they found that for users with a stronger consumer–brand relationship, the effect of perceived benefits on the value of information is larger than that for users with a weaker consumer–brand relationship. Study III proposed that the relationship between targeting knowledge and users' perceived benefits and privacy concerns would be moderated by their existing attitudes towards Instagram as a platform. Finally, this study investigated an exploratory outcome for users who experienced the benefits of personalization or targeting. According to prior research, users might not be as surrendered to algorithms as one might think. In an exploratory study, Kapsch (2022) found that some users influence what content they see by interacting with profiles, liking posts, commenting, or even texting via direct messages (DMs). Users who exercise these techniques try to gain autonomy by actively showing the algorithm of a platform what they like, which is potentially a proxy for their willingness to actively use algorithms to their advantage. Study III investigated this construct as *algorithmic user agency*, a construct that could be positively related to users' perceived benefits.

In this study, a simulation of microtargeting was used in the same manner as in Study I, in which participants received statements on the stimulus material that were more likely to be in line with their own beliefs after they answered a battery of statements regarding their views on several topics but only measured their beliefs regarding climate change regulations (see Article I). In the online experiment, 450 adult German Instagram users were recruited and randomly assigned to one of the experimental conditions. The study used a one-factorial (control vs. targeting disclosure) between-subjects design. In line with Study II, statistical analyses were performed using R (version 4.2.1; R Core Team, 2021) and jamovi (version 2.0.0.0; The jamovi project, 2022), and differences in the groups as well as relationships between the constructs were investigated through structural equation modeling with the variable mean scores as observed variables. Participants in both conditions were exposed to an Instagram advertisement by a fake political party. In the control condition, the ad contained a sponsored disclosure in line with the disclosures Instagram currently uses, whereas in the experimental condition, a more salient targeting disclosure was used. This disclosure was based on the false information disclosures that Instagram/Meta used during the COVID-19 pandemic. However, this disclosure was adapted to contain information about the post being targeted ("This sponsored message is targeted at you based on your age, gender, and online behavior").

The results showed that users who were exposed to a targeting disclosure had higher levels of targeting knowledge than those who were exposed to a sponsored disclosure. Moreover, targeting knowledge was positively related to perceived benefits but not to privacy concerns, meaning that users perceived the benefits of microtargeting but did not experience the privacy concerns that could arise with the implementation of the technique. Furthermore, these relationships were not moderated by users' attitudes towards Instagram as a platform. In addition, the results showed no relationship between benefit perceptions and the intention to engage in privacy protection behavior. However, the results showed a relationship between privacy concerns and users' intention to engage in privacy protection behavior. Finally, the results showed a positive correlation between perceived benefits and the exploratory construct of algorithmic user agency.

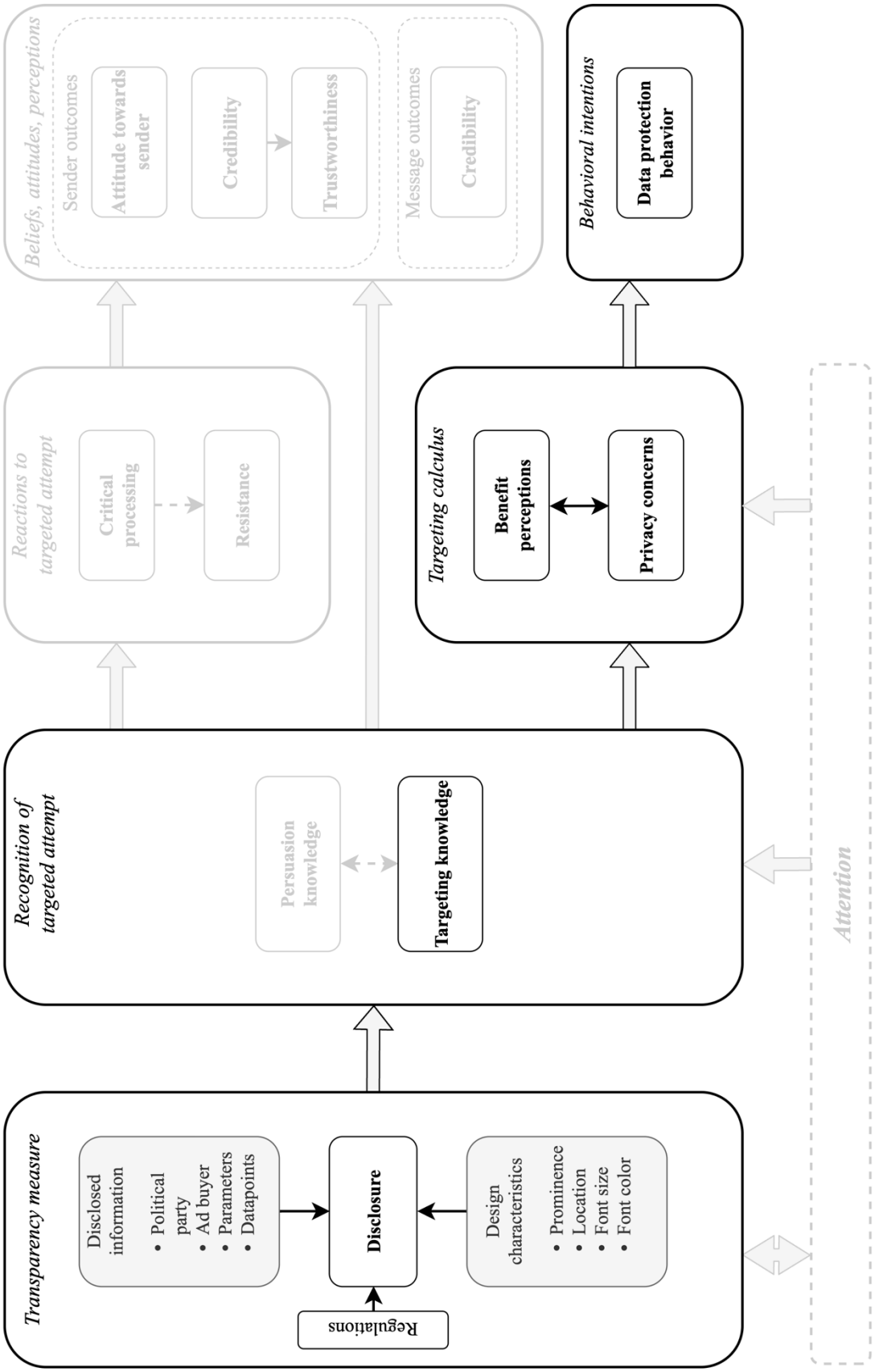
In Studies I and II, a large similarity was that not all participants—specifically in the experimental conditions with a newly developed targeting disclosure—recalled the correct disclosure. However, Study III showed different results regarding the manipulation check. That is, 84% of participants in the experimental condition recalled the correct disclosure, whereas only 32% of participants in the control condition did so. This finding not only contrasts Studies I and II but also prior research that found that recall regarding disclosures is usually lower than

that found in Study III (Binder et al., 2022; Jansen & Krämer, 2023; Kruikemeier et al., 2016). To an extent, the recall of the targeting disclosure might also be the reason why the results showed a positive effect of our manipulation on targeting knowledge. Moreover, the results confirmed that when users are aware of the advertisement being targeted at them, they have higher levels of perceived benefits, which is in line with earlier work that found more positive evaluations of content when it was personalized (Brinson & Eastin, 2016; Tam & Ho, 2006). A potential mechanism behind this finding could be that the fluency of the personalized content makes users disregard the disclosure as a warning. However, another explanation could be that users, in general, are more used to receiving personalized content and advertisements and that they mainly perceive the benefits of tailoring as a technique. This would also explain another result of this study: the lack of a relationship between targeting knowledge and privacy concerns. While existing research shows that if people know that they are exposed to personalized messages, this affects privacy risk perceptions (Jain & Purohit, 2022) or that users perceive these messages as intrusive or creepy (Segijn et al., 2022; Segijn & Van Ooijen, 2022), the current study was unable to confirm this. However, recent work on microtargeting similarly did not find a relationship between exposure to transparency information and users' privacy concerns (Dobber, Kruikemeier, Helberger, et al., 2023). Furthermore, it is possible that users had a higher level of privacy concerns regarding PMT, as the mean score for this construct was relatively high. This could mean that a certain privacy threat awareness already existed, which would mean that explicitly informing users about the targeting would not make a difference. In addition, a ceiling effect might have been active, which would limit the statistical increase or decrease in the relationship strength between targeting knowledge and privacy concerns, as the default for all participants might already be high.

Furthermore, this study was unable to detect a relationship between perceived benefits and users' intention to engage in privacy protection behavior, which conforms to work on personalization (Van Ooijen, 2022) and SNSs (Meier et al., 2020). Nevertheless, the results did show a positive relationship between users' privacy concerns and their intention to engage in privacy protection behavior, which confirms earlier findings regarding social media (Dienlin & Metzger, 2016; Meier, Schäwel, et al., 2021) and PMT (Binder et al., 2022; Stubenvoll et al., 2022). The main rationale for these two predictions was the 'calculation' in the privacy calculus theory (Culnan & Armstrong, 1999), which states that users weigh perceived benefits and privacy risks to (dis-) engage in privacy protection behavior. However, it is possible that benefits, risks, and privacy protection behavior are independent constructs and that the

intention to engage in privacy protection behavior might not be the outcome of the proposed calculation. Figure 4 depicts a schematic overview of this study.

Figure 4. Schematic overview of Study III



Article IV: For your eyes only? An eye-tracking experiment investigating microtargeting transparency, visual attention, and critical processing (Jansen & Van Ooijen, 2023)

A substantial problem with PMT is that receivers of targeted ads are not aware that these ads are targeted at them. Being aware that they are targeted is a prerequisite to understanding that their personal data are used in employing the technique. Disclosures can contribute to transparency and help users to become aware that they are being targeted. However, this improvement in transparency potentially leads to users processing the information in the ad more critically. Moreover, the use of disclosures could also lead to differences in users' attention to the advertisement itself.

One of the main implications of current disclosure research is that the prominence, positioning, and degree to which disclosures stand out, in contrast to regular content on a timeline, are important (Amazeen & Wojdynski, 2020; Wojdynski & Evans, 2016). Moreover, a combination of text and a logo or symbol (compared to just text or just a logo) leads to the most visual fixation (Boerman et al., 2015). When disclosures are more prominent, they are more likely to stand out from the timeline than regular content. This leads to higher levels of attention and clearer perceptions of the information contained in the disclosure, thereby contributing to transparency (Binder et al., 2022; Kruikemeier et al., 2016). Current disclosures, mostly designed in gray and thin font sentences (i.e., lacking salience), seem to be insufficient in informing users, and thus in increasing transparency (Binford et al., 2021; Dobber, Kruikemeier, Helberger, et al., 2023; Jansen & Krämer, 2023; Jost et al., 2022). However, since 2020, a differently designed disclosure has been used by platforms. During the COVID-19 pandemic, Meta implemented measures to counter the spread of misinformation on its platforms. One of these measures was a disclosure on posts and stories (i.e., disappearing posts) that informed users that the information in that post could be false or not yet proven. The disclosure was located below the content and contained informational text and a large "I" symbol to indicate that it contained information. The design choices made regarding the salience and prominence of these disclosures might be a solution for platforms to adhere to DSA regulations.

When users are made aware of a persuasive attempt, which is supposed to be one of the primary results of using disclosures, they must determine how to effectively manage their responses to this persuasive attempt (Friestad & Wright, 1994). Their awareness of persuasive attempts encourages more systematic and biased processing of advertisements (Janssen et al.,

2010). This method of processing a message can be considered critical processing, which is defined as the adoption of an evaluative style of processing in which the content is criticized (Boerman et al., 2014). This style of processing has indeed been found to be elicited through the use of sponsorship disclosures (compared to no disclosure) in the case of sponsored content on television shows (Boerman et al., 2014). In the case of targeted ads, Zarouali et al. (2017) found that adolescents engage in more critical reflection of targeted ads if they are exposed to a cue that triggers their critical elaboration. Other work confirms this and shows that if adolescents are exposed to a disclosure in a timeline that explains the targeted nature of a Facebook ad, they engage in significantly more critical processing compared to a dropdown menu containing that information or an ad without any transparency measure (Zarouali et al., 2018). In addition, awareness of a persuasive attempt, in general, has been shown to lead to less favorable perceptions of the communication attempt (Allyn & Festinger, 1961; Boerman et al., 2014; Campbell & Kirmani, 2000).

Disclosures can have different consequences than just informing and shielding users from the persuasive effects of advertisements. They can also unintentionally increase, rather than decrease, persuasive effects (Janssen et al., 2016). The main reason for this is that when users allocate more attention to the disclosure and read the disclosure message, they might be curious about what a political party or advertiser thinks they know about them and, subsequently, with what message they aim to persuade them. If this means that attention to a disclosure would lead to more attention to the ad—assuming the ad stays as persuasive as it is intended to be—this might indeed be a negative unintended effect. Other work shows that viewers' recognition of an ad increases their brand memory, which could be seen as one of the aims of a persuasive message (Boerman et al., 2015). In a study on disclosures, Evans et al. (2019) found that sponsorship transparency can have a positive effect on receivers' attitudes towards the ad and their purchase intention. Pan and Zinkhan (2006) found that a privacy disclosure message in a web shop environment leads to greater trust in the retailer. In addition, when platforms and advertisers are transparent about personalization practices, the perception of vulnerability decreases, which results in consumers being more inclined to engage with the ad (Aguirre et al., 2015).

In a one-factorial (control vs. targeting disclosure) between-subjects laboratory eye-tracking experiment, 134 Instagram users were exposed to an Instagram timeline containing five posts and one political ad containing either a sponsored disclosure in the control condition or a more informative disclosure that stated that the advertisement was targeted at them in the targeting condition. To ensure that most participants could have been targeted by the

advertisement, this study used the same political advertisement from two Dutch parties, one center-left and one center-right (i.e., the Labour Party (Partij van de Arbeid) and Democrats 66). Since it was predicted that most participants would be students, the advertisement displayed a statement regarding the housing market in the Netherlands and stated "Housing is a fundamental right, not something that should be speculated with". Randomization checks showed no differences regarding which political party respondents were exposed to between the experimental conditions. Statistical analyses were performed using R (version 4.2.1.; R Core Team, 2021) and jamovi (version 2.3; The jamovi project, 2022). Differences between the experimental groups and relationships between the constructs and mediation analysis (5,000 bootstrapped samples) were investigated using structural equation modeling, with the variable mean scores as observed variables.

With regard to the manipulation check, participants were asked which disclosure label they recalled. In the control condition, most participants correctly recalled the sponsored disclosure; in the targeting condition, an equal number of participants recalled the targeting or sponsored disclosure, while fewer participants recalled no disclosure, which in turn was lower than the number of participants who did not recall a disclosure in the control condition. Interestingly, since this study employed eye-tracking, a comparison could be made between participants' recall of the different disclosures and their visual attention towards the disclosures. This comparison shows that, across the groups, participants who did not recall a disclosure had the lowest visual attention to the disclosure, followed by those who recalled a sponsored disclosure. Finally, participants who recalled a targeting disclosure were fixated on the disclosure the longest.

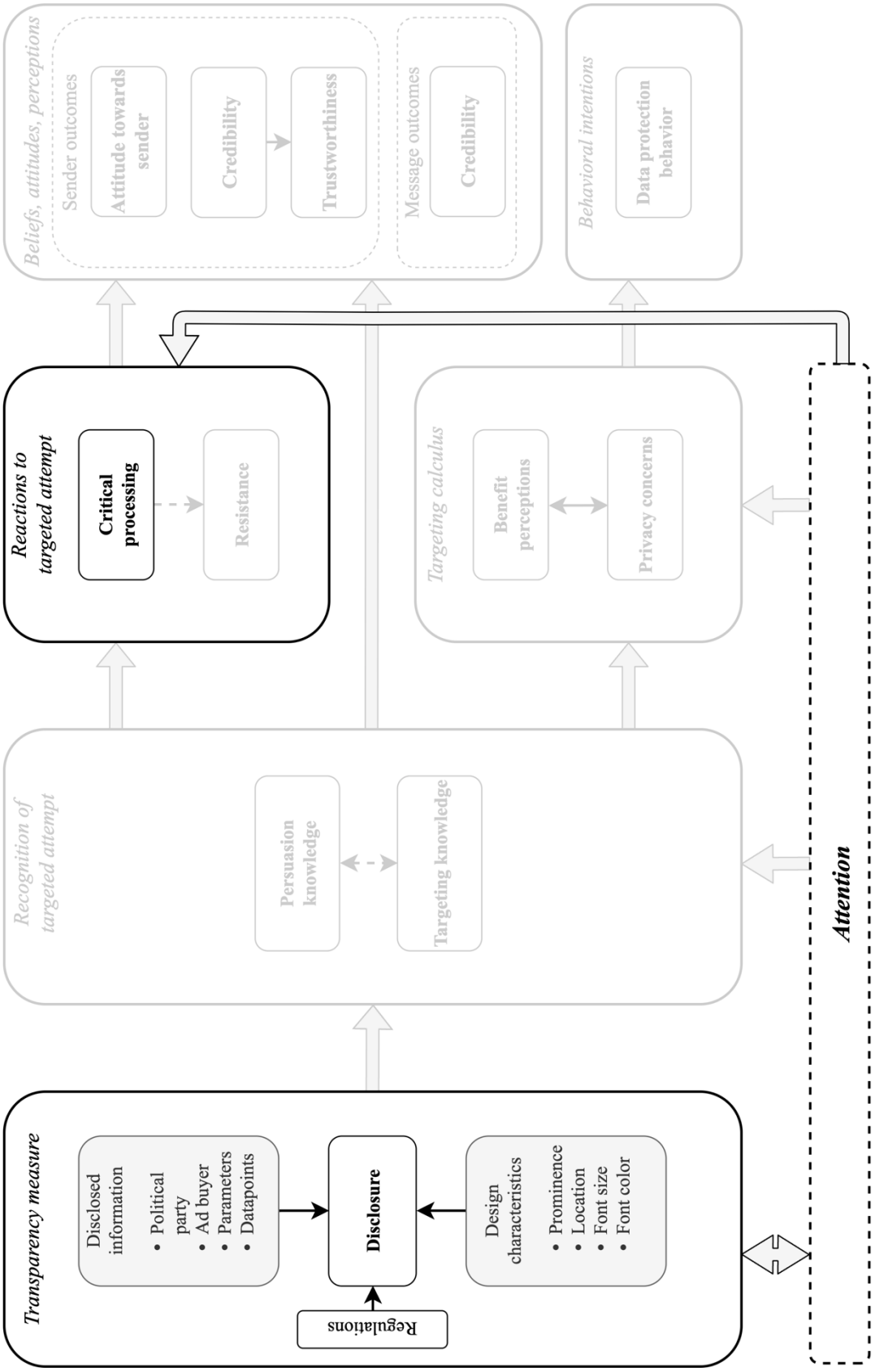
The results showed that if an ad contained a targeting disclosure, it significantly affected participants' visual attention to the disclosure. However, this effect size was below the set threshold for the SESOI and thus could not be interpreted meaningfully. Furthermore, as predicted, the results showed that participants' visual attention to the disclosure was positively related to their critical processing of the message, as well as their visual attention towards the advertisement. Nevertheless, no mediation of visual attention to the advertisement on the effect of visual attention to the disclosure on critical processing was found.

In line with existing work (Binford et al., 2021; Boerman et al., 2015), this study expected that exposure to a targeting disclosure would lead to more attention towards the disclosure itself. The results were in favor of this, but the effect size was too small to be meaningfully interpreted. However, because this was a laboratory experiment with a relatively high number of participants, an experiment in an online setting, or at least on a larger scale,

may be able to replicate these effects while interpreting smaller effect sizes. In line with this result, not all participants recalled the correct disclosure. While this agrees with existing work regarding targeting disclosures (Binford et al., 2021; Jost et al., 2022; Kruikemeier et al., 2016), the finding that participants who recalled the targeting disclosure had more than one-and-a-half times the visual attention compared to those who recalled the sponsored disclosure is a step forward compared to the existing body of work on targeting disclosures. In addition, this result, even though it was not the main research question of this study, could imply that the implementation of the adapted version of the COVID-19 disclosures could be effective, even if it would be on a speculative basis. Furthermore, in line with other studies (Boerman et al., 2014; Zarouali et al., 2017), this study showed that users' visual attention to the disclosure was positively related to their critical processing of the advertised message. While the results show no meaningful increase in visual attention to the disclosure due to the different disclosures, the fact that their users' critical processing is positively related to the visual attention might indicate that the disclosure affects the subsequent processing of the advertisement. While the disclosure could have been effective in helping users distinguish the ad from regular content, it could have helped them become aware of the persuasive attempt (Friestad & Wright, 1994) and subsequently manage their response by systematically and critically evaluating the advertisement (Boerman et al., 2014; Janssen et al., 2010). However, the low and non-significant correlation between the conditions and critical processing shows that this could only be speculated about in this study.

Moreover, this study found that one of the unintentional effects of disclosures occurred in this experiment, where users' visual attention to the disclosure was positively related to their visual attention to the ad itself. This effect may mean that in addition to protecting users from the persuasive effects of an ad, disclosures can also increase rather than decrease these effects (Janssen et al., 2016). Visual attention toward the advertisement might be seen as a proxy for attitude towards the advertisement, which has been found to increase through sponsorship transparency measures (Evans et al., 2019). However, this could also mean that users become interested in what advertisers think they know about them, which might lead to them being more curious and having a surveillance motive (Choi, 2016; Flavián & Gurrea, 2009). Figure 5 depicts a schematic overview of this study.

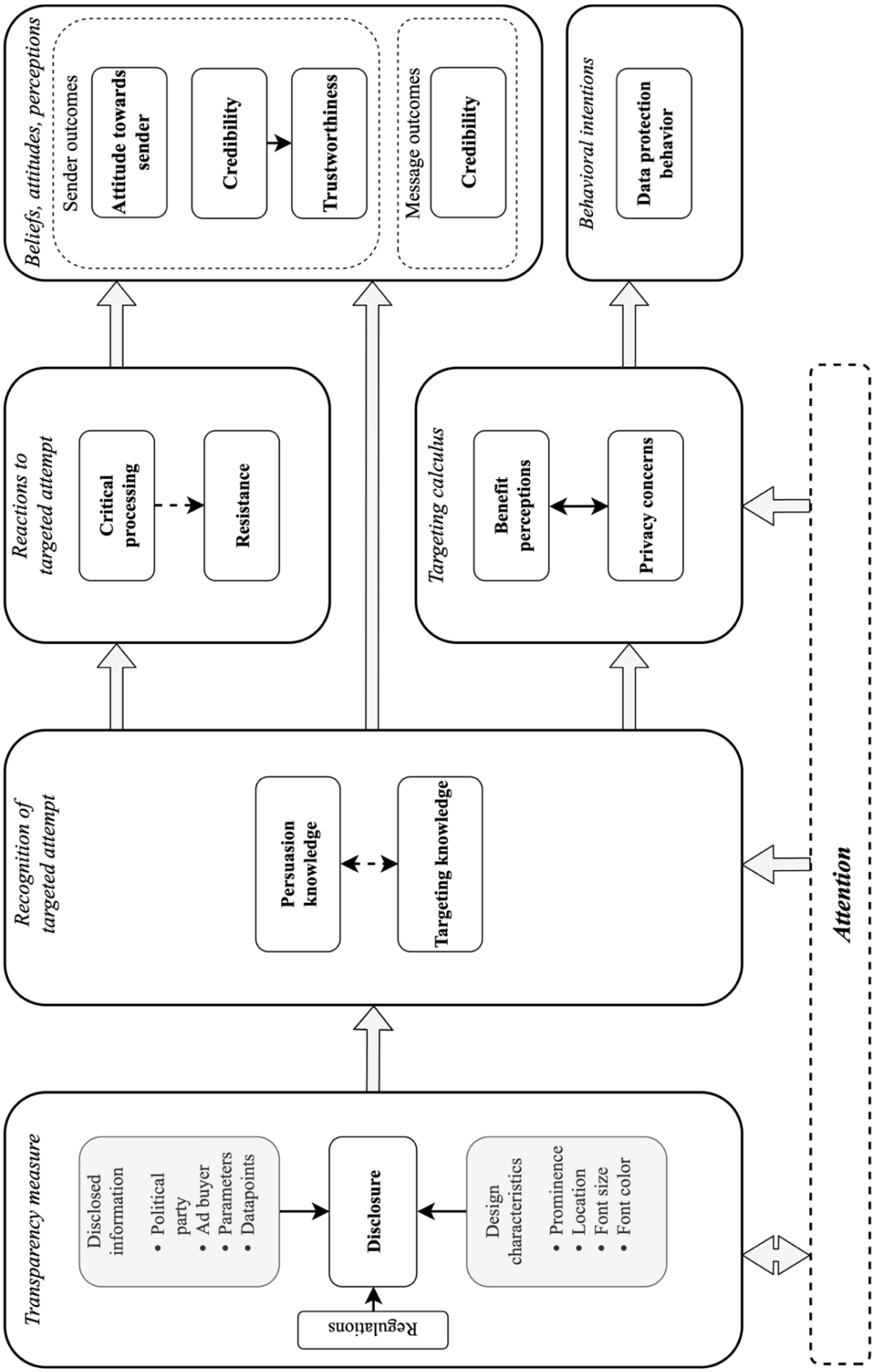
Figure 5. Schematic overview of Study IV



IV General Discussion

The studies presented in this dissertation investigated the effects of targeting disclosures and the psychological processes that occur when receivers are exposed to these disclosures. This section is structured according to the four aspects of this dissertation. Accordingly, the four studies in this dissertation are discussed alongside these four aspects. A graphical overview of these aspects and their relationships is shown in the conceptual model in Figure 6. Before discussing different aspects, the paragraph on disclosure characteristics discusses the characteristics that increase users' recall of disclosures and, thereby, their potential effectiveness. The following paragraph describes the first aspect of this dissertation by discussing how receivers recognize targeted attempts. The second aspect discusses the psychological reactions that receivers can have to the targeted attempt if they are aware of the message being a targeted advertisement. The third aspect discusses how receivers form new beliefs, attitudes, and perceptions after becoming aware of the targeted nature of microtargeted advertisements. Finally, the fourth aspect in this dissertation discusses the targeting calculus based on the rationale described in the privacy calculus theory (Culnan & Armstrong, 1999). This rationale describes how receivers weigh their benefits and privacy concerns in order to engage in different levels of privacy protection behavior. The following section provides a broader overview of the four studies discussed in this dissertation and compares and contextualizes their implications.

Figure 6. Conceptual model of the four aspects in this dissertation



8. Overview of the findings

8.1 Disclosure characteristics

Based on existing work in the fields of persuasive and political communication, all the studies in this dissertation investigated targeting disclosures. These disclosures were based on those used in persuasive communication, such as influencer marketing, advertorials, and search engine results. Existing research shows diverse results regarding the effectiveness of disclosures, as did the studies in this dissertation. Regarding the disclosures used in these studies, it is important to recognize that both Facebook and Instagram posts were used as stimulus materials. The disclosures used in the studies were either in line or inspired by disclosures on those platforms. Study I compared an organic Facebook post with the same post displayed as an advertisement. Subsequently, the advertisement contained either sponsorship disclosure or a targeting disclosure. The sponsored disclosure replicated the disclosures that Facebook currently uses; the targeting disclosure was designed in the same way but disclosed different information (i.e., the fact that the message was targeted at the receiver). The findings of this study showed that most participants did not recall either of these disclosures correctly. In line with Study I, Study II built on the disclosures that Facebook currently uses, but compared the currently used sponsored disclosure with a more prominent targeting disclosure that contained red and underlined text, exposing that receivers could hover over the terms to see the data used to fill in those parameters. Prominence has been found to increase participants' recall and, subsequently, the effectiveness of disclosures (Amazeen & Wojdynski, 2020; Wojdynski & Evans, 2016). In addition, the color red might attract more visual attention (Baik et al., 2013). Study II found higher levels of recall for the sponsored disclosure across the entire sample. However, recall in the targeting condition was higher than in the first study.

Regarding the stimulus material, the third study switched from Facebook to Instagram. Accordingly, the disclosures used in this study were two that were used on Instagram. This study compared Instagram's current sponsorship disclosure, a thin font on the top left of the post, to the disclosure that the platform used to inform users of potential misinformation during the COVID-19 pandemic. This disclosure is larger and more prominent than Instagram's currently used sponsorship disclosure. In addition, the disclosure contained both text and an icon to inform users that more information was available. Both higher prominence and the combination of text and an icon have been found to improve the effectiveness of disclosures (Beckert & Koch, 2022; Boerman et al., 2015). In Study III, the majority of participants who were shown a targeting disclosure recalled the correct disclosure, indicating that increased

prominence and the combination of text and an icon were effective in improving the recall of the disclosure. Study IV compared the same disclosure as the one used in the experimental condition in Study III but differentiated between a text in the disclosure that stated that the message was sponsored or a text stating that the sponsored message was targeted at the receiver. Most participants in the targeting disclosure condition recalled either a sponsored or targeting disclosure (compared to no disclosure). However, because the experiment used eye tracking, it was shown that participants who did not recall any disclosure spent the least number of seconds looking at the disclosure, followed by those who recalled a sponsored disclosure. Finally, the participants across both conditions who recalled a targeting disclosure (disregarding whether this was the disclosure they were shown) spent the most time visually fixating on the disclosure.

The results regarding the recall of disclosures in the studies within this dissertation indicate that increasing prominence is an effective way to improve receivers' recall of disclosures and, in turn, the effectiveness of the disclosure. In addition, the combination of text and an icon was found to be a factor in improving the recall of disclosures. Moreover, the studies in this dissertation relied not only on self-reported measures but also scrutinized users' behavior by employing eye-tracking in Study IV. The findings show that participants' self-reports do not always align with their gazing behavior, as has been found in other research (Vraga et al., 2016). If users report that they recall the incorrect disclosure, especially in an experimental study, this might simply mean that they do not recall the correct disclosure, not that they did not see it, or looked at it longer. In line with this, future research would benefit from a combination of self-reported measures and behavioral data.

For most users, the goal of using social media will not be to look at disclosures; their attention will initially not be on the disclosures, but on the content they would like to consume. It remains important to be transparent about targeting practices occurring on social media, where advertising might be more covert than in other types of media (Wojdynski & Evans, 2020). However, the diverse findings in existing research might be explained by users focusing on the content first and foremost, and subsequently possibly focusing on disclosures. Conversely, existing research, including the studies in this dissertation, may overestimate the effectiveness of disclosures. While Study IV used eye tracking and showed participants a mock-up timeline containing six Instagram posts, other studies and most existing research show participants short bits of content or just one social media post containing a disclosure. This method of exposing participants to a disclosure may overestimate the attention that users allocate to the disclosures used in a study because there is no social media platform that can be scrolled endlessly, which is often the case in the 'real world'. In line with this, another reason

that users might not allocate attention to, or even recognize, disclosures might be *banner blindness* (Boerman & Kruikemeier, 2016; Hervet et al., 2011). This concept explains behavior in which users consciously seem to avoid looking at banners, as most websites at the beginning of the Internet carried banners as their prime means of advertising (Benway, 1998). Banner blindness explains that users recognize advertisements compared to regular content on websites, and therefore do not look at the advertisements' content further. If users still employ this behavior, but nowadays do so on social media, this could explain why they do not see disclosures to begin with. If users would recognize an advertisement because they, for instance, have seen it in the past or recognize the sender as a political party they do not follow, and therefore decide not to look at it any further, there is a chance that users will not look at the disclosures that are placed on the advertisement. Nevertheless, disclosures regarding targeted messages might do more than simply inform people that an advertisement is sponsored or targeted. The targeting disclosures described in the DSA will provide users with information about who is targeting them, with which information or data, and why *they* are targeted specifically (European Commission, 2022). Moreover, most results show that at least one disclosure is recalled, even if it is not always the correct one. It is possible that after an adaptation period, disclosures will be more effective in informing receivers of targeted messages, as repeated exposure might lead to a familiarity effect, helping receivers recognize disclosures and what information could be available in them (Montoya et al., 2017).

There might be another reason why research on disclosures has found diverse results. While this would be on a speculative basis, it might be that research stays too close to the disclosures that platforms use, and therefore, does not always find results that can confirm the effectiveness of these disclosures. It is reasonable that research aims to achieve high ecological validity, but at the same time, it should be noted that these disclosures have not always succeeded in informing users; thus, researchers have a greater chance of not finding effects. However, it is vital to note that there is a reason why researchers may aim for ecological validity regarding disclosures. It is possible to make disclosures in experimental studies as prominent as they can be, for instance, by making them pink, with huge borders and bright letters, and then find effects in those studies regarding recognition, as people will notice those disclosures. However, a certain ecological validity should be present in studies, as the disclosures should be developed in a way that fits for platforms as well. If disclosures do not fit, there is a chance that platforms simply do not allow political targeting anymore, as was the case with X (previously Twitter) in the past.

In conclusion, while earlier research shows that users do not always recall or recognize targeting disclosures (Binder et al., 2022; Jost et al., 2022; Kruikemeier et al., 2016), the studies in this dissertation show that certain alterations to these disclosures can positively affect their effectiveness. The largest implication is that the increased prominence of disclosures compared to the less prominent ones that are currently used has the potential to increase users' recognition and recall of these labels (Evans et al., 2019; Jost et al., 2022; Van Reijmersdal et al., 2012). For instance, the prominence of disclosures can be improved by using text colors that do not match the platform's corporate design or attract more attention, such as red or other warm colors, in the case of Facebook's blue design. Moreover, prominence can be improved by using less thin light gray text and more bold text that is easier to visually recognize. Finally, using both text and a larger icon or symbol in the disclosure, for instance an 'I' to show that additional information can be found in the disclosure, can help attract users' visual attention to the disclosure, which is important as attention to the disclosure can be seen as a prerequisite for its effectiveness.

8.2 Recognition of the targeted attempt

While the previous paragraph discussed disclosures and their characteristics, the following paragraph discusses the recognition of the targeted attempt. Prior research shows that using disclosures on persuasive messages can lead to the activation of persuasion and targeting knowledge (Binder et al., 2022; Boerman et al., 2012; Kruikemeier et al., 2016). However, these types of knowledge about (targeted) advertising are not only activated by disclosures, but can also be activated by users' mere recognition of the content being an advertisement. Therefore, from the first aspect of this dissertation, users' recognition of the targeted attempt was discussed. Throughout the four studies, the concept of recognition was investigated using three different constructs. Based on prior research, every study predicted that a targeting disclosure, which was more prominent than a sponsored disclosure, would lead to higher levels of recognition of the targeted persuasive attempt. Studies I and II investigated users' recognition through persuasion knowledge, and the third study focused on the adaptation of that construct through targeting knowledge. Study IV used a different approach because eye-tracking was utilized in this experiment, which allowed for the registration of users' behavior in a laboratory setting. In that study, the participants' visual attention to disclosure in seconds was measured.

In the first study, no effects of the different types of disclosures on persuasion knowledge were found. In this study, the targeting disclosure was strongly based on the

sponsored disclosure that Facebook currently uses. The only difference was that the text did not state that the advertisement was sponsored and who paid for it, but that the content was targeted at users through their online behavior. While it was never the sole goal of this study, the largest implication was that the disclosures that Meta currently uses on Facebook are not sufficiently prominent and, thus, are not noticed by users. In Study II, the more prominent disclosure containing underlined red text did not lead to higher levels of persuasion knowledge. Conversely, the more prominent disclosure led to lower levels of persuasion knowledge, but the effect size was too small to be meaningfully interpreted. Reflecting on this, it became clear that a reason for this could be that in the control condition, participants were shown the disclosure that Meta now uses on Facebook, stating that content is sponsored. However, in the experimental condition, participants were shown a disclosure stating only that the message was targeted at them and based on which data this was. The participants were never shown information that the content was sponsored and targeted at the same time. In this study, targeting knowledge was investigated as an exploratory construct and compared between the two experimental groups. The results showed that, for the experimental group compared to the control group, persuasion knowledge was lower, on average, while targeting knowledge was higher. The third study in this dissertation focused on targeting knowledge as a measure of users' recognition of a targeted attempt. The results confirmed that a more salient and larger disclosure stating that the information was sponsored and targeted at the user leads to higher levels of targeting knowledge. In the fourth study, the targeting disclosure led to higher levels of visual attention to the disclosure compared to a sponsored disclosure that was designed in the same way, but the effect size was just below the set threshold to be meaningfully interpreted.

In summary, the studies discussed in this dissertation confirm the diversity of results found in existing research. Some studies have found that disclosures are effective in enabling users to recognize persuasive and targeted attempts (Binder et al., 2022; Binford et al., 2021; Jost et al., 2022). However, most studies that do find this result have disclosures that are more prominent than those used in other studies. If disclosures are more prominent, users will begin to see them more, which might lead to a familiarity effect (Montoya et al., 2017). This effect could eventually enable users to better recognize disclosures, which agrees with the findings of the studies in this dissertation, where the sponsored disclosure that platforms currently use is often recalled better than newly developed targeting disclosures. Subsequently, disclosures can help users better recognize persuasive and targeted attempts by incorporating this information into their persuasion (or targeting) knowledge. As Friestad and Wright (1994)

explained, persuasion knowledge is a construct that is constantly updated through the incorporation of new information regarding new advertising techniques. If users' persuasion knowledge is updated regarding the techniques, it will consequently help them recognize persuasive targeted attempts in the future. However, this does not mean that the disclosures are obsolete. As marketing techniques evolve, they sometimes become more covert, making it more difficult for users to recognize them (Wojdynski & Evans, 2020). If this is the case, disclosures can still enable users to recognize persuasive attempts, even if they are already knowledgeable regarding that type of advertising or specific technique. In addition, the results of Study II suggest that, in line with other research (Binder et al., 2022; Dobber, Kruikemeier, Helberger, et al., 2023), it might be easier for receivers to activate persuasion knowledge than specific targeting knowledge. As argued by Dobber et al. (2023), this might be because microtargeting is an ill-understood phenomenon (Binder et al., 2022; Dobber et al., 2019). Therefore, specific targeting knowledge might be more challenging to activate than regular persuasion knowledge, and future research should investigate whether receivers learn something from the disclosures they are exposed to compared to solely investigating whether receivers recall a certain disclosure.

In conclusion, although not every user has the same level of knowledge about persuasive and targeted attempts, disclosures still enable users to recognize the intentions of the message's source through persuasion or targeting knowledge activation, disregarding their pre-existing knowledge. Thereafter, these users would be able to incorporate the experience with that particular persuasive episode into their persuasion or targeting knowledge to recognize attempts in the future. Finally, even individuals who already have higher levels of knowledge might still benefit from disclosures to activate their persuasion or targeting knowledge, and thereby recognize persuasive attempts more easily.

8.3 Reactions to the targeted attempt

The second aspect of this dissertation concerns users' reactions to targeted attempts. These reactions are processes that occur after users are made aware that the content is persuasive and targeted at them through persuasion or targeting knowledge. These reactions are not reactions regarding the source of the message or users' attitudes but rather processes that users can employ to cope with the targeted attempt. In Study II, resistance as a psychological mechanism was investigated. This study distinguished between cognitive and affective resistance and thus developed two 'routes' (Van Reijmersdal et al., 2016). Each route led to a direct effect of each type of resistance on source credibility and users' attitude towards

the politician. As predicted, persuasion knowledge was positively related to both types of resistance. However, the strength of the relationship between persuasion knowledge and affective resistance was below the threshold set for meaningful interpretation. Interestingly, all other predicted relationships on the affective resistance route were non-significant and below the threshold for meaningful interpretation. Study IV investigated users' critical processing of a targeted message. The study showed that, as predicted, the visual attention that users had for the disclosure would positively relate to more visual attention to the advertisement itself, as well as to higher levels of critical processing. Finally, no predicted mediation of visual attention to the advertisement on the relationship between attention to the disclosure and critical processing was found.

The operationalizations of resistance in the second study were counterarguing for cognitive resistance and negative affect for affective resistance. As stated above, only the relationships via the cognitive route were significant and interpretable. While the study found no relationships for affective resistance, it did correlate with cognitive resistance, and thus, might be part of other psychological processes. This agrees with research that has found that receivers react differently to stimuli, depending on which dimension of resistance is activated (Di Plinio et al., 2022). The relationship between persuasion knowledge and cognitive resistance indicates that participants resisted the message when they were aware that they were being persuaded by actively counterarguing with it (Fransen, Smit, et al., 2015; Fransen, Verlegh, et al., 2015; Youn & Kim, 2019). Thus, individuals who are aware that a political message is an advertisement could have more doubts about its content and reject it rather than engage with it. An example of this is a message that a user can perceive as dishonest and only aimed at attracting voters or being a sales pitch for a candidate. Moreover, political communication can be seen as a high-involvement process, which may be another reason why users counterargue more with the content. An explanation for this might be that political communication contains information that demands more cognitive resources to be processed, and users may experience more involvement with it, which may result in more resistance to the message than a commercial that advertises a (lower-involvement) product (Van Reijmersdal et al., 2016).

The finding in Study IV that attention to a disclosure is positively related to critical processing shows that while there was no effect of the specific targeting disclosure on the attention to the disclosure, there might still be a subsequent effect of the disclosures. The disclosures could have been effective in helping users distinguish the ad from regular content in the timeline, which could have helped them become aware of the persuasive attempt

(Friestad & Wright, 1994). Consequently, they would have been able to manage their responses by systematically and critically evaluating the advertisement (Boerman et al., 2014; Janssen et al., 2010). While Study IV investigated users' critical and skeptical responses to advertisements, it is also possible that users are skeptical at another level. This study shows that users' visual attention to the disclosure is positively related to their visual attention to the advertisement. A potential reason for this is that they might be skeptical about what the platform thinks it knows about them as individuals and why they receive a certain advertisement. This can be explained as a more investigative surveillance motive, where they are curious about the advertisement and the targeting taking place (Choi, 2016; Flavián & Gurrea, 2009).

Additionally, the politician and party that were advertised in Study II were nonexistent. While this was done to eliminate potential confounding from existing attitudes and ensure that the effects in the study would be related to the manipulation, it might also have over- and underestimated participants' levels of resistance. If participants in the study allocated more resources to the image in the posts, it is possible that they did not allocate as much resources to the statement in the post as they would in a real-world situation, leading to them counterarguing with it less. In contrast, it is also possible that participants counterargued with the statement more because they allocated more cognitive resources to recognize that the party and politician were something they had not seen before. This could have led to more systematic processing, and thus the activation of a state in which all information was processed more systematically and critically. This might have led them to think a lot about the stimulus material in general, disregarding the potential activation of persuasion knowledge and, in turn, critical processing or resistance. Moreover, in both Study II and Study IV, regardless of whether the party was an existing party, it could also be that the participants did not feel as if the party or subject of the message was congruent with their political attitudes, leading to more critical processing of the message (Binder et al., 2022).

In conclusion, exposure to a persuasive and targeted attempt that includes a disclosure might lead to higher levels of recognition and, in turn, users may employ resistance strategies to cope with these attempts. Furthermore, recognition of a targeted advertisement might lead users to process that message in a more critical way, potentially leading to the activation of resistance. In this dissertation, the results only support the activation of cognitive resistance; it is important to acknowledge that affective resistance could be activated as it might be part of other psychological reactions to targeted advertisements.

8.4 Beliefs, attitudes, and perceptions

The third aspect of this dissertation discusses users' beliefs, attitudes, and perceptions of the sender and their message. Recognition of and reactions to microtargeted political advertisements can activate different psychological mechanisms for users. The mere recognition of microtargeting directly, but also the psychological mechanisms that can be activated as a response, may influence the way users observe the source and the message. The effects that recognition and reactions have on the way users perceive the sender and message are discussed below.

8.4.1 Sender-related outcomes

If political parties or actors send microtargeted messages to receivers and receivers are informed that these messages are targeted through a disclosure, it is possible that this projects onto their perceptions of the source. For the receiver, it could feel, for instance, that the source is attempting to manipulate them. The first concept in this dissertation that can be distinguished regarding this is users' attitude towards the politician advertised in the message. Study II investigated this concept. It was predicted that this concept would be related to users' cognitive and affective resistance as well as their perceived credibility of the source (Van Reijmersdal et al., 2016). As mentioned earlier in this discussion section, predicted relationships on the affective resistance route were not found in this study (see Section 8.3 Reactions to the targeted attempt). However, the results showed that there was a negative relationship between cognitive resistance and users' attitude towards the politician. This means that when users counterargue with the message, this relates to less favorable attitudes towards the politician depicted in the content (Carr & Hayes, 2014; Deng et al., 2020). Moreover, a positive relationship between source credibility and attitude towards the politician was found, which means that if users have more favorable perceptions towards the source of the message, they also have those towards the politician depicted in the message (Kumkale et al., 2010). In a multiparty system, the source of a message is often the political party or its social media account. Thus, there is a distinction between the source of the message and the politician, as they are often not the same. Furthermore, there might be a difference between attitudes towards the source and the politician; it does not always have to be that an individual with a preference for the Democratic Party in the United States is, was, or will be a fan of Joe Biden.

Study II showed significant results for the route of cognitive resistance but not for affective resistance. The use of microtargeting has been found to reinforce party evaluations and make voters feel more strongly connected to the party (Lavigne, 2021). It has also been

found that if people are made aware of the targeting taking place, this positively relates to their party evaluations, but concurrently, users' perceptions of the source's manipulative intentions are negatively related to party evaluations (Hirsch, Binder, et al., 2023). Taking this into account, it is possible that users resist the message cognitively by counterarguing with it at the moment of receiving the message, but that an affective reaction, such as anger, does not occur because of a single message on a social media platform. Moreover, it might even be that the use of disclosures can be seen as a positive signal instead of a signal that prompts negative affective reactions, as research has shown that it leads to more favorable attitudes towards sources and messages (Stubb et al., 2019). It is possible that transparency about the persuasive intentions of a source may project positively on the source, and that it is perceived as more honest by enabling users to integrate the intentions of the source or message into their perceptions of the source and reduce their uncertainty (Carr & Hayes, 2014).

Another reason for the result that the relationship between cognitive resistance and source credibility is stronger than the relationship between cognitive resistance and attitude towards the politician might be that the name of the political party was more prevalent for participants in the experiment. While both the party and politician were fake, participants were at least able to read the name of the party. The image of the person in the picture did not lead to any recognition in the participants' minds, because the photo used was a stock photo. Therefore, it might have been the case that participants allocated more attention to reading the name of the party and the statement in the stimulus material and did not look at the politician that much. In addition, the name of the party is depicted above the image, as is the case on Facebook. However, in line with the previous argument, this might have led to participants focusing on the name of the party first and thereafter looking at the rest of the post. Moreover, these results might indicate that users perceive the source of the message as the one actually presenting the persuasive message, whereas the politician is only advertised in it because they are part of that party. This might lead to certain perceptions being activated more in the reaction to that party, but less towards the advertised politician. However, this interpretation is on a speculative basis, since in a real-world scenario, it is very likely that existing attitudes towards both the politician and the party do not always align; therefore, the effects may shift in the other direction if people appreciate the party, but not the politician, or vice versa. This could happen if there is a large discrepancy between attitudes towards the politician and attitudes towards the party, which might be a rare situation but is not impossible. Through this discrepancy, it is probable that the factor that participants appreciate the most, either the party or the politician, is perceived as the source of the message. In times of campaigning, this might be less

problematic because most people are exposed to multiple persuasive campaigning messages, but for research, this makes it more complicated to ensure that participants correctly distinguish between the party and the politician.

A concept that has also been widely adopted to investigate the effects of sender–receiver relations is source credibility. Source credibility influences how individuals utilize and process information from specific sources (Madsen, 2019). If sources are perceived as having high expertise and trustworthiness, political studies show that people are more likely to vote for them and increase their intention to vote in the election (Hetherington, 1999; Housholder & LaMarre, 2014). In this dissertation, Studies I and II investigated source credibility. While Study I found no relationship between the different types of disclosures or persuasion knowledge and users' perceived credibility of the source, Study II found that cognitive resistance was negatively related to this construct. Moreover, the study found that higher levels of source credibility were positively related to attitudes towards the politician.

If a message comes from a credible source, it becomes more persuasive (Pornpitakpan, 2004). However, as stated earlier, the campaigning and voting processes are very complex. Although the process is a high-involvement activity, there is also a less constant flow of information. In regular advertising, companies initially run a campaign to promote their newest products, after which the product is usually readily available (e.g., a new sports drink in grocery stores). In political communication, it is common that surrounding elections there are several peaks in communication and a larger spread of persuasive messages (Eren, 2021; Hmielowski et al., 2020). Nevertheless, campaign ads are found to have both short-term and long-lasting persuasive effects, which means that communication at a certain point in time can still have effects later on (Gerber et al., 2011). Fluctuations in the amount of communication when countries are not in election years may influence receivers' source credibility. For instance, the source could be perceived as less credible when there is less communication. However, this may only be the case for members of the public who are less politically interested but still vote in elections. For people with a higher interest in politics, the flow of information about politics may be constant (e.g., because they watch political television shows or actively engage with political channels on social media), which could also lead to fewer fluctuations in their perceptions of political sources. Additionally, persuasive messages from parties aim to present the party or its candidates in the most positive way. If the message fulfills its goals, certain voters will have a more favorable credibility perception of the party or politician.

In conclusion, source credibility remains an important factor in the process of voting, as it not only relates to voters' attitudes regarding candidates but also enables research to

identify the specific dimensions of credibility (e.g., trust) that people find more important in political candidates. However, the potential fluctuations in the levels of perceived credibility make it challenging and more important for all dimensions of the concept to continue to be investigated regarding the effects of PMT. Finally, as the use of disclosures as transparency measures has been found to have direct effects on users' credibility perceptions, it is important to consider this factor in future research as disclosures are regulated and become mandatory.

8.4.2 Message-related outcomes

In this dissertation, Study I investigated the effects of different disclosure labels on users' perceived credibility of the message targeted at the user. In that study, no effect of the different types of disclosures on message credibility was found. In line with this, no mediating effect of persuasion knowledge on the direct effect was found. In addition, Bayesian analyses showed substantial evidence in favor of the absence of any effects of disclosures on message credibility. While other work regarding transparency and political communication finds that indeed, message and source credibility are distinguishable constructs, and that the most prominent and red-colored warning leads to lower credibility of the message but not the source, this work found no effects that resemble something like that (Dobber, Kruikemeier, Votta, et al., 2023). The most consistent finding throughout Study I was that there were no effects of the different manipulations on any of the measures, including message credibility. One reason for this is that the participants in that study did not notice the disclosure labels to which they were exposed. However, looking at the mean scores per group, the message credibility was higher than the source credibility, but around the same level as source trustworthiness. Moreover, the results show that the effects of the manipulations on message credibility were small but non-significant, whereas the effects of the manipulations on source credibility were close to nothing and non-significant. Even on a speculative basis, this finding underlines the results from other studies that show that message credibility is a construct distinct from source credibility. For instance, it is possible that the receiver thinks that a political advertisement is too persuasive or that it is just promoting a political actor instead of a party ideology; this might even be the case for voters who already prefer that specific party. The lack of recall of the different disclosures in this study might have given the participants the idea that they were looking at a regular Facebook post instead of a sponsored or targeted advertisement. While it was never the main objective of this study, the results show that the disclosures that Facebook currently uses are not prominent enough to be noticed and therefore do not contribute to transparency, and thus did not lead to differences in message credibility.

In conclusion, contrary to existing research, this work found no relationship between targeting or persuasion knowledge and users' perceived credibility of the messages that they received. While the findings of the studies in this dissertation did not reveal any changes in message credibility through the use of disclosures, the usage of these labels may still have consequences for the way receivers interpret the message. Moreover, this finding might indicate that source and message credibility are separate constructs, as the previous sections show that relations between the usage of disclosures and source credibility were found.

8.5 Targeting calculus

The fourth and final aspect of this dissertation provides insights into users' privacy perceptions through the rationale of the privacy calculus theory (Culnan & Armstrong, 1999). By doing so, users' perceived benefits and privacy concerns were investigated and subsequently scrutinized to determine whether these perceptions led to differences in their intentions to engage in privacy protection behavior. The results of Study III show that if users are made aware of the targeting practices occurring on a platform, this activates targeting knowledge. In turn, targeting knowledge is positively related to users' perceived benefits, but not their privacy concerns. These results agree with work that found higher levels of positive evaluations of content when it is personalized (Brinson & Eastin, 2016; Tam & Ho, 2006). It is possible that the fluency of the content, as it is personalized, leads to easier absorption, which would require less cognitive resources and, therefore, less attention to the content in general. Hereby, the receiver potentially disregards the disclosure, as they focus more on the content itself, which is why people browse social media platforms. However, examining the recall of the disclosures in this study, the levels are substantially higher than those in the other studies in this dissertation. This again underlines the argument that people do not mind receiving targeted advertisements. However, it should be noted that because this experiment was conducted in a controlled setting, it might be that participants were aware that it was not their personal data that were used to target the message at them. Even though participants answered statements indicating their point of view on climate change regulations to replicate microtargeting, this is not the same as how microtargeting is expected to be implemented on social media platforms. In addition, users may not care much about the use of their data. While personalized advertisements might bring a sort of immediate gratification if they are personalized correctly, the use of data by platforms is not something that most people are aware of. While disclosures may contribute to that awareness by informing people that they are targeted by means of their data, it is also possible that users do not care about their data usage

by platforms, even if they are informed. This would align with recent work on transparency regarding political targeting that did not find a relationship between exposure to transparency information and privacy concerns (Dobber, Kruikemeier, Helberger, et al., 2023). Regarding the absence of a relationship between targeting knowledge and privacy concerns, it is important to note that, while disclosures could evoke privacy concerns, they might also decrease existing concerns. Platforms and advertisers can be viewed as sincerer and transparent, which could in turn decrease users' privacy concerns. This agrees with the finding that being transparent about personalization using disclosures decreases the feelings of vulnerability that occur when users suspect personalization (Aguirre et al., 2015; Van Ooijen, 2022).

As an outcome of the 'calculation' in the privacy calculus, Study III investigated users' intention to engage in privacy protection behavior. The results show that users' perceived benefits do not relate to a lower intention to engage in privacy protection behavior, whereas privacy concerns do relate positively to these intentions. As described earlier, this result agrees with work on personalization (Van Ooijen, 2022). Although it is still important to improve transparency regarding targeting practices, it is possible that this does not relate to users' intention to engage in privacy protection behavior. Moreover, the study did not investigate what measures people have already taken to protect their data and privacy online. It is possible that the baseline for some people is already high, as small steps regarding privacy protection behavior can be easily taken (e.g., setting DuckDuckGo as a default search engine or using a browser's built-in tracker prevention). Conversely, the study showed that, as predicted, privacy concerns relate to users' intention to engage in privacy protection behavior. This confirms the results of earlier research on microtargeting and privacy (Binder et al., 2022; Stubenvoll et al., 2022). However, it should be noted that intentions to engage in privacy protection behavior do not always translate directly to actual behavior, and that the subjects in most measurement instruments regarding privacy protection behavior do not always translate the same for all users. For instance, changing social media settings to hide advertisements or changing advertising preferences may not always be possible. In line with this, deregistering from a platform to protect data might be a harsh measure for some people, as the number of users on social networks continues to increase (Pew Research Center, 2021). Furthermore, the benefit perceptions and privacy concerns of users might depend on the congruency they experience with the political party or actor who sent the message. While the targeting techniques that parties employ might overlap, it is possible that users do not mind being targeted by the parties they prefer in the first place. This would mean that they perceive more benefits and fewer concerns, which in turn could still lead to lower or higher intentions to engage in privacy

protection behavior. Finally, the fact that Study III found a relationship between privacy concerns, but not perceived benefits, and the intention to protect oneself online does not necessarily imply that users weigh the benefits and privacy concerns regarding targeting. This merely shows that when users have more privacy concerns, disregarding whether they perceive the targeted advertisement to be beneficial, they might aim to protect themselves online. This emphasizes that privacy concerns, perceived benefits, and privacy protection behavior might be independent constructs, and that protection behavior might not be the outcome of a calculation that people make between benefits and risks, as discussed in earlier research (Dienlin & Metzger, 2016; Meier, Schäwel, et al., 2021).

In conclusion, this dissertation finds that receivers of targeted messages do not seem to mind that they are being targeted. While it is possible that they appreciate the personalized message because it might fit their needs at that specific moment, it could be that this overrules the potential privacy concerns that could come with the message. Moreover, perceived benefits do not relate to lower levels of intention to engage in privacy protection behavior, whereas privacy concerns do relate to higher levels of these intentions. It is possible that users do not weigh these constructs, as described in the privacy calculus, but that the constructs are independent of each other. Finally, users may have different benefit perceptions and concerns depending on the congruency of the message and not microtargeting as a technique.

9. Theoretical implications

Through the empirical results of this dissertation it was investigated what happens if users are exposed to microtargeted messages labeled with disclosures. This was accomplished by building on two existing theories: the persuasion knowledge model and source credibility. This dissertation extends these theories through their application in the context of political targeting. The implications of the overarching theories in this dissertation are described below.

9.1 Persuasion and targeting knowledge

In Studies I and II of this dissertation, the persuasion knowledge model was used to investigate how receivers of microtargeted messages recognize these messages based on their prior experience and existing knowledge (Friestad & Wright, 1994). Moreover, it was scrutinized whether targeting disclosures enabled users to activate higher levels of persuasion knowledge. In Studies II and III, a recently adopted version of persuasion knowledge was used in the form of targeting knowledge (Binder et al., 2022). Based on the results of these studies, it can be concluded that, in line with other research, the effects of disclosures on persuasion

knowledge are diverse and largely dependent on the prominence of the disclosure. This finding suggests that the persuasion knowledge model is a fitting theory to scrutinize receivers' recognition of persuasive attempts and that, as an extension, targeting knowledge can be applied in the case of microtargeting (disregarding whether it is political or not). Moreover, the results of Study IV show that measuring people's visual attention or self-reported recall of a disclosure does not always indicate whether they understood the information that was disclosed. While in current research the most used measure of the effectiveness of disclosures is the recall of the participant, it would be informative if it would be investigated if participants actually understand what processes take place regarding their personal information and how it is used to deliver advertisements to them. However, the PKM itself distinguishes between the *target*, *agent*, and *persuasive episodes*. It would be appropriate to extend it with a fourth perspective in the case of personalization and targeted messages.

While the PKM describes what the target and agent know about each other, the impact of the technique used is not present in the model in its current form. However, it could be argued that while the model describes the persuasion attempt, more focus should be placed on the *techniques* currently used in the process of persuasion. While it is difficult to replicate personalized advertising in research, the model is not only used to measure people's recognition of a persuasive attempt but also to explain the process of influencing targets by advertisers or parties. Although the persuasion knowledge model initially focused on users' recognition, users need to increasingly employ knowledge to recognize advertising since it has become increasingly covert (Wojdynski & Evans, 2020). While the measurement of targeting knowledge is a great step forward in extending the measurements typically used to measure persuasion knowledge, the persuasion knowledge model would benefit from the inclusion of the technical aspects of persuasion, as more persuasion occurs online.

In line with this, the aspect of techniques can also be described from both roles used in the PKM (i.e., *agent* and *target*). The PKM states that both roles have knowledge of the topic, persuasion, and each other. Not only the knowledge of the technique by the target or receiver will shape the persuasive episode, but also the knowledge of the technique by the sender will influence the persuasive episode, as the attempt can differ depending on the sender's knowledge of personalization in advertising. An example of this can be mistargeting, which explains a situation in which political parties aim to microtarget users but fail, leading to mistargeted advertisements that do not reflect the interest of the receiver (Hirsch, Stubenvoll, et al., 2023). Thus, while knowledge of the technique is a fitting aspect in the PKM regarding the role of the target, the role of the agent should also include knowledge of the technique, as

this knowledge will also shape the persuasive episode and make a difference for both the agent and target of the advertisement.

Finally, following the argument of Dobber et al. (2023), this dissertation suggests that general persuasion knowledge might be easier to activate than specific targeting knowledge, as the general public does not always have previous experience with microtargeting or at least might not be aware of those experiences. Additionally, targeting might not be properly understood by the public. The challenge regarding targeting knowledge might be that it is harder to activate than persuasion knowledge, which usually refers to regular advertising. As the authors argued, it would be informative to measure what individuals know before and after exposure to transparency-improving information (Dobber, Kruikemeier, Helberger, et al., 2023).

9.2 Source credibility

In Studies I and II of this dissertation, the construct of source credibility was examined to determine whether the utilization of microtargeting and increased transparency influenced the perceived credibility of the source among receivers. The significance of source credibility as a predictor of voting behavior has been well established in existing research (Hetherington, 1999; Housholder & LaMarre, 2014; Madsen, 2019), and it is thus necessary to investigate the impact of microtargeting as a new form of political communication. The results of Study II revealed that when individuals counterargue with a message, their perceived source credibility decreases. However, it was not found that negative affect alone decreases source credibility. Although receivers of microtargeted messages may not affectively resist these messages, it is still possible that users' source credibility may decrease if a message is more strongly suggestive or polarizing. In addition, if users receive a message containing a disclosure label, they may process it more critically (as found in Study IV), leading to lower levels of perceived credibility as they become more aware of the persuasive nature of the message. This perceived persuasiveness can be seen as a change in the meaning of the message. If disclosures are effective in making receivers aware of the targeting that takes place on a platform, this might lead to yet another change in the meaning of the messages when the disclosures also incorporate information about what data the source used to deliver this targeted advertisement. In summary, although the results of this dissertation are diverse, source credibility remains an important concept for investigating PMT.

It would be beneficial to examine the relationship between source credibility and individuals' attitudes towards politicians and voting intentions in future research. While the

intention to vote was not investigated in the current work, it was found that if people perceive a source as more credible, they will have a more favorable attitude towards that candidate. However, if senders of messages employ PMT as a technique, this might also lead to negative feelings for the receivers, which, in turn, might lead to lower credibility perceptions and as an extension of that a difference in the intention to vote for the party of the sender. Thus, it is implied that it would be beneficial to investigate both concepts simultaneously, as while the concepts are strongly related, the source of the message is not always the same as the politician that is advertised.

9.3 Resistance

In Study II of this dissertation, users' resistance to a microtargeted message was investigated. It was predicted that if users were aware of the message being persuasive and a targeted advertisement, this would lead to more resistance. Resistance in this dissertation was viewed as a way to cope with targeted advertisements. This way of coping was predicted to occur in two ways (Knowles & Linn, 2004, p. 7). The first was a cognitive way in which users would think critically about the message and would not believe it. The second was an affective way in which users would not like the message and the emotions it would make them feel.

It was predicted that awareness of the message's nature through persuasion knowledge would relate to users counterarguing (cognitive resistance) with the message and also have a negative affect (affective resistance) towards the message. Only the relationship between persuasion knowledge and cognitive resistance was found in the second study. This result shows that if people are more aware of the persuasive nature of a message, they tend to resist the message cognitively by counterarguing with it. However, negative emotions such as anger and sadness do not seem to be experienced by receiving a microtargeted advertisement.

It is possible that if users are aware of the political message being an advertisement, they try to reject the message by counterarguing with it, as they believe it is not honest and that its sole purpose is to influence them. Moreover, it is possible that users actively counterargue with these messages as they actively process the information in the message, as political communication can be seen as a high-involvement process. Conversely, it is possible that the study found no relationships of affective resistance, as users might simply not become angry or sad because of an advertisement they receive. The construct of affective resistance might be operationalized through concepts that are very strong emotions, such as sadness and anger. Yet, other affective reactions such as feeling annoyed might still occur after receiving a targeted message, but might be measured differently, as the emotional strength of feeling annoyed might

be lower compared to feeling angry or sad. Furthermore, affectively resisting advertisements might not happen in the case of social media. On platforms, it is easy to scroll to the next post and try to ignore the advertisement out of a caution to avoid experiencing negative emotions. Conversely, counterarguing with content might be something that some users automatically do if the political advertisement contains a statement, as shown in Study II. These implications show that counterarguing, as an operationalization of cognitive resistance to an advertisement, seems to be a fitting concept to at least partially unravel how users react to targeted political advertisements they might receive on social media.

9.4 Targeting calculus

In Study III, the privacy aspect of microtargeting was investigated using Culnan and Armstrong's privacy calculus theory (1999). Using this theory, it was scrutinized how receivers of targeted messages perceive benefits and privacy concerns regarding the technique, and if subsequently, this related to fluxes regarding their intentions to engage in behavior that would protect their data online. Microtargeting is a technique that primarily relies on user data, either self-disclosed or gathered by platforms. However, as most individuals were previously unaware that they were being targeted, they were not expected to have perceptions concerning their data and its protection. However, as disclosures are beginning to become regulated by institutions, this has the potential to inform users about the targeting occurring on platforms, potentially leading to differences regarding their privacy perceptions.

While it was predicted that greater awareness of the targeting used to deliver the message would positively relate to benefit perceptions as well as privacy concerns, proof for only the first relationship was found. Moreover, it was predicted that benefit perceptions would be negatively related to users' intention to engage in privacy protection behavior, and conversely, that privacy concerns would be positively related to these intentions. Nevertheless, only proof of the relationship between privacy concerns and protection behavior was found. These results may indicate that users do not mind that they are being targeted and that they perceive the benefits of targeting, but do not have concerns. Concurrently, it should be noted that the level of privacy concerns was already high in Study III, which might have influenced the result that no significant relationship was found between targeting knowledge and privacy concerns. Accordingly, users may primarily focus on the privacy concerns of targeting by default. While this might have led to the absence of a significant relationship with targeting knowledge, it does not mean that users do not experience privacy concerns regarding microtargeting. However, as the results showed a relationship between concerns and users'

intention to engage in privacy protection behavior, it might be possible that indeed, higher levels of targeting knowledge do not relate to more privacy concerns, as these might already be high.

These results imply that users might not make a calculated decision to engage in less or more data protection behavior. While the benefits of receiving a microtargeted message might exist, they might coexist with users' privacy concerns. Privacy perceptions regarding microtargeting might be described using a duality approach rather than a calculus approach, which assumes that users outweigh benefits and concerns. In this duality approach, users may experience both benefits and risks without weighing them to engage in certain behaviors. Moreover, contradicting the calculus rationale, it is possible that if benefits increase, concerns also increase, particularly if users are made aware that they are targeted. If individuals perceive a message as beneficial but concurrently to receiving that message, receive a disclosure that provides the information that they are targeted by means of their information, this does not necessarily mean they perceive the message as less beneficial. This might lead to them having more privacy concerns while still perceiving the message as beneficial. In contrast, returning to the example of mistargeting used earlier, it is possible that this works in both directions. If users are mistargeted or receive a message that is not fitting, this message will be less beneficial. However, this might also lead to fewer privacy concerns, disregarding the disclosure that will be on the message, as the data and information that platforms or advertisers use might be seen as invalid or simply wrong. This might lead to users being less cautious about the fact that they are targeted, as platforms or advertisers do not use the information to effectively microtarget the user, making the message feel less creepy or intrusive.

These implications emphasize that users' privacy regarding PMT might be a concept that is better described as having a positive and negative side (i.e., a benefit and risk side) which can coexist rather than outweigh one another. In certain situations, the positive aspects may outweigh the negative aspects, or vice versa. However, it seems more appropriate to investigate both positive and negative aspects separately without assuming that a negative relationship exists, where an increase on one side will automatically lead to a decrease in the other.

10. Practical implications

In addition to the theoretical implications of this dissertation, practical implications arise from the implementation of disclosures as a means of improving microtargeting transparency. The practical implications are divided into three subsections, describing specific

implications for designing disclosures in an effective way, for political parties that are using microtargeting, and finally for regulatory bodies that are implementing disclosures.

10.1 Implications for designing effective disclosures

First, the implications of using disclosures to inform users about targeting practices are discussed. In line with existing research, the studies in this dissertation emphasize the importance of prominence (Amazeen & Wojdynski, 2020). While prominence is an important aspect of disclosures, as they need to be seen in order to inform users, it should be stressed that this dissertation found multiple factors that can increase the prominence of disclosures. First, the use of at least a regular (if not bolder) font, compared to a thin light-gray version of a font, will increase the prominence of the disclosure. Second, if there is specific information that is not the same for every receiver (e.g., data used to fill in parameters such as gender and online behavior), these terms can be underlined to emphasize that additional information can be found there. Finally, using a color that stands out compared to the corporate design, or a color that is effective in attracting attention in general, like red, will also increase the prominence of the disclosure if at least a part of the text in the disclosure is that color.

Moreover, the location of the disclosure differed among the studies in this dissertation. In Studies I and II, the disclosure was placed above the content, while in Studies III and IV, it was placed below the content. The reason for this was that the disclosures replicated the use of disclosures on the platforms used in these studies, namely Facebook and Instagram. Although existing research indicates that disclosures above the content are more effective, this dissertation cannot replicate these results (Wojdynski & Evans, 2016). The largest implication regarding location in this dissertation is that the larger disclosure used on Instagram seems to be more effective than the disclosures above the content on both Instagram and Facebook. However, it should be noted that the location of a disclosure should be a place that makes sense along with users' scrolling behavior by putting the disclosures in a timeline and not in an external space or on a different page that would make the user leave the timeline.

Finally, the combination of text and a logo is important. In this dissertation, Studies III and IV used a combination of an icon ("I") and text in the disclosure. This was done to show users that there was extra information about the advertisement in this location, as this has been found to be an effective strategy for improving the attention that users have towards disclosures and, eventually, their effectiveness (Binford et al., 2021; Boerman et al., 2015). The recall in these studies was higher than that in the other two studies in this dissertation. This implies that

it is important that a combination of a symbol and text is used to draw users' attention, as simply including textual information may not be sufficient.

10.2 Implications for political parties using microtargeting

The second section of these practical implications concerns political parties using or planning to use microtargeting as a way of communicating with their target audience. The regulation of disclosures that inform users means that, for political parties and other actors that send messages, larger parts of their target groups may be aware that they are using microtargeting. This can lead to a change in meaning in the receiver's perception; for instance, a message can be perceived as less sincere or more manipulative. However, awareness may also activate other processes.

Microtargeted users could resist messages more if they were aware of the targeting. In addition, users may employ a more critical and skeptical way of processing advertisements if they are aware that they are being targeted. In addition, this resistance not only decreases perceived credibility but also users' attitudes towards the advertised politician. For political parties and advertisers working with them, this implies that using PMT should be a well-considered choice, as it might indirectly be counterproductive for affecting voters (Hetherington, 1999; Housholder & LaMarre, 2014).

However, the awareness of being targeted that can be activated through a disclosure might also lead to more positive party evaluations (Dobber, Kruikemeier, Helberger, et al., 2023; Hirsch, Binder, et al., 2023). In parallel, the current dissertation indeed finds that exposure to a disclosure can lead to higher levels of recognition of microtargeting, which in turn can lead to more benefit perceptions but not privacy concerns. Thus, while research finds diverse processes that can occur in users' minds, it is possible that receivers of targeted messages appreciate personalization, and that this feeling is stronger than the potential risks they experience regarding their privacy and data.

10.3 Implications for regulatory bodies implementing disclosures

Finally, the third section focuses on the practical implications of this dissertation for institutions that regulate the use of disclosures. In different parts of the world, these might be different organizations, governments, or regulatory bodies. First, it should be noted that regarding the European Union, the way in which disclosures, particularly targeting disclosures, are described in the DSA is a major step forward. However, the factors within the DSA are up for interpretation, both on the level of granularity (e.g., how specified the targeting information

is displayed) and design. There are some potential improvements to the Act that other institutions might benefit from as well. As research shows that prominence is a determining factor for the attention that users pay to disclosures, it is important to improve regulations and add design characteristics which can improve prominence. For instance, a colored box around the disclosure will improve the attention it attracts. As shown in Study II, the color red, as it attracts attention more easily, might lead to better effectiveness of disclosures than colder colors or colors in the corporate design of the platform. Additionally, as described earlier, the combination of an icon and text might be a feasible implementation for the design of disclosures, as it has been found to improve the recall of disclosures in Studies III and IV.

In addition to improving the potential attention-increasing characteristics of disclosures, receivers would also benefit from uniformity regarding disclosures. For example, the EU could regulate disclosures using a standardized disclosure, as has been done in food labeling with the Nutri-Score, which has been voluntarily adopted in most Western European countries. The major difference is that the Nutri-Score cannot be forced upon countries and has been voluntarily adopted; disclosures regarding (targeted) advertising could be implemented in a less voluntary way, as the DSA has been accepted (European Commission, 2022). The studies in this dissertation showed that the recall of a sponsored disclosure, which was based on the actual disclosures that platforms use nowadays, is often higher than the recall of the targeting disclosure. One of the reasons for this is that receivers might need time to adapt to and get to know this new disclosure. Through repeated exposure to the newly developed targeting disclosure, a familiarity effect may occur (Montoya et al., 2017), which may lead users to recognize the disclosure more easily. This would improve the transparency of targeting practices. To improve this, there might be an opportunity to use uniform disclosures across all platforms. This would mean that users may be exposed to the disclosure more often, potentially speeding up the phase of becoming familiar with the disclosure, meaning that transparency might improve sooner. This disclosure should contain all important factors regarding the parameters used to target individuals or groups and the most effective design characteristics. This disclosure could cross continents and country borders and would display the receivers of targeted messages how platforms enable advertisers to show users specific advertisements. Potentially shielding citizens, improving transparency, and informing the public.

11. Limitations

This section describes the general limitations of the studies included in this dissertation. First, all studies in this dissertation used cross-sectional data. While this provides information

about a certain point in time, trends and developments over time are not considered. For instance, the first and second studies were conducted just before and just after the German federal election of 2021, and longitudinal data could have accounted for shifts in political interest and the general increase in political communication that usually occurs around elections (Eren, 2021; Hmielowski et al., 2020). Moreover, the use of cross-sectional data yields only limited causality inferences. All the studies measured the effects of disclosures on at least one dependent variable. The studies were only able to investigate the relationships between other variables that were not manipulated. Although based on theories and previous research, it is possible to speculate on some causal effects, real causality was not often tested in the experiments. In particular, when employing disclosures to improve the transparency of microtargeting, it is important to use longitudinal designs to test their effectiveness, as repeated exposure might help users recognize targeted advertisements.

Another limitation of all studies is the fact that besides Study IV, all studies used an online experimental design in which participants only obtained one specific social media advertisement. Moreover, these studies used timers to ensure that the participants viewed the advertisement for a minimum amount of time. While this was done to ensure that participants actually looked at the advertisement, this measure might have decreased the ecological validity, while also potentially overestimating the results of the experiments. The external validity of these experiments was impaired because social media users typically consume large amounts of content on a platform. Moreover, users scroll through timelines at a faster pace than the number of seconds they were forced to look at the advertisements in the studies (Keib et al., 2022). This potentially overestimated the measurements in these studies; for instance, more persuasion knowledge activation might have occurred if participants looked at the advertisement longer (Boerman et al., 2012). In line with this, Study IV allowed participants to browse a timeline containing six posts, and while this study would be on the shorter side regarding the number of questions that participants were asked, the average time of completion was suggestively lower than for the other three studies.

Furthermore, none of the studies used participants' own social media timelines. Only Study IV showed participants an actual timeline on Instagram. However, while the content of the timeline contained posts that would not have been strange to see in the timelines of the participants in that sample, it was not the timeline that a participant would see if they had opened social media on their preferred device. While designing an experiment that includes participants' own social media would improve ecological validity, it would also come with ethical/privacy considerations, as it would be hard to record what the participants saw without

recording their own social media timeline. However, certain usage and gratification strategies may be employed by users of different social media channels or platforms to satisfy various needs.

Finally, the first three studies in this dissertation used stimulus material consisting of a made-up political party and a made-up politician (i.e., an altered stock photo). While this design choice aimed to minimize the effects of existing attitudes towards politicians and parties, it also meant that the ecological validity of this study was lower. In particular, the voting process is very complex, and attitudes are usually developed over longer periods of time. However, Study IV showed participants an advertisement for an existing political party. The downside is that, in this study, two political parties were used and randomized among participants. While this study statistically controlled for which political party was shown to the participants, it is important to note that not every participant might have seen the party to which they mostly relate. However, in microtargeting, this might also happen, as sometimes people still receive more general advertisements that are less tailored or mistargeted (Hirsch, Stubenvoll, et al., 2023).

12. Future research

While microtargeting in its current form is a relatively new technique, transparency measures and disclosures regarding PMT have not been extensively researched. However, research can partially rely on work that has been conducted regarding regular persuasive communication and apply this to the context of targeting. Nevertheless, with the constant development of targeting strategies and platforms, open questions remain. Based on the results of this dissertation, future studies could provide benefits to both research and society by investigating the specific characteristics that would improve the prominence and effectiveness of disclosures. While institutions aim to regulate transparency and disclosures, it is important to conduct research on recently founded social media platforms (e.g., TikTok) and recent targeting possibilities to develop fitting disclosures that are truly effective in improving transparency. Moreover, if the elements that successfully increase prominence are further scrutinized, this would also address one of the largest drawbacks of disclosures, which is that some individuals across most studies do not see or remember them. In line with earlier arguments (see Section 9.1 Persuasion and targeting knowledge), future work should focus on the distinction between what individuals know about targeting in general and about a targeted attempt before and after exposure to a transparency measure. This is important because targeting as a construct is harder to understand than regular advertising, and individuals might

not have developed a certain literacy regarding microtargeting (Dobber, Kruikemeier, Helberger, et al., 2023). While regular persuasive messages are something that people have been exposed to for years now, microtargeted messages might be new for many people, especially because transparency regarding these practices was nonexistent in the beginning. In line with this, it might be that, years from now, literacy regarding microtargeting is higher, as advertisers might adopt the technique more heavily. Therefore, in addition to focusing on disclosures as transparency measures, it would be appropriate to also focus on literacy campaigns and how to improve their effectiveness (Zarouali, Verdoodt, et al., 2020). However, the results of this dissertation also imply that people do not mind being targeted, and mostly perceive the benefits of the technique. Looking across the borders of political communication, microtargeting may also be used for prosocial causes (e.g., in health communication). It is possible that the technique provides benefits as people might be shown information on how to improve their health or why they should take part in certain vaccination campaigns (Gaysynsky et al., 2022). Finally, while attitudes and credibility perceptions have been shown to be important in the voting process, it is important to recognize that these concepts are also subject to change over time, not only by incidents with or by politicians, but also because political communication increases around election periods, which can influence receivers' attitudes (Eren, 2021; Hmielowski et al., 2020). Longitudinal research could answer questions regarding these potential fluctuations but also investigate social media platforms and how they develop over time and keep adding (and deleting) functions (Debatin et al., 2009). In conclusion, research should adopt a practical point of view and focus on the development of disclosures in their most effective form. Moreover, research should focus on the user and investigate how they internalize and learn from information about targeting, while also investigating the behavior of users' attitudes over time. Finally, while the consensus about microtargeting and the potential risks associated with using the technique is rather negative, research should scrutinize the potential of the technique regarding prosocial causes, such as health communication.

13. Conclusion

As persuasion and targeting techniques evolve, it is essential to inform receivers of targeted advertisements as much as possible. Whereas only when they are informed, voters are able to make the best possible decision and choose the party or actor that fits their needs and viewpoints. While disclosures are a solution that would not distract users who want to consume content on social media too much, advances should be made in the design of disclosures. The

most important factor of the design is the prominence of the disclosures, which needs to be improved to subsequently improve their effectiveness.

In addition to investigating targeting disclosures, this dissertation applied four aspects to investigate psychological mechanisms and scrutinize what happens when receivers are exposed to an advertisement that contains a targeting disclosure. The first aspect was the recognition of the targeted nature of a message. The findings indicated that disclosures can enable receivers to recognize that a message is persuasive and targeted at them. However, prominence is found to be an important determinant of disclosure recognition, which, in turn, is a prerequisite for the disclosure's effectiveness. Additionally, this prominence seemed to be lacking in the sponsorship disclosures currently utilized on social media. Receivers' reactions to targeted messages were scrutinized in the second aspect of this dissertation. It was found that receivers, when made aware of the nature of the message, were able to cope with it by processing it critically. In addition, this processing, or even merely the awareness of the targeted nature, was found to lead to receivers resisting the message cognitively, but not affectively. The third aspect in this work described how recognition and reactions to the targeted advertisement influence receivers' beliefs and attitudes regarding the message and the source that sent it. It was discovered that if receivers are aware of the targeting used to deliver the message to them, this can lead to less favorable attitudes and lower credibility perceptions regarding the message's source. The fourth and final aspect of this dissertation described the targeting calculus. Building on the privacy calculus theory rationale, the findings showed that if receivers are aware of the targeted nature of the advertisement, they do not develop more privacy concerns, but recognize the benefits of the technique. In turn, the privacy concerns they might have, relate to intentions to engage in behavior to protect their data online, while the benefit perceptions do not.

In conclusion, as disclosures become more regulated than they were before, this dissertation provides vital information on the most important design characteristics. Disclosures are most effective if they are more prominent than they currently are. This prominence can be improved by using text that is larger and bolder and text that is red and underlined. Disclosures with this improved prominence would benefit from an icon combined with the text, which will attract more attention and will show receivers of messages that additional information can be found within the text disclosed in the label. If disclosures are designed like this, they will improve transparency regarding microtargeting and thus inform receivers of the occurrence of the technique, instead of just being present but not noticed.

References

Ackermann, K. (2022, July 5). *Alcohol by Volume (ABV): Beer, Wine, & Liquor*. Alcohol.Org. https://alcohol.org/statistics-information/abv/

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing*, *91*(1), 34–49. https://doi.org/10.1016/j.jretai.2014.09.005

Allyn, J., & Festinger, L. (1961). The effectiveness of unanticipated persuasive communications. *The Journal of Abnormal and Social Psychology*, *62*(1), 35–40. https://doi.org/10.1037/h0048493

Amazeen, M. A., & Wojdynski, B. W. (2019). Reducing Native Advertising Deception: Revisiting the Antecedents and Consequences of Persuasion Knowledge in Digital News Contexts. *Mass Communication and Society*, *22*(2), 222–247. https://doi.org/10.1080/15205436.2018.1530792

Amazeen, M. A., & Wojdynski, B. W. (2020). The effects of disclosure format on native advertising recognition and audience perceptions of legacy and online news publishers. *Journalism*, *21*(12), 1965–1984. https://doi.org/10.1177/1464884918754829

Appelman, A., & Sundar, S. S. (2016). Measuring message credibility: Construction and validation of an exclusive scale. *Journalism and Mass Communication Quarterly*, *93*(1), 59–79. https://doi.org/10.1177/1077699015606057

Ayeh, J. K., Au, N., & Law, R. (2013). "Do We Believe in TripAdvisor?" Examining Credibility Perceptions and Online Travelers' Attitude toward Using User-Generated Content. *Journal of Travel Research*, *52*(4), 437–452. https://doi.org/10.1177/0047287512475217

Baik, M., Suk, H.-J., Lee, J., & Choi, K. (2013). *Investigation of eye-catching colors using eye tracking* (B. E. Rogowitz, T. N. Pappas, & H. De Ridder, Eds.; p. 86510W). https://doi.org/10.1117/12.2001141

Banerjee, S. C., & Greene, K. (2012). Role of Transportation in the Persuasion Process: Cognitive and Affective Responses to Antidrug Narratives. *Journal of Health Communication*, *17*(5), 564–581. https://doi.org/10.1080/10810730.2011.635779

Barbu, O. (2014). Advertising, Microtargeting and Social Media. *Procedia - Social and Behavioral Sciences*, *163*, 44–49. https://doi.org/10.1016/j.sbspro.2014.12.284

Barocas, S. (2012). The price of precision: Voter microtargeting and its potential harms to the democratic process. *Proceedings of the First Edition Workshop on Politics, Elections and Data*, 31–36. https://doi.org/10.1145/2389661.2389671

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, *24*(2), 117–150. https://doi.org/10.1023/A:1021226224601

Beckert, J., & Koch, T. (2022). Why context matters in a changing social media environment: The role of placement prominence as a moderator of disclosure effects in sponsored influencer content. *Studies in Communication Sciences*, *22*(1). https://doi.org/10.24434/j.scoms.2022.01.3061

Ben-David, A. (2020). Counter-archiving Facebook. *European Journal of Communication*, *35*(3), 249–264. https://doi.org/10.1177/0267323120922069

Benway, J. P. (1998). Banner Blindness: The Irony of Attention Grabbing on the World Wide Web. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(5), 463–467. https://doi.org/10.1177/154193129804200504

Binder, A., Stubenvoll, M., Hirsch, M., & Matthes, J. (2022). Why Am I Getting This Ad? How the Degree of Targeting Disclosures and Political Fit Affect Persuasion Knowledge, Party Evaluation, and Online Privacy Behaviors. *Journal of Advertising*, 1–17. https://doi.org/10.1080/00913367.2021.2015727

Binford, M. T., Wojdynski, B. W., Lee, Y.-I., Sun, S., & Briscoe, A. (2021). Invisible transparency: Visual attention to disclosures and source recognition in Facebook political advertising. *Journal of Information Technology & Politics*, *18*(1), 70–83. https://doi.org/10.1080/19331681.2020.1805388

Bodó, B., Helberger, N., & De Vreese, C. H. (2017). Political micro-targeting: A manchurian candidate or just a dark horse? *Internet Policy Review*, *6*(4). https://doi.org/10.14763/2017.4.776

Boerman, S. C., & Kruikemeier, S. (2016). Consumer responses to promoted tweets sent by brands and political parties. *Computers in Human Behavior*, *65*, 285–294. https://doi.org/10.1016/j.chb.2016.08.033

Boerman, S. C., Kruikemeier, S., & Zuiderveen Borgesius, F. J. (2017). Online Behavioral Advertising: A Literature Review and Research Agenda. *Journal of Advertising*, *46*(3), 363–376. https://doi.org/10.1080/00913367.2017.1339368

Boerman, S. C., & Van Reijmersdal, E. A. (2016). Informing Consumers about "Hidden" Advertising: A Literature Review of the Effects of Disclosing Sponsored Content. In P. De Pelsmacker (Ed.), *Advertising in New Formats and Media* (pp. 115–146).

Emerald Group Publishing Limited. https://doi.org/10.1108/978-1-78560-313-620151005

Boerman, S. C., Van Reijmersdal, E. A., & Neijens, P. C. (2012). Sponsorship Disclosure: Effects of Duration on Persuasion Knowledge and Brand Responses. *Journal of Communication*, *62*(6), 1047–1064. https://doi.org/10.1111/j.1460-2466.2012.01677.x

Boerman, S. C., Van Reijmersdal, E. A., & Neijens, P. C. (2014). Effects of Sponsorship Disclosure Timing on the Processing of Sponsored Content: A Study on the Effectiveness of European Disclosure Regulations: EFFECTS OF SPONSORSHIP DISCLOSURE TIMING. *Psychology & Marketing*, *31*(3), 214–224. https://doi.org/10.1002/mar.20688

Boerman, S. C., Van Reijmersdal, E. A., & Neijens, P. C. (2015). Using Eye Tracking to Understand the Effects of Brand Placement Disclosure Types in Television Programs. *Journal of Advertising*, *44*(3), 196–207. https://doi.org/10.1080/00913367.2014.967423

Boerman, S. C., Willemsen, L. M., & Van Der Aa, E. P. (2017). "This Post Is Sponsored" Effects of Sponsorship Disclosure on Persuasion Knowledge and Electronic Word of Mouth in the Context of Facebook. *Journal of Interactive Marketing*, *38*, 82–92. https://doi.org/10.1016/j.intmar.2016.12.002

Bohner, G., Crow, K., Erb, H.-P., & Schwarz, N. (1992). Affect and persuasion: Mood effects on the processing of message content and context cues and on subsequent behaviour. *European Journal of Social Psychology*, *22*(6), 511–530. https://doi.org/10.1002/ejsp.2420220602

Bol, N., Dienlin, T., Kruikemeier, S., Sax, M., Boerman, S. C., Strycharz, J., Helberger, N., & de Vreese, C. H. (2018). Understanding the Effects of Personalization as a Privacy Calculus: Analyzing Self-Disclosure Across Health, News, and Commerce Contexts†. *Journal of Computer-Mediated Communication*, *23*(6), 370–388. https://doi.org/10.1093/jcmc/zmy020

Brehm, S. S., & Brehm, J. W. (2014). *Psychological Reactance: A Theory of Freedom and Control.* Elsevier Science. http://qut.eblib.com.au/patron/FullRecord.aspx?p=1839510

Brinson, N. H., & Eastin, M. S. (2016). Juxtaposing the persuasion knowledge model and privacy paradox: An experimental look at advertising personalization, public policy and public understanding. *Journal of Psychosocial Research*, *10*.

Brown-Devlin, N., Lim, H. S., & Tao, J. (2022). Examining the Influence of Algorithmic Message Personalization on Source Credibility and Reputation. *International Journal of Business Communication*, 232948842211264. https://doi.org/10.1177/23294884221126489

Büchi, M., Just, N., & Latzer, M. (2017). Caring is not enough: The importance of Internet skills for online privacy protection. *Information, Communication & Society*, *20*(8), 1261–1278. https://doi.org/10.1080/1369118X.2016.1229001

Buijzen, M., Van Reijmersdal, E. A., & Owen, L. H. (2010). Introducing the PCMC Model: An Investigative Framework for Young People's Processing of Commercialized Media Content. *Communication Theory*, *20*(4), 427–450. https://doi.org/10.1111/j.1468-2885.2010.01370.x

Burgoon, M., Miller, M. D., Cohen, M., & Montgomery, C. L. (1978). AN EMPIRICAL TEST OF A MODEL OF RESISTANCE TO PERSUASION. *Human Communication Research*, *5*(1), 27–39. https://doi.org/10.1111/j.1468-2958.1978.tb00620.x

Cadwalladr, C. (2018, March 18). 'I made Steve Bannon's psychological warfare tool': Meet the data war whistleblower. *The Guardian*. https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump

Campbell, M. C. (1995). When Attention-Getting Advertising Tactics Elicit Consumer Inferences of Manipulative Intent: The Importance of Balancing Benefits and Investments. *Journal of Consumer Psychology*, *4*(3), 225–254. https://doi.org/10.1207/s15327663jcp0403_02

Campbell, M. C., & Kirmani, A. (2000). Consumers' Use of Persuasion Knowledge: The Effects of Accessibility and Cognitive Capacity on Perceptions of an Influence Agent. *Journal of Consumer Research*, *27*(1), 69–83. https://doi.org/10.1086/314309

Carr, C. T., & Hayes, R. A. (2014). The Effect of Disclosure of Third-Party Influence on an Opinion Leader's Credibility and Electronic Word of Mouth in Two-Step Flow. *Journal of Interactive Advertising*, *14*(1), 38–50. https://doi.org/10.1080/15252019.2014.909296

Choi, J. (2016). Why do people use news differently on SNSs? An investigation of the role of motivations, media repertoires, and technology cluster on citizens' news-related activities. *Computers in Human Behavior*, *54*, 249–256. https://doi.org/10.1016/j.chb.2015.08.006

Culnan, M. J., & Armstrong, P. K. (1999). Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation. *Organization Science*, *10*(1), 104–115. https://doi.org/10.1287/orsc.10.1.104

De Keyzer, F., Dens, N., & De Pelsmacker, P. (2022). Let's get personal: Which elements elicit perceived personalization in social media advertising? *Electronic Commerce Research and Applications*, *55*, 101183. https://doi.org/10.1016/j.elerap.2022.101183

De Keyzer, F., Van Noort, G., & Kruikemeier, S. (2022). Going Too Far? How Consumers Respond to Personalized Advertising From Different Sources. *Journal of Electronic Commerce Research*, *23*(3), 22.

Debatin, B., Lovejoy, J. P., Horn, A.-K., & Hughes, B. N. (2009). Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. *Journal of Computer-Mediated Communication*, *15*(1), 83–108. https://doi.org/10.1111/j.1083-6101.2009.01494.x

Decker, H., & Krämer, N. (2023). Is Personality Key? Persuasive Effects of Prior Attitudes and Personality in Political Microtargeting. *Media and Communication*, *11*(3), 250–261. https://doi.org/10.17645/mac.v11i3.6627

Deng, X., Li, M., & Suh, A. (2020). Recommendation or advertisement? The influence of advertising-disclosure language with pictorial types on influencer credibility and consumers' brand attitudes. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12427 LNCS*, 234–248. https://doi.org/10.1007/978-3-030-60152-2_19

Di Plinio, S., Aquino, A., Haddock, G., Alparone, F. R., & Ebisch, S. J. H. (2022). Brain and behavioral contributions to individual choices in response to affective–cognitive persuasion. *Cerebral Cortex*, bhac213. https://doi.org/10.1093/cercor/bhac213

Dienlin, T., & Metzger, M. J. (2016). An Extended Privacy Calculus Model for SNSs: Analyzing Self-Disclosure and Self-Withdrawal in a Representative U.S. Sample: THE EXTENDED PRIVACY CALCULUS MODEL FOR SNSs. *Journal of Computer-Mediated Communication*, *21*(5), 368–383. https://doi.org/10.1111/jcc4.12163

Dillard, J. P., & Shen, L. (2005). On the Nature of Reactance and its Role in Persuasive Health Communication. *Communication Monographs*, *72*(2), 144–168. https://doi.org/10.1080/03637750500111815

Dinev, T., & Hart, P. (2006). An Extended Privacy Calculus Model for E-Commerce Transactions. *Information Systems Research*, *17*(1), 61–80. https://doi.org/10.1287/isre.1060.0080

Dobber, T. (2021). *Effectiveness of Online Political Ad Disclosure Labels: Empirical Findings*. 6.

Dobber, T., Kruikemeier, S., Helberger, N., & Goodman, E. (2023). Shielding citizens? Understanding the impact of political advertisement transparency information. *New Media & Society*. https://doi.org/10.1177/14614448231157640

Dobber, T., Kruikemeier, S., Votta, F., Helberger, N., & Goodman, E. P. (2023). The effect of traffic light veracity labels on perceptions of political advertising source and message credibility on social media. *Journal of Information Technology & Politics*, 1–16. https://doi.org/10.1080/19331681.2023.2224316

Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2021). Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes? *The International Journal of Press/Politics*, *26*(1), 69–91. https://doi.org/10.1177/1940161220944364

Dobber, T., Trilling, D., Helberger, N., & de Vreese, C. (2019). Spiraling downward: The reciprocal relation between attitude toward political behavioral targeting and privacy concerns. *New Media & Society*, *21*(6), 1212–1231. https://doi.org/10.1177/1461444818813372

Dobber, T., Trilling, D., Helberger, N., & de Vreese, C. H. (2017). Two crates of beer and 40 pizzas: The adoption of innovative political behavioural targeting techniques. *Internet Policy Review*, *6*(4). https://doi.org/10.14763/2017.4.777

Dommett, K. (2021). The inter-institutional impact of digital platform companies on democracy: A case study of the UK media's digital campaigning coverage. *New Media & Society*, *25*(10), 2763–2780. https://doi.org/10.1177/14614448211028546

Edell, J. A., & Burke, M. C. (1987). The Power of Feelings in Understanding Advertising Effects. *Journal of Consumer Research*, *14*(3), 421. https://doi.org/10.1086/209124

Elswah, M., & Howard, P. N. (2020). *Tunisia-memo-English.pdf* [PDF]. The Challenges of Monitoring Social Media in the Arab World: The Case of the 2019 Tunisian Elections. https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2020/03/Tunisia-memo-English.pdf

Endres, K., & Kelly, K. J. (2018). Does microtargeting matter? Campaign contact strategies and young voters. *Journal of Elections, Public Opinion and Parties*, *28*(1), 1–18. https://doi.org/10.1080/17457289.2017.1378222

Eren, E. (2021). Education Policies in the Context of Political Communication in Turkey. *European Journal of Educational Research*, *10*(1), 227–240. https://doi.org/10.12973/eu-jer.10.1.227

European Commission. (2022). *Regulation of the European Parliament and of The Council on a Single Market For Digital Services (Digital Services Act) and amending Directive*. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN

Evans, N. J., Phua, J., Lim, J., & Jun, H. (2017). Disclosing Instagram Influencer Advertising: The Effects of Disclosure Language on Advertising Recognition, Attitudes, and Behavioral Intent. *Journal of Interactive Advertising*, *17*(2), 138–149. https://doi.org/10.1080/15252019.2017.1366885

Evans, N. J., Wojdynski, B. W., & Grubbs Hoy, M. (2019). How sponsorship transparency mitigates negative effects of advertising recognition. *International Journal of Advertising*, *38*(3), 364–382. https://doi.org/10.1080/02650487.2018.1474998

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, *15*(2), 115–141. https://doi.org/10.1080/0269993004200024

Lutte contre la manipulation de l'information, 2018, 1202, 0297 (2018). https://www.senat.fr/dossier-legislatif/ppl17-623.html

Flanagin, A., & Metzger, M. J. (2014). *Digital Media and Perceptions of Source Credibility in Political Communication* (K. Kenski & K. H. Jamieson, Eds.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199793471.013.65

Flavián, C., & Gurrea, R. (2009). Users' motivations and attitude towards the online press. *Journal of Consumer Marketing*, *26*(3), 164–174. https://doi.org/10.1108/07363760910954109

Fransen, M. L., Mollen, S., Rains, S. A., Das, E., & Vermeulen, I. (2023). Sixty Years Later: A Replication Study of McGuire's First Inoculation Experiment. *Journal of Media Psychology*, 1864-1105/a000396. https://doi.org/10.1027/1864-1105/a000396

Fransen, M. L., Smit, E. G., & Verlegh, P. W. J. (2015). Strategies and motives for resistance to persuasion: An integrative framework. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01201

Fransen, M. L., Verlegh, P. W. J., Kirmani, A., & Smit, E. G. (2015). A typology of consumer strategies for resisting advertising, and a review of mechanisms for

countering them. *International Journal of Advertising*, *34*(1), 6–16. https://doi.org/10.1080/02650487.2014.995284

Friestad, M., & Wright, P. (1994). The Persuasion Knowledge Model: How People Cope with Persuasion Attempts. *Journal of Consumer Research*, *21*(1), 1–31. https://doi.org/10.1086/209380

Funk, C. L. (1999). Bringing the Candidate into Models of Candidate Evaluation. *The Journal of Politics*, *61*(3), 700–720. https://doi.org/10.2307/2647824

Gandy, O. H. (2000). Dividing Practices: Segmentation and Targeting in the Emerging Public Sphere. In W. L. Bennett & R. M. Entman (Eds.), *Mediated Politics* (1st ed., pp. 141–159). Cambridge University Press. https://doi.org/10.1017/CBO9780511613852.008

Gaysynsky, A., Heley, K., & Chou, W.-Y. S. (2022). An Overview of Innovative Approaches to Support Timely and Agile Health Communication Research and Practice. *International Journal of Environmental Research and Public Health*, *19*(22), 15073. https://doi.org/10.3390/ijerph192215073

Gerber, A. S., Gimpel, J. G., Green, D. P., & Shaw, D. R. (2011). How Large and Long-lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment. *American Political Science Review*, *105*(1), 135–150. https://doi.org/10.1017/S000305541000047X

Ghosh, D. (2018, October 4). *What is microtargeting and what is it doing in our politics?* Mozilla. https://blog.mozilla.org/internetcitizen/2018/10/04/microtargeting-dipayan-ghosh/

Giasson, T., Le Bars, G., & Dubois, P. (2019). Is Social Media Transforming Canadian Electioneering? Hybridity and Online Partisan Strategies in the 2012 Quebec Election. *Canadian Journal of Political Science*, *52*(2), 323–341. https://doi.org/10.1017/S0008423918000902

Gorton, W. A. (2016). Manipulating citizens: How political campaigns' use of behavioral social science harms democracy. *New Political Science*, *38*(1), 61–80. https://doi.org/10.1080/07393148.2015.1125119

Greer, J. D. (2003). Evaluating the Credibility of Online Information: A Test of Source and Advertising Influence. *Mass Communication and Society*, *6*(1), 11–28. https://doi.org/10.1207/S15327825MCS0601_3

Ham, C.-D., & Nelson, M. R. (2016). The role of persuasion knowledge, assessment of benefit and harm, and third-person perception in coping with online behavioral

advertising. *Computers in Human Behavior*, *62*, 689–702. https://doi.org/10.1016/j.chb.2016.03.076

Ham, C.-D., Ryu, S., Lee, J., Chaung, U.-C., Buteau, E., & Sar, S. (2022). Intrusive or Relevant? Exploring How Consumers Avoid Native Facebook Ads through Decomposed Persuasion Knowledge. *Journal of Current Issues & Research in Advertising*, *43*(1), 68–89. https://doi.org/10.1080/10641734.2021.1944934

Hass, R. G. (2014). Effects of source characteristics on cognitive responses and persuasion. In *Cognitive responses in persuasion* (pp. 141–172). Psychology Press.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (Second edition). Guilford Press.

Hayes, J. L., Brinson, N. H., Bott, G. J., & Moeller, C. M. (2021). The Influence of Consumer–Brand Relationship on the Personalized Advertising Privacy Calculus in Social Media. *Journal of Interactive Marketing*, *55*(1), 16–30. https://doi.org/10.1016/j.intmar.2021.01.001

Heim, J., Asting, T., & Schliemann, T. (2002). Medium effects on persuasion. *Proceedings of the Second Nordic Conference on Human-Computer Interaction - NordiCHI '02*, 259. https://doi.org/10.1145/572020.572060

Hervet, G., Guérard, K., Tremblay, S., & Chtourou, M. S. (2011). Is banner blindness genuine? Eye tracking internet text advertising. *Applied Cognitive Psychology*, *25*(5), 708–716. https://doi.org/10.1002/acp.1742

Hetherington, M. J. (1999). The Effect of Political Trust on the Presidential Vote, 1968–96. *American Political Science Review*, *93*(2), 311–326. https://doi.org/10.2307/2585398

Hirsch, M., Binder, A., & Matthes, J. (2023). The Influence of Political Fit, Issue Fit, and Targeted Political Advertising Disclosures on Persuasion Knowledge, Party Evaluation, and Chilling Effects. *Social Science Computer Review*, 08944393231193731. https://doi.org/10.1177/08944393231193731

Hirsch, M., Stubenvoll, M., Binder, A., & Matthes, J. (2023). Beneficial or Harmful? How (Mis)Fit of Targeted Political Advertising on Social Media Shapes Voter Perceptions. *Journal of Advertising*, 1–17. https://doi.org/10.1080/00913367.2023.2175081

Hmielowski, J. D., Kelvin, W. F., Hutchens, M. J., Silva, D. E., Beam, M. A., Donaway, R. R., & York, C. (2020). Communication Behaviors During Presidential Elections. *Public Opinion Quarterly*, *84*(S1), 309–331. https://doi.org/10.1093/poq/nfaa013

Holt, K., Shehata, A., Strömbäck, J., & Ljungberg, E. (2013). Age and the effects of news media attention and social media use on political interest and participation: Do social

media function as leveller? *European Journal of Communication*, *28*(1), 19–34. https://doi.org/10.1177/0267323112465369

Housholder, E. E., & LaMarre, H. L. (2014). Facebook Politics: Toward a Process Model for Achieving Political Source Credibility Through Social Media. *Journal of Information Technology & Politics*, *11*(4), 368–382. https://doi.org/10.1080/19331681.2014.951753

Infante, D. A., & Grimmett, R. A. (1971). Attitudinal effects of utilizing a critical method of analysis. *Central States Speech Journal*, *22*(4), 213–217. https://doi.org/10.1080/10510977109363075

Jain, S., & Purohit, H. C. (2022). Privacy concerns and avoidance behaviour towards data-driven online behavioural advertising. *Business Analyst Journal*, *43*(1), 1–12. https://doi.org/10.1108/BAJ-08-2022-0024

Jamieson, K. H. (2013). Messages, micro-targeting, and new media technologies. *Forum (Germany)*, *11*(3), 429–435. https://doi.org/10.1515/for-2013-0052

Jansen, M.-P., & Krämer, N. C. (2023). Empty Transparency? The Effects on Credibility and Trustworthiness of Targeting Disclosure Labels for Micro-Targeted Political Advertisements. *Weizenbaum Journal of the Digital Society*, *3*(1). https://doi.org/10.34669/WI.WJDS/3.1.5

Janssen, L., Fennis, B. M., & Pruyn, A. Th. H. (2010). Forewarned is forearmed: Conserving self-control strength to resist social influence. *Journal of Experimental Social Psychology*, *46*(6), 911–921. https://doi.org/10.1016/j.jesp.2010.06.008

Janssen, L., Fransen, M. L., Wulff, R., & Van Reijmersdal, E. A. (2016). Brand placement disclosure effects on persuasion: The moderating role of consumer self-control: Self-control and brand placement disclosure. *Journal of Consumer Behaviour*, *15*(6), 503–515. https://doi.org/10.1002/cb.1587

Jost, P., Kruschinski, S., Sülflow, M., Haßler, J., & Maurer, M. (2022). Invisible transparency: How different types of ad disclaimers on Facebook affect whether and how digital political advertising is perceived. *Policy & Internet*, poi3.333. https://doi.org/10.1002/poi3.333

Jung, A. R., & Heo, J. (2019). Ad Disclosure vs. Ad Recognition: How Persuasion Knowledge Influences Native Advertising Evaluation. *Journal of Interactive Advertising*, *19*(1), 1–14. https://doi.org/10.1080/15252019.2018.1520661

Kapsch, P. H. (2022). Exploring user agency and small acts of algorithm engagement in everyday media use. *Media International Australia*, *183*(1), 16–29. https://doi.org/10.1177/1329878X211067803

Keib, K., Wojdynski, B. W., Espina, C., Malson, J., Jefferson, B., & Lee, Y.-I. (2022). Living at the Speed of Mobile: How Users Evaluate Social Media News Posts on Smartphones. *Communication Research*, *49*(7), 1016–1032. https://doi.org/10.1177/00936502211018542

Kerem, K., & Ulla, M. (2018). Perceived Intrusiveness of Personalized Marketing. *Digital Transformation – Meeting the Challenges*, 147–158. https://doi.org/10.18690/978-961-286-170-4.9

Knowles, E. S., & Linn, J. A. (Eds.). (2004). *Resistance and persuasion*. Lawrence Erlbaum Associates.

König, P. D. (2020). Why Digital-Era Political Marketing is Not the Death Knell for Democracy: On the Importance of Placing Political Microtargeting in the Context of Party Competition. *Statistics, Politics and Policy*, *11*(1), 87–110. https://doi.org/10.1515/spp-2019-0006

Krouwer, S., Poels, K., & Paulussen, S. (2017). To Disguise or to Disclose? The Influence of Disclosure Recognition and Brand Presence on Readers' Responses Toward Native Advertisements in Online News Media. *Journal of Interactive Advertising*, *17*(2), 124–137. https://doi.org/10.1080/15252019.2017.1381579

Kruglanski, A. W., Webster, D. M., & Klem, A. (1993). Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of Personality and Social Psychology*, *65*(5), 861–876. https://doi.org/10.1037/0022-3514.65.5.861

Kruikemeier, S., Sezgin, M., & Boerman, S. C. (2016). Political Microtargeting: Relationship between Personalized Advertising on Facebook and Voters' Responses. *Cyberpsychology, Behavior, and Social Networking*, *19*(6), 367–372. https://doi.org/10.1089/cyber.2015.0652

Kumkale, G. T., Albarracín, D., & Seignourel, P. J. (2010). The Effects of Source Credibility in the Presence or Absence of Prior Attitudes: Implications for the Design of Persuasive Communication Campaigns: SOURCE CREDIBILITY AND PERSUASION. *Journal of Applied Social Psychology*, *40*(6), 1325–1356. https://doi.org/10.1111/j.1559-1816.2010.00620.x

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Lavigne, M. (2021). Strengthening ties: The influence of microtargeting on partisan attitudes and the vote. *Party Politics*, *27*(5), 965–976. https://doi.org/10.1177/1354068820918387

Online Advertising and Social Media (Transparency) Bill, 150–2017 (2021). https://data.oireachtas.ie/ie/oireachtas/bill/2017/150/eng/initiated/b15017d.pdf

Leerssen, P., Ausloos, J., Zarouali, B., Helberger, N., & de Vreese, C. H. (2019). Platform ad archives: Promises and pitfalls. *Internet Policy Review*, *8*(4), 1–21. https://doi.org/10.14763/2019.4.1421

Lubianco, J. (2022, April 12). *Fake News bill gets stuck in Brazilian Congress and it's unlikely to be voted on before the elections; remuneration proposal for journalistic organizations is a sensitive topic*. LatAm Journalism Review by the Knight Center. https://latamjournalismreview.org/articles/fake-news-brazil-payment-journalism/

Madsen, J. K. (2019). *The psychology of micro-targeted election campaigns*. Palgrave Macmillan.

Main, K., Dahl, D., & Darke, P. (2007). Deliberative and Automatic Bases of Suspicion: Empirical Evidence of the Sinister Attribution Error. *Journal of Consumer Psychology*, *17*(1), 59–69. https://doi.org/10.1207/s15327663jcp1701_9

Markus, G. B. (1982). Political Attitudes during an Election Year: A Report on the 1980 NES Panel Study. *American Political Science Review*, *76*(3), 538–560. https://doi.org/10.1017/S0003055400188392

Matthes, J., Hirsch, M., Stubenvoll, M., Binder, A., Kruikemeier, S., Lecheler, S., & Otto, L. (2022). Understanding the democratic role of perceived online political micro-targeting: Longitudinal effects on trust in democracy and political interest. *Journal of Information Technology & Politics*, 1–14. https://doi.org/10.1080/19331681.2021.2016542

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48), 12714–12719. https://doi.org/10.1073/pnas.1710966114

McGuire, W. J. (1961). Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *The Journal of Abnormal and Social Psychology*, *63*(2), 326–332. https://doi.org/10.1037/h0048344

Meier, Y., Meinert, J., & Krämer, N. C. (2021). Investigating factors that affect the adoption of COVID-19 contact-tracing apps: A privacy calculus perspective. *Technology, Mind, and Behavior*, *2*(3), 1–10. https://doi.org/10.1037/tmb0000040

Meier, Y., Schäwel, J., & Krämer, N. C. (2021). Between protection and disclosure: Applying the privacy calculus to investigate the intended use of privacy-protecting tools and self-disclosure on different websites. *Studies in Communication and Media*, *10*(3), 283–306. https://doi.org/10.5771/2192-4007-2021-3-283

Meier, Y., Schäwel, J., Kyewski, E., & C. Krämer, N. (2020). Applying Protection Motivation Theory to Predict Facebook Users' Withdrawal and Disclosure Intentions. *International Conference on Social Media and Society*, 21–29. https://doi.org/10.1145/3400806.3400810

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & Mccann, R. M. (2003). Credibility for the 21st Century: Integrating Perspectives on Source, Message, and Media Credibility in the Contemporary Media Environment. *Annals of the International Communication Association*, *27*(1), 293–335. https://doi.org/10.1080/23808985.2003.11679029

Montoya, R. M., Horton, R. S., Vevea, J. L., Citkowicz, M., & Lauber, E. A. (2017). A re-examination of the mere exposure effect: The influence of repeated exposure on recognition, familiarity, and liking. *Psychological Bulletin*, *143*(5), 459–498. https://doi.org/10.1037/bul0000085

Murray, G. R., & Scime, A. (2010). Microtargeting and electorate segmentation: Data mining the American National Election Studies. *Journal of Political Marketing*, *9*(3), 143–166. https://doi.org/10.1080/15377857.2010.497732

Nelson, M. R., Ham, C. D., & Haley, E. (2021). What Do We Know about Political Advertising? Not Much! Political Persuasion Knowledge and Advertising Skepticism in the United States. *Journal of Current Issues & Research in Advertising*, 1–25. https://doi.org/10.1080/10641734.2021.1925179

Ohanian, R. (1990). Construction and Validation of a Scale to Measure Celebrity Endorsers' Perceived Expertise, Trustworthiness, and Attractiveness. *Journal of Advertising*, *19*(3), 39–52.

Ostrom, T. M. (1969). The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology*, *5*(1), 12–30. https://doi.org/10.1016/0022-1031(69)90003-1

Pan, Y., & Zinkhan, G. M. (2006). Determinants of retail patronage: A meta-analytical perspective. *Journal of Retailing*, *82*(3), 229–243. https://doi.org/10.1016/j.jretai.2005.11.008

Papakyriakopoulos, O., Hegelich, S., Shahrezaye, M., & Serrano, J. C. M. (2018). Social media and microtargeting: Political data processing and the consequences for Germany. *Big Data & Society*, *5*(2), 2053951718811844. https://doi.org/10.1177/2053951718811844

Petty, R. E., & Cacioppo, J. T. (1977). Forewarning, cognitive responding, and resistance to persuasion. *Journal of Personality and Social Psychology*, *35*(9), 645–655. https://doi.org/10.1037/0022-3514.35.9.645

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205. https://doi.org/10.1016/S0065-2601(08)60214-2

Pew Research Center. (2021, April 5). Use of online platforms, apps varies – sometimes widely – by demographic group. *Pew Research Center: Internet, Science & Tech*. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/pi_2021-04-07_social-media_0-03/

Pilditch, T., & Madsen, J. K. (2021). Targeting *Your* Preferences: Modelling Micro-Targeting for an Increasingly Diverse Electorate. *Journal of Artificial Societies and Social Simulation*, *24*(1), 5. https://doi.org/10.18564/jasss.4452

Pornpitakpan, C. (2004). The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x

Quinn, J. M., & Wood, W. (2004). Forewarnings of Influence Appeals: Inducing Resistance and Acceptance. In *Resistance and persuasion.* (pp. 193–213). Lawrence Erlbaum Associates Publishers.

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Richmond, V. P., & McCroskey, J. C. (1975). Whose Opinion Do You Trust? *Journal of Communication*, *25*(3), 42–50. https://doi.org/10.1111/j.1460-2466.1975.tb00604.x

Roozenbeek, J., Van Der Linden, S., & Nygren, T. (2020). Prebunking interventions based on the psychological theory of "inoculation" can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016//mr-2020-008

Rosenberg, M. J., Hovland, C. I., McGuire, W. J., Abelson, R. P., & Brehm, J. W. (1960). *Attitude organization and change: An analysis of consistency among attitude components. (Yales studies in attitude and communication.), Vol. III* (pp. x, 239). Yale Univer. Press.

Rosselli, F., Skelly, J. J., & Mackie, D. M. (1995). Processing Rational and Emotional Messages: The Cognitive and Affective Mediation of Persuasion. *Journal of Experimental Social Psychology*, *31*(2), 163–190. https://doi.org/10.1006/jesp.1995.1008

Rozendaal, E., Lapierre, M. A., Van Reijmersdal, E. A., & Buijzen, M. (2011). Reconsidering Advertising Literacy as a Defense Against Advertising Effects. *Media Psychology*, *14*(4), 333–354. https://doi.org/10.1080/15213269.2011.620540

Sagarin, B. J., & Cialdini, R. B. (2004). Creating Critical Consumers: Motivating Receptivity by Teaching Resistance. In *Resistance and persuasion.* (pp. 259–282). Lawrence Erlbaum Associates Publishers.

Sagarin, B. J., Cialdini, R. B., Rice, W. E., & Serna, S. B. (2002). Dispelling the illusion of invulnerability: The motivations and mechanisms of resistance to persuasion. *Journal of Personality and Social Psychology*, *83*(3), 526–541. https://doi.org/10.1037/0022-3514.83.3.526

Schäwel, J., Frener, R., & Trepte, S. (2021). Political Microtargeting and Online Privacy: A Theoretical Approach to Understanding Users' Privacy Behaviors. *Media and Communication*, *9*(4), 158–169. https://doi.org/10.17645/mac.v9i4.4085

Schwarz, N., Bless, H., & Bohner, G. (1991). Mood and Persuasion: Affective States Influence the Processing of Persuasive Communications. In *Advances in Experimental Social Psychology* (Vol. 24, pp. 161–199). Elsevier. https://doi.org/10.1016/S0065-2601(08)60329-9

See, Y. H. M. (2018). Cognitive-Affective Processing System. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1–7). Springer International Publishing. https://doi.org/10.1007/978-3-319-28099-8_1788-2

Segijn, C. M., Opree, S. J., & Van Ooijen, I. (2022). The validation of the Perceived Surveillance Scale. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *16*(3), Article 3. https://doi.org/10.5817/CP2022-3-9

Segijn, C. M., & Van Ooijen, I. (2022). Differences in consumer knowledge and perceptions of personalized advertising: Comparing online behavioural advertising and synced advertising. *Journal of Marketing Communications*, *28*(2), 207–226. https://doi.org/10.1080/13527266.2020.1857297

Snyder, J. M., & Ting, M. M. (2002). An Informational Rationale for Political Parties. *American Journal of Political Science*, *46*(1), 90. https://doi.org/10.2307/3088416

Sparks, J. R., & Rapp, D. N. (2011). Readers' reliance on source credibility in the service of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 230–247. https://doi.org/10.1037/a0021331

Stubb, C., & Colliander, J. (2019). "This is not sponsored content" – The effects of impartiality disclosure and e-commerce landing pages on consumer responses to social media influencer posts. *Computers in Human Behavior*, *98*, 210–222. https://doi.org/10.1016/j.chb.2019.04.024

Stubb, C., Nyström, A.-G., & Colliander, J. (2019). Influencer marketing: The impact of disclosing sponsorship compensation justification on sponsored content effectiveness. *Journal of Communication Management*, *23*(2), 109–122. https://doi.org/10.1108/JCOM-11-2018-0119

Stubenvoll, M., Binder, A., Noetzel, S., Hirsch, M., & Matthes, J. (2022). Living is Easy With Eyes Closed: Avoidance of Targeted Political Advertising in Response to Privacy Concerns, Perceived Personalization, and Overload. *Communication Research*, 009365022211308. https://doi.org/10.1177/00936502221130840

Sundar, S. S., & Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*.

Susser, D., Roessler, B., & Nissenbaum, H. F. (2018). Online Manipulation: Hidden Influences in a Digital World. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3306006

Tam, K. Y., & Ho, S. Y. (2006). Understanding the Impact of Web Personalization on User Information Processing and Decision Outcomes. *MIS Quarterly*, *30*(4), 865–890. https://doi.org/10.2307/25148757

The jamovi project. (2022). *Jamovi* (2.3) [Computer software]. http://www.jamovi.org

Tormala, Z. L., Briñol, P., & Petty, R. E. (2006). When credibility attacks: The reverse impact of source credibility on persuasion. *Journal of Experimental Social Psychology*, *42*(5), 684–691. https://doi.org/10.1016/j.jesp.2005.10.005

Tormala, Z. L., & Petty, R. E. (2002). What doesn't kill me makes me stronger: The effects of resisting persuasion on attitude certainty. *Journal of Personality and Social Psychology*, *83*(6), 1298–1313. https://doi.org/10.1037/0022-3514.83.6.1298

Tormala, Z. L., & Petty, R. E. (2004). Resistance to Persuasion and Attitude Certainty: The Moderating Role of Elaboration. *Personality and Social Psychology Bulletin*, *30*(11), 1446–1457. https://doi.org/10.1177/0146167204264251

Trepte, S. (2021). The Social Media Privacy Model: Privacy and Communication in the Light of Social Media Affordances. *Communication Theory*, *31*(4), 549–570. https://doi.org/10.1093/ct/qtz035

Van Ooijen, I. (2022). When Disclosures Backfire: Aversive Source Effects for Personalization Disclosures on Less Trusted Platforms. *Journal of Interactive Marketing*, *57*(2), 178–197. https://doi.org/10.1177/10949968221080499

Van Reijmersdal, E. A., Fransen, M. L., van Noort, G., Opree, S. J., Vandeberg, L., Reusch, S., van Lieshout, F., & Boerman, S. C. (2016). Effects of Disclosing Sponsored Content in Blogs: How the Use of Resistance Strategies Mediates Effects on Persuasion. *American Behavioral Scientist*, *60*(12), 1458–1474. https://doi.org/10.1177/0002764216660141

Van Reijmersdal, E. A., Rozendaal, E., & Buijzen, M. (2012). Effects of Prominence, Involvement, and Persuasion Knowledge on Children's Cognitive and Affective Responses to Advergames. *Journal of Interactive Marketing*, *26*(1), 33–42. https://doi.org/10.1016/j.intmar.2011.04.005

Van Reijmersdal, E. A., Rozendaal, E., Hudders, L., Vanwesenbeeck, I., Cauberghe, V., & van Berlo, Z. M. C. (2020). Effects of Disclosing Influencer Marketing in Videos: An Eye Tracking Study Among Children in Early Adolescence. *Journal of Interactive Marketing*, *49*, 94–106. https://doi.org/10.1016/j.intmar.2019.09.001

Van Reijmersdal, E. A., Tutaj, K., & Boerman, S. C. (2013). The effects of brand placement disclosures on skepticism and brand memory. *Communications - The European Journal of Communication Research*, *38*(2). https://doi.org/10.1515/commun-2013-0008

Van Reijmersdal, E. A., Opree, S. J., & Cartwright, R. F. (2021). Brand in focus: Activating adolescents' persuasion knowledge using disclosures for embedded advertising in music videos. *Communications*. https://doi.org/10.1515/commun-2019-0168

Vraga, E., Bode, L., & Troller-Renfree, S. (2016). Beyond Self-Reports: Using Eye Tracking to Measure Topic and Style Differences in Attention to Social Media Content. *Communication Methods and Measures*, *10*(2–3), 149–164. https://doi.org/10.1080/19312458.2016.1150443

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & Van Der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426–432. https://doi.org/10.1037/a0022790

Wilson, E. J., & Sherrell, D. L. (1993). Source effects in communication and persuasion research: A meta-analysis of effect size. *Journal of the Academy of Marketing Science*, *21*(2), 101–112. https://doi.org/10.1007/BF02894421

Winter, S., & Krämer, N. C. (2014). A question of credibility—Effects of source cues and recommendations on information selection on news sites and blogs. *Communications*, *39*(4), 435–456. https://doi.org/10.1515/commun-2014-0020

Winter, S., Maslowska, E., & Vos, A. L. (2021). The effects of trait-based personalization in social media advertising. *Computers in Human Behavior*, *114*. https://doi.org/10.1016/j.chb.2020.106525

Wojdynski, B. W. (2019). Advertorials and Native Advertising. In T. P. Vos, F. Hanusch, D. Dimitrakopoulou, M. Geertsema-Sligh, & A. Sehl (Eds.), *The International Encyclopedia of Journalism Studies* (1st ed., pp. 1–6). Wiley. https://doi.org/10.1002/9781118841570.iejs0062

Wojdynski, B. W., & Evans, N. J. (2016). Going Native: Effects of Disclosure Position and Language on the Recognition and Evaluation of Onlin. *Journal of Advertising*, *45*(2), 157–168. https://doi.org/10.1080/00913367.2015.1115380

Wojdynski, B. W., & Evans, N. J. (2020). The Covert Advertising Recognition and Effects (CARE) model: Processes of persuasion in native advertising and other masked formats. *International Journal of Advertising*, *39*(1), 4–31. https://doi.org/10.1080/02650487.2019.1658438

Wood, M. L. M., Nelson, M. R., Atkinson, L., & Lane, J. B. (2008). Social Utility Theory: Guiding Labeling of VNRs as Ethical and Effective Public Relations. *Journal of*

Public Relations Research, *20*(2), 231–249. https://doi.org/10.1080/10627260801894405

Wood, R. V., Bradac, J. J., Barnhart, S. A., & Kraft, E. (1970). The effect of learning about techniques of propaganda on subsequent reaction to propagandistic communications. *The Speech Teacher*, *19*(1), 49–53. https://doi.org/10.1080/03634527009377792

Wright, P. (1975). Factors Affecting Cognitive Resistance to Advertising. *Journal of Consumer Research*, *2*(1), 1. https://doi.org/10.1086/208610

Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y., & Chen, Z. (2009). How much can behavioral targeting help online advertising? *Proceedings of the 18th International Conference on World Wide Web - WWW '09*, 261. https://doi.org/10.1145/1526709.1526745

Yoon, K., Pinkleton, B. E., & Ko, W. (2005). Effects of Negative Political Advertising on Voting Intention: An Exploration of the Roles of Involvement and Source Credibility in the Development of Voter Cynicism. *Journal of Marketing Communications*, *11*(2), 95–112. https://doi.org/10.1080/1352726042000315423

Youn, S., & Kim, S. (2019). Newsfeed native advertising on Facebook: Young millennials' knowledge, pet peeves, reactance and ad avoidance. *International Journal of Advertising*, *38*(5), 651–683. https://doi.org/10.1080/02650487.2019.1575109

Zarouali, B., Dobber, T., De Pauw, G., & de Vreese, C. (2020). Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media. *Communication Research*. https://doi.org/10.1177/0093650220961965

Zarouali, B., Poels, K., Ponnet, K., & Walrave, M. (2018). "Everything under control?": Privacy control salience influences both critical processing and perceived persuasiveness of targeted advertising among adolescents. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *12*(1). https://doi.org/10.5817/CP2018-1-5

Zarouali, B., Ponnet, K., Walrave, M., & Poels, K. (2017). "Do you like cookies?" Adolescents' skeptical processing of retargeted Facebook-ads and the moderating role of privacy concern and a textual debriefing. *Computers in Human Behavior*, *69*, 157–165. https://doi.org/10.1016/j.chb.2016.11.050

Zarouali, B., Verdoodt, V., Walrave, M., Poels, K., Ponnet, K., & Lievens, E. (2020). Adolescents' advertising literacy and privacy protection strategies in the context of targeted advertising on social networking sites: Implications for regulation. *Young Consumers*, *21*(3), 351–367. https://doi.org/10.1108/YC-04-2020-1122

Zuiderveen Borgesius, F. J., Möller, J., Kruikemeier, S., Fathaigh, R., Irion, K., Dobber, T., Bodo, B., & de Vreese, C. (2018). Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, *14*(1), 82–96. https://doi.org/10.18352/ulr.420

Zuwerink Jacks, J., & Cameron, K. A. (2003). Strategies for Resisting Persuasion. *Basic and Applied Social Psychology*, *25*(2), 145–161. https://doi.org/10.1207/S15324834BASP2502_5

V Research articles

In the following, the four research articles in this dissertation are presented. The order is in correspondence with the order of presentation in the dissertation. Article I has been published in an international journal. Article II has been revised and resubmitted at an international journal. Article III has been accepted for publication at an international journal. Article IV is currently under review at an international journal.

Article I

Jansen, M.-P., & Krämer, N. C. (2023). Empty Transparency? The Effects on Credibility and Trustworthiness of Targeting Disclosure Labels for Micro-Targeted Political Advertisements. *Weizenbaum Journal of the Digital Society*, *3*(1). https://doi.org/10.34669/WI.WJDS/3.1.5

weizenbaum journal\ OF THE DIGITAL SOCIETY

J\03 ISSUE 01

WEIZENBAUM JOURNAL OF THE DIGITAL SOCIETY
Volume 3 \ Issue 1 \ w3.1.5 \ 08-09-2023
ISSN 2748-5625 \ DOI 10.34669/WI.WJDS/3.1.5

Information on this journal and its funding can be found on its website:
https://wjds.weizenbaum-institut.de

KEYWORDS

micro-targeting
disclosure labels
trustworthiness
credibility
persuasion knowledge

RESEARCH PAPER

Empty Transparency?

The Effects on Credibility and Trustworthiness of Targeting Disclosure Labels for Micro-Targeted Political Advertisements

Martin-Pieter Jansen [1*] \ Nicole C. Krämer [1,2]

[1] Social Psychology: Media and Communication, University of Duisburg-Essen
[2] Research Center Trustworthy Data Science and Security
[*] Corresponding author, martin-pieter.jansen@uni-due.de

ABSTRACT

Political micro-targeting describes the use of data to identify members of a target audience and send messages designed to fit their views and resonate with them. The practice has received considerable attention of late, especially around questions of transparency. This study explores one potential solution to this quandary, namely, disclosure labels. Adopting a pre-registered online one-factorial three-group between-subjects experimental design, we have investigated how different types of disclosure labels for micro-targeted advertisements impact source and message credibility, as well as source trustworthiness. Furthermore, we have investigated the potential mediating effect of persuasion knowledge on these effects. We exposed 227 German Facebook users to either a Facebook advertisement without a disclosure label, a sponsored disclosure label, or a targeting disclosure label that stated they were targeted based on their online behavior. The results demonstrate small and non-significant differences between groups regarding source and message credibility and source trustworthiness, with no mediation by persuasion knowledge observed. Additionally, most participants did not recall the disclosure we exposed them

to, potentially explaining these small effects within our sample. In conclusion, our targeting disclosure approaches were insufficiently informative. Hence, we argue that platforms should put more effort into improving transparency for their users than they currently do.

1 Introduction

In recent years, political campaigns have moved with their audiences from face-to-face contact to a focus on digital connections and interactions and relationship development. In the current digital media landscape, among the most common ways that political parties contact potential voters is through social networking sites (SNSs) (Giasson et al., 2019). Parties and political consultancy firms use these platforms to direct their messages to the target groups that they predict to be most susceptible to them. The information parties use to target these individuals is oftentimes a product of their online behavior (Matz et al., 2017). The large volumes of information that platforms gather from usage and the information that users voluntarily share – such as demographics, likes, interests, and location data – enable these targeting activities to be personalized and, in some cases, extremely narrowly targeted (i.e., at small groups of people with the same interests) (Dobber et al., 2017). These messages are developed to resonate with the specific target audience. The demarcation and targeting based on this data describe the concept of *micro-targeting* or, in the case of politics, *political micro-targeting* (PMT) (Kruikemeier et al., 2016; Zarouali et al., 2020; Zuiderveen Borgesius et al., 2018).

In the most generally known illustration of PMT, political consultancy firm Cambridge Analytica allegedly gathered and used data from more than 50 million Facebook users to establish psychological profiles and target users with messages that would persuade them as strongly as possible (Cadwalladr, 2018). According to whistleblowers, this contributed to Donald Trump's presidential victory and the Leave campaign's success in the Brexit referendum (Kaiser, 2019; Wylie, 2019).

Although PMT can mobilize potential voters that might otherwise have been left out, personalized content and the sharing of relevant information (Zuiderveen Borgesius et al., 2018) imply risks for society and democracy. PMT allows senders of messages to contribute to selective information exposure, which sees parties present themselves as single-issue parties to different individuals (with different issues). In such cases, PMT could lead to biased perceptions regarding parties if voters are not aware of other issues that the party focuses on, potentially threatening the marketplace of ideas within our democratic society (Barocas, 2012; Zuiderveen Borgesius et al., 2018). Furthermore, PMT could deliberately ignore certain target groups: Because it is possible to

mobilize voters that are more likely to vote for a certain party, those who would likely vote for the other parties are purposely neglected, decreasing scrutiny of the democratic process (Jamieson, 2013). In turn, this could expand the gap in representation in governments by making those who are targeted more strongly represented (Endres & Kelly, 2018). Finally, PMT could undermine the public sphere by helping to shield potential voters from information or viewpoints that might challenge their beliefs and values (Gorton, 2016, p. 69).

One solution to these risks could be increasing transparency by using disclosure labels. These labels are often used for regular marketing activities – such as advertorials (Boerman & van Reijmersdal, 2016), sponsored vlogs (Van Der Goot et al., 2021), and influencer marketing (van Reijmersdal et al., 2020) – and could represent a chance to inform the public of the nature of advertisements and the targeting that took place. Simultaneously, disclosures could provide users with more information about the party that pays for the advertisement, which could differ from the political party itself. In an example of a new regulatory approach, the EU Digital Services Act states that targeted information and advertisements should include information about when and on whose behalf content is displayed (European Commission, 2022).

In one of the first experiments on the use of disclosure labels on PMT advertisements, Kruikemeier et al. (2016) showed that if individuals notice disclosure labels on a micro-targeted Facebook post, they will better understand that the post is an advertisement. This promotes the activation of *persuasion knowledge* (PK), a mechanism that helps users identify persuasion attempts based on previous exposure and experience and can be influenced by awareness of the advertisement's topic and knowledge of the message's sender (Friestad & Wright, 1994; Jung & Heo, 2019). Furthermore, PK has been shown to effectuate a more critical style of processing a message, which might be problematic for the sender of a message, especially when it is a political party (Boerman & Kruikemeier, 2016; Campbell, 1995; Main et al., 2007; Wentzel et al., 2010).

Building on the work of Kruikemeier et al. (2016), this work investigates the effects of different types of disclosures on source trustworthiness while also investigating the potential mediating role of PK on these effects. However, we move beyond the measures used by Kruikemeier et al. (2016) by also including source and message credibility as dependent variables. Both of these constructs are found to be important predictors in the voting process (Carr & Hayes, 2014; Hetherington, 1999; Housholder & LaMarre, 2014; Madsen, 2019; Main et al., 2007). Furthermore, we simulate PMT differently by letting participants indicate whether they agree with certain statements before exposing them to our conditions, thereby aligning advertisements with their beliefs. In addition, we use a different disclosure label for our targeting disclosure. This label is in line with the disclosure labels Facebook currently uses but includes more salience about the targeting practices being used. Hence, we arrive at our study's central research question:

How does placing targeting disclosures labels above micro-targeted political advertisements impact source credibility, message credibility, and source trustworthiness, and what is the mediating role of persuasion knowledge on these effects?

2 Theoretical Background

2.1 Micro-Targeting

SNSs such as Facebook and Instagram and search engines such as Google or Bing offer considerable marketing potential. Constant usage and online behavior enable users of these sites to help create databases that are perfect for brand promotion (Barbu, 2014). For senders of messages, these databases provide detailed information about what a user likes and dislikes, which messages they are more given to interacting with, and, hence, which messages are potentially more influential (Winter et al., 2021; Yan et al., 2009). Companies and other parties that utilize advertisements can target specific messages at specific target groups where those messages are likely to be more effective, a practice known as *behavioral targeting* (Matz et al., 2017; Yan et al., 2009). At a time when people are more online than ever before, leaving breadcrumbs as they go from website to website, this practice has become more efficient than ever. These breadcrumbs that people unconsciously leave behind include information about their personal lives and their (online) behavior and can sometimes even be used to predict personality traits based on previous interactions with websites and other forms of online content (e.g., Facebook likes) (Kosinski et al., 2013; Matz et al., 2017; Yan et al., 2009; Zarouali et al., 2020). Furthermore, by applying intensive algorithms, senders can automatically cluster users that share attributes by using machine learning to process these new types of data (Papakyriakopoulos et al., 2018). According to Wilson (2017), this is something that could be automated further by having artificial intelligence processes move from personality profiling to specialized content generation and delivery.

Within the political realm, the goal is not to sell products but to sell the story and ideology of a politician or a party to receive votes. With techniques borrowed from commercial companies and marketing agencies, political actors aim to persuade voters that are unsure about their partisanship, voters that are unsatisfied with previously preferred candidates, and potential new voters. Of course, the concept of dividing different types of potential voters and targeting them with certain messages is not new: Before the internet, parties used canvassing strategies that relied on different fliers in different states or for different zip codes (Barbu, 2014; Gandy, 2000; Murray & Scime, 2010). However, the amount of data gathered by platforms and advertisers makes it possible

2.2 Countermeasures

However, instead of focusing on these roadblocks and the sender side of PMT, the SNSs that consultancy firms use to influence their audiences could represent a solution. Researchers use different countermeasures to minimize the potentially harmful effects of PMT described earlier in this work. One starting point for transparency could be the development of *ad archives or ad libraries* that store all of the advertisements that have run on an SNS (including potential targeting measures) stored and make them openly available to the public (Leerssen et al., 2019). This is something that Meta provides with its ad library and CrowdTangle platform.[1] Nonetheless, Meta has received criticism about these tools not being accessible to everybody and the fact that the company itself manages the mechanisms (Elswah & Howard, 2020). Ben-David (2020) has even argued that Meta's sustained control over this public data enables the company to keep citizens away from information until its contents become the past. Other commentators have suggested that until the companies behind SNSs start regulating the political messages that appear on their platforms, the best solution for voters is to rely on themselves, be cautious, and check their information diet, which the aforementioned tools might make possible (Ghosh, 2018).

Among the most transparent approaches to countering these potentially harmful effects – and an approach that seems convenient – involves implementing targeting disclosure labels (Binford et al., 2021; Kruikemeier et al., 2016). Targeting disclosure labels could be a perfect middle-ground between total platform transparency and user self-reliance. Platforms need not give up their precise targeting information – a substantial part of their business models – and users need not invest in becoming so self-reliant. By providing users with information regarding the nature of an advertisement, platforms demonstrate the transparency that seems critical for users to distinguish between different types of messages and advertisements (Amazeen & Wojdynski, 2020; Binford et al., 2021).

Furthermore, when advertisers undertake covert information-collection strategies for targeting purposes, consumers may experience feelings of vulnerability. Informational cues could be used to offset these effects (Aguirre et al., 2015). In recent attempts to be more transparent about the nature of advertisements that look like general content in a newsfeed, some SNSs and search engines have started using disclosure labels on advertisements and other sponsored content to show that a sender paid to place the message in question (Binford et al., 2021; Jung & Heo, 2019; van Reijmersdal et al., 2016). Disclosure labels on advertisements exist in many forms. Within the realm of marketing, more prominent disclosure labels are seemingly more effective. Prominent disclosure labels are those that are more easily seen by users and therefore lead

[1] CrowdTangle is a public insights tool from Facebook that gives people with access (i.e., journalists, researchers, and social media professionals) insights into public content on the platform (Bleakley, n.d.).

to higher levels of recognition that an advertisement is indeed an advertisement (Amazeen & Wojdynski, 2020). The level of prominence has been found to be important because users are less likely to recognize advertisements when impartial disclosure labels are used (i.e., labels that vaguely state something is sponsored, or just use an icon without explanations; Stubb & Colliander, 2019).

Existing research has not observed differences in the length (in seconds) of exposure to a disclosure label (Boerman et al., 2012). The source of a persuasive message is perceived to be more credible when explicit disclosure labels are used, that is, labels with complete and exact descriptions of what is sponsored and by whom. In such cases, users are aware of potential biases of the source and can integrate that information into their perception of the source and counter uncertainty regarding the source's intentions (Carr & Hayes, 2014). Nonetheless, recent research on disclosure labels for advertisements on SNSs shows that disclosure labels are likely to negatively impact the credibility of both the source and the message (Deng et al., 2020). However, these works both investigated disclosures in the context of regular advertising and sponsored content, which does not incorporate the same levels of personalization or utilize user data to show them specific advertisements. This is pertinent because other work has demonstrated that personalization can be perceived as intrusive (Segijn & van Ooijen, 2022) or even creepy (De Keyzer et al., 2022).

Perceived credibility influences how we use and process information from a particular source (Madsen, 2019). Credibility and trust are more important when the subject of discussion is something the receiver of a message has less knowledge about and when they lack access to information regarding that subject. Although politicians are oftentimes considered experts who specialize in a certain subject, their knowledge will be understood as more valid if they are perceived as more credible and trustworthy. Furthermore, trust directly influences the choice of a political candidate (Hetherington, 1999) and increases the intention to vote for them (Housholder & LaMarre, 2014). This means that when a potential voter deems a candidate more trustworthy and capable, they are more likely to vote for them. Therefore, credibility and trustworthiness both represent important constructs, especially in political communication. Furthermore, when communicating via persuasive messages on SNSs, the receiver's recognition of the persuasion attempt is also important. Understanding the persuasive intent of a message might undermine the trustworthiness of the sender because their aims become known (Main et al., 2007). Furthermore, disclosure labels regarding a message's sponsorship influence perceptions of the opinion leader's (e.g., a political actor) credibility (Carr & Hayes, 2014).

Instead of regular disclosure labels, this study focuses on targeting disclosure labels. These disclosure labels provide users with more information about the fact that they have been targeted by a sender of a persuasive message. This is done while remaining consistent with how SNSs design their regular "sponsored" disclosure labels. This study distinguishes between Facebook posts

without disclosure labels, Facebook advertisements or posts, with the regular disclosure labels that the platform currently uses, and posts featuring a targeting disclosure label that provides users with more salient information about the targeting practices that have taken place. Furthermore, because the targeting disclosure label includes more words, users might notice it more. This could increase the central processing of the disclosure label, encouraging better comprehension of the targeting occurring (Kruikemeier et al., 2016; Petty & Cacioppo, 1986). Although the extant studies previously discussed have demonstrated both increases and decreases in credibility via the use of disclosures, we expect that the transparency that disclosure provides about the targeting and personalization taking place (i.e., the use of user data) decreases the credibility of the source. Therefore, we focus on targeting disclosure labels and investigating their effects on the credibility of the source, the credibility of the message, and the trustworthiness of the source. Given multiple definitions of source and message credibility and source trustworthiness exist, we would like to emphasize the interpretations used in this work. We understand source credibility according to the work of Winter and Krämer (2014), who use competence as an operationalization of credibility that "indicates if the sender is *able* to provide valid statements on a topic" (p. 437). Next, we use the work of Appelman and Sundar (2016) to understand message credibility as "an individual's judgment of the veracity of the content of communication" (p. 5). Finally, we interpret source trustworthiness as the honesty of the communicator, following its use in existing research around PMT (Kruikemeier et al., 2016). Taking the above into account, we assume that:

> H1: A micro-targeted political Facebook message with a targeting disclosure label (vs. a regular disclosure label or no disclosure label) will lower (a) the perceived credibility of both the source and the message and (b) the perceived trustworthiness of the source.

2.3 Persuasion Knowledge

Modern internet users are exposed to many advertisements and influential messages. Previous exposure allows them to develop personal beliefs and knowledge about those messages in order to manage them. In the context of classic persuasion attempts, this has encouraged the development of the term PK, defined as personal knowledge and beliefs about the motives and tactics of an advertisement and the advertisement's source (Friestad & Wright, 1994). The development of PK depends on previous exposure to and experiences with advertisements and persuasive messages. Additionally, knowledge of the topic and knowledge of the message's sender (also known as the agent) both influence PK (Friestad & Wright, 1994; Jung & Heo, 2019). In one of the first studies on targeting disclosure labels for PMT messages on Facebook, Kruikemeier et al. (2016) found that if users notice the disclosure labels on a PMT Facebook

post, they better understand that the Facebook post is an advertisement sent by a political party, activating PK. Targeting disclosure labels can be considered an attempt to inform users about the targeting being practiced. Furthermore, a disclosure label will activate PK, which advertisements without disclosure labels are less likely to do (Boerman et al., 2012; Kruikemeier et al., 2016). Based on these findings, we propose our second hypothesis:

> H2: A micro-targeted political Facebook advertisement with a targeting disclosure label (vs. a regular disclosure label or no disclosure label) will activate stronger persuasion knowledge.

PK has also been shown to lead users to a more critical style of processing. This, in turn, influences the sender's evaluations, possibly lessening their perception of sincerity and trustworthiness (Boerman & Kruikemeier, 2016; Campbell, 1995; Main et al., 2007; Wentzel et al., 2010). Targeting transparency is not substantially addressed by existing regulations: There is no limit to the extent of targeting and only limited oversight concerning the information used to target users (Dommett, 2020). Beyond this lack of regulation, a gap in the literature regarding transparency disclosure labels for PMT is apparent. Some authors argue that previous research regarding PMT suffers from mono-theoretical blindness, providing only an overview of regulations or focusing on campaign practices or big data analytics (Bodó et al., 2017; Zuiderveen Borgesius et al., 2018). This has prompted calls for more studies integrating more theoretical concepts and providing more insights into PMT-related questions (Bodó et al., 2017).

In its current form, PMT is a recent development in the context of political advertising and campaigning, and there is limited research concerning solutions that can counter PMT and inform users about micro-targeted messages. Nonetheless, some researchers have acknowledged concerns about the potential harms to democracy and the marketplace of ideas that PMT could ultimately represent (Barocas, 2012; Gorton, 2016; Zuiderveen Borgesius et al., 2018). For example, the already discussed work of Kruikemeier et al. (2016) revealed that respondents appeared to resist personalized content when they noticed a disclosure label. Furthermore, they concluded that the opportunity of personalizing ads to reach possible voters might not always be as beneficial in practice, leading them to call for future work investigating negative implications, such as political content avoidance. Based on these findings, we attempt to further investigate potential differences in message credibility, source credibility, and source trustworthiness associated with the mediating role of PK. Although research on disclosure labels, PK, source trustworthiness, and credibility has been conducted in the context of marketing activities (Carr & Hayes, 2014; Deng et al., 2020; Stubb & Colliander, 2019), it is important to investigate these presumed relationships within the context of PMT. Additionally, because both PK activation and transparency disclosure labels have been demonstrated to decrease credibility and trustworthiness in the regular marketing context, it

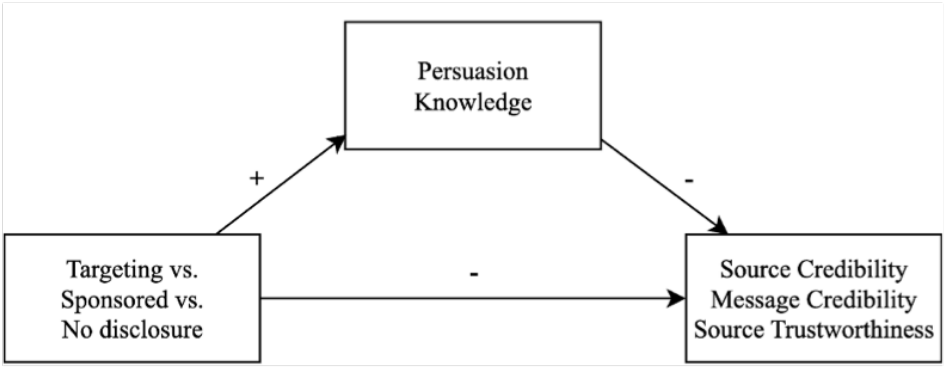
is critical to interrogate the possible existence of these undesirable effects in the political context. Furthermore, research has demonstrated that lower credibility and trustworthiness tend to decrease willingness to vote for a certain party or candidate, making such practices counter-productive for political actors (Hetherington, 1999; Housholder & LaMarre, 2014; Madsen, 2019). Based on these findings, we propose our third hypothesis:

> H3: A micro-targeted political Facebook advertisement with a targeting disclosure label (vs. a regular disclosure label or no disclosure label) will activate stronger persuasion knowledge, negatively impacting (a) source and message credibility and (b) source trustworthiness.

Returning to the work of Kruikemeier et al. (2016), that study saw them examine the relationship between exposure to personalized political ads on Facebook and voter intention to engage in electronic word of mouth and the perceived trustworthiness of those ads. They also investigated the mediating role of PK within this relationship. During their experiment, they exposed respondents ($N = 122$) to either a regular Facebook message, a Facebook advertisement with a disclosure label, or a Facebook message with a disclosure label and an explanation about personalized advertising on Facebook. Within their sample, they found that exposure to a personalized ad from a political party activates PK. Although this, in turn, lowers the intention to engage in electronic word of mouth, this only holds for participants that recall the disclosure label. No effects on source trustworthiness were observed, and adding the text about the practice of personalized advertising did not increase PK or encourage different responses to the message.

Building on the work of Kruikemeier et al. (2016), the current study implements new dependent variables (i.e., source and message credibility), a different PMT simulation, and a different type of targeting disclosure labeling. Where Kruikemeier et al. (2016) used a “sponsored” label as their personalized advertisement condition, the current research will ask participants about their view on climate change regulations and show them a Facebook advertisement that is either for or against these regulations. As such, we try to show participants an advertisement that aligns with their views to simulate PMT. Furthermore, instead of a training condition with an explanation of micro-targeting that may take longer to read, we focus on a potential new *targeting disclosure* label that provides the receiver with more salient information in the form of a short sentence. This corresponds to the disclosure labels that Facebook itself uses and hereby fits the design of the advertisements on that platform.

Figure 1: Proposed Mediation Model



3 Method

The study was approved by the ethics committee of the University of Duisburg-Essen. We pre-registered this study before collecting data: https://osf.io/nbtc4. All participants gave informed consent before participation. Supplementary materials are publicly accessible on OSF (https://osf.io/z4wb8).

3.1 Analysis

To test differences in the effects of a targeting disclosure label (vs. regular disclosure label and no disclosure label), we used the PROCESS macro, model 4, by Hayes (2018) using 5,000 bootstrap samples. Hypothesis 1 represents the total effects, or path c, within the mediation model. Hypothesis 2 represents path a, the first half of the indirect effects, and Hypothesis 3 represents the product of path a and path b, the indirect effects. The mediation analyses were run a total of six times, once with the targeting disclosure label group dummy as an independent variable, the regular disclosure label category as the reference category, and the no disclosure label condition as a covariate (Hayes & Preacher, 2014). The second time the mediation analysis was run with the targeting disclosure label group dummy as an independent variable, the no disclosure label category as the reference category, and the regular disclosure label category as a covariate. Furthermore, the analysis was run separately for each of the three dependent variables (i.e., source credibility, message credibility, and source trustworthiness).

3.2 Sample

We determined the sample size using MedPower by Kenny (2017). Based on the work of Kruikemeier et al. (2016), we assumed a small effect size. We added approximately 17% to the indicated sample size ($N=171$) to arrive at a target of at least $N=200$.

We recruited 280 German Facebook users via the online non-probability access panel of German company Respondi AG (which, following a merger, now pertains to Bilendi & Respondi) during the period September 20–24, 2021. Thirteen respondents did not agree to our informed consent, 13 respondents were not Facebook users, and seven failed to correctly answer our attention check items. Additionally, 20 participants did not finish the experiment, leaving a sample of 227 respondents. Participant age ranged from 20 to 69 years old ($M=43.84$, $SD=14.03$). One hundred and twelve participants identified as female and 115 as male. Regarding education, most participants reported obtaining a secondary school certificate ($n=90$), 57 reported a university entrance qualification, 39 reported a master's degree, 32 reported a bachelor's degree, 28 reported a qualifying middle school diploma, and 15 reported an advanced technical college entrance diploma, 2 indicated having no diploma, 2 indicated a doctorate, and 5 indicated that their educational background differed from all the available responses. Randomization checks showed that the experimental groups did not differ in terms of age ($F(2, 224)=1.79$, $p=0.169$), gender ($\chi^2(2, N=227)=2.71$, $p=.260$) or highest level of education ($\chi^2(16, N=227)=18.01$, $p=.320$).

3.3 Study Design

To test our hypotheses, we conducted an online experiment with a factorial between-subjects design. The experiment entailed three conditions: participants were exposed to either a regular Facebook post without any disclosure label, a regular Facebook advertisement with the disclosure label that Facebook uses on its platform ("Sponsored, paid for by …"), or a Facebook advertisement with a more salient targeting disclosure label ("This content has been targeted at you based on your online behavior").

To at least simulate micro-targeting, we tried to expose participants to advertisements in line with their own beliefs. To accomplish this, we asked the participants to indicate whether they agreed or disagreed with a set of statements on climate change regulations. These statements were pre-tested in a set of eight statements ($N=100$) to determine which statements clearly described the view of someone who is for climate change regulations (the view of someone who wants to counter climate change and could be considered more eco-friendly, hereafter called *pro*) or against climate change regulations (the view of

someone who does not believe in climate change and thinks the regulations to counter this do not work, hereafter called *against*). Four of these pre-tested statements – those most clearly pro or against – were used in the main study.

3.4 Procedure

Of the ten statements used in the main study, four statements were either pro or against climate change regulations (two each). Six statements served to counter testing effects. After answering the statements, participants were randomly assigned to one of the experimental conditions. Depending on whether they agreed or disagreed with the statements, we measured their views on climate change regulations, and participants were shown a Facebook post featuring a statement that aligned with their answers to the previous statements. In this way, we tried to recreate at least a degree of (micro) targeting by showing them something a well-developed algorithm might also show them. After exposure to our stimuli, the participants answered the questions regarding our variables.

Thereafter, we asked participants to state the degree to which they found the advertisements targeted and in line with their beliefs regarding climate change regulations. Finally, we asked for demographic information (see the measures section) before debriefing and thanking the participants. The average completion time was about 8 minutes.

3.5 Stimuli

We created a total of six Facebook advertisements for a non-existent male politician from a non-existent political party (see Figure 2). For each experimental condition, one Facebook post was developed: a Facebook post with no disclosure label, a Facebook post with a regular disclosure label (“sponsored”), and a Facebook post with a targeting disclosure label. We developed one of each of these posts for pro participants and one of each of these posts for against participants. Other than the disclosure labels and the statement used in the advertisement, all stimuli were identical (i.e., in terms of the number of likes and number of comments). On Facebook, a regular disclosure label is normally placed in the same position as the time of posting on a non-sponsored post. The chosen posting time for the no disclosure label condition was “Yesterday, 09:32.” This was chosen because we did not have to change the date based on when the participant took part in the experiment and because the time is during office hours. We used the same locations that Facebook uses for our disclosure labels to increase the study’s external validity.

Figure 2: Examples of the Sponsored Disclosure Label, the No-Disclosure Label, and the Total Post Including the Targeting Disclosure Label as Used in the Experiment (in German)

3.6 Measures

All measured constructs were tested regarding their factor validity in confirmatory factor analysis. The output of this analysis and an overview of our items can be found as online supplementary materials and on OSF (https://osf.io/z4wb8). The factor analysis revealed a near-adequate model fit for all indicators (Alavi et al., 2020; Hair et al., 2014): $\chi^2(98)=256.03$, $p<.001$, $\chi^2/df=2.61$, CFI=0.95, TLI=0.94, RMSEA = .08, 90% CI [.07, .10], SRMR=.04. (CFI=Comparative fit index; TLI=Tucker-Lewis index; RMSEA=root-mean-square error of approximation; SRMR=standardized root-mean-square residual; CI=confidence interval).

We measured the mediator PK using five items developed by Kruikemeier et al. (2016) based on Ham et al. (2015) that have been validated in the context of prior research on micro-targeting. The items (e.g., "The post feels like an ad") were measured on a seven-point Likert scale ranging from 1 (= strongly disagree) to 7 (= strongly agree). The mean score of these items was calculated and used as a measure for PK (Cronbach's $\alpha = .83$, McDonald's $\omega = .84$, Average Variance Extracted = .53).

In addition, following Winter and Krämer (2014), we measured the dependent variable source credibility using three items with five-point semantic scales to establish a credibility score (ranging from, for example, "competent" to "incompetent"). Here, the mean score of these items was calculated and used as a measure for source credibility (Cronbach's $\alpha = .93$, McDonald's $\omega = .93$, Average Variance Extracted = .82).

Furthermore, we measured message credibility using three items developed by Appelman and Sundar (2016). Participants were asked to assess how well the adjectives "accurate," "authentic," and "believable" described the content they had just read using a seven-point Likert scale ranging from 1 (= describes very poorly) to 7 (= describes very well). The mean score of these items was calculated and used to measure message credibility (Cronbach's $\alpha = .94$, McDonald's $\omega = .94$, Average Variance Extracted = .84).

Finally, following Ohanian (1990) and, more recently, the PMT research of Kruikemeier et al. (2016), we measured source trustworthiness via four items using seven-point semantic differential scales to establish a trustworthiness score (ranging from, for example, "dishonest" to "honest"). Due to an overlap in translation into German, the item for "dependable–undependable" was not measured in this study. We have extended the scale with one item used by Winter and Krämer (2014), namely, "sincere–insincere," which is also measured on a seven-point semantic differential scale. Again, the mean score of these items was calculated and used to measure source trustworthiness (Cronbach's $\alpha = .93$, McDonald's $\omega = .93$, Average Variance Extracted = .74).

As a manipulation check, we asked participants whether they recalled seeing the disclosure label ("On the Facebook post, there was a disclosure message about it being sponsored"), seeing the targeting disclosure label ("On the Facebook post, there was a disclosure message about it being tailored to my online behavior"), or not seeing any disclosure label ("There was no disclosure message on the Facebook post"). We also included several control variables and asked participants for demographic information (age, gender, and highest level of education).

Finally, the pre-registered variable *perceived targeting* was measured by two items: "I think the Facebook message was tailored to me" and "The Facebook message matched with my personal views." These items were dichotomously measured due to a mistake in data collection. However, the variable showed no variance and was therefore omitted from the analyses.

4 Results

This study's focus concerned how positioning disclosure labels above a micro-targeted political Facebook post impacts source and message credibility and source trustworthiness. Furthermore, we investigated the mediating role of PK on these direct effects. Our preliminary analyses are publicly accessible on OSF (https://osf.io/zf83k). The bivariate correlations, means, and standard deviations for our measured variables appear in Table 1.

Table 1: Means, Standard Deviations, and Bivariate Correlations of the Measured Constructs

Measured construct	***M (SD)***	**1**	**2**	**3**
1 Persuasion Knowledge	4.02 (1.12)	-		
2 Source Credibility	3.11 (1.04)	-0.05	-	
3 Message Credibility	4.33 (1.51)	-0.07	0.73***	-
4 Source Trustworthiness	4.23 (1.40)	-0.07	0.82***	0.81***

Note. $*p < .05$, $**p < .01$, $***p < .001$.

According to our manipulation check, a substantial proportion of participants did not recall the correct disclosure (see Table 2). Because the assumption for normality could not be met and the sub-sample sizes for the groups (including recollection) were not close to equal, we compared the differences in the means of our groups for the mediator and the dependent variables between the participants who recalled the label correctly and the participants who did not. This involved using the Mann-Whitney U test for PK ($U = 6012$, $p = .962$), message credibility ($U = 5638.5$, $p = .405$), source credibility ($U = 5587.5$, $p = .345$), and source trustworthiness ($U = 5796.5$, $p = .617$). After comparing the groups, we found no significant differences and chose to keep our whole sample for further analyses.

Means and standard deviations appear in Table 3, and means and standard deviations for participants that correctly recalled the label can be found in the online supplementary materials.

Table 2: Results of the Manipulation Check

Condition	*n*	Recall	Percentage
Control (no disclosure)	76	54	71
Sponsored disclosure	75	18	24
Targeting disclosure	76	13	17

Table 3: Mean Scores (with Standard Deviations Between Parentheses)

Measure	No disclosure (n=76)	Sponsored disclosure (n=75)	Targeting disclosure (n=76)
Persuasion Knowledge	4.09 (1.16)	3.94 (1.14)	4.02 (1.07)
Source Credibility	3.00 (1.06)	3.14 (1.07)	3.18 (1.00)
Message Credibility	4.35 (1.53)	4.41 (1.57)	4.23 (1.44)
Source Trustworthiness	4.21 (1.40)	4.24 (1.53)	4.23 (1.27)

Note. N=227. All constructs were measured on a 7-point Likert or semantic scale.

Contrary to our hypothesis (H_1), there were only small – but non-significant – increases in (H_{1a}) source credibility for the targeting disclosure label group compared to the regular disclosure label group (b=0.04, t(224)=0.27, p=.788) and the no disclosure label group (b=0.18, t(224) = 0.17, p=.276) when controlling for the other experimental group. Furthermore, there were only minor – and non-significant – reductions in message credibility for the targeting disclosure label group compared to the regular disclosure label group (b=-0.18, t(224)=-0.73, p=.463) and the no disclosure label group (b=-0.12, t(224)=-0.48, p=.630) when controlling for the other experimental group. Additionally, there were no reductions in (H_{1b}) source trustworthiness for the targeting disclosure label group compared to the regular disclosure label group (b=-0.01, t(224)=-0.04, p=.971) and almost no increases in the targeting disclosure label compared to the no disclosure label group (b=0.02, t(224)=0.10, p=.917) when controlling for the other experimental group.

Regarding our second hypothesis, we found small – but non-significant – increases in persuasion knowledge (H_2) for the targeting disclosure label group compared to the regular disclosure label group (b=0.08, t(224)=0.44, p=.663). By contrast, we observed a small – but non-significant – decrease in PK for the targeting disclosure label group compared to the no disclosure label group (b=-0.07, t(224)=-0.39, p=.697) when controlling for the other experimental group.

To test the third hypothesis, we considered the mediation – or indirect effects (path a * path b) – within the analysis conducted using PROCESS Macro, which included bootstrapping with 1,000 samples (Hayes, 2018). Contrary to what we hypothesized, but aligning with the results of our first two hypotheses, we found no results that would lead us to accept the hypothesis. Results of the mediation analyses appear online (https://osf.io/xv7kt). Notably, although we observed no direct effects, we recognize that this does not always mean that there are no indirect effects (Hayes, 2009, p. 413). Instead, it can indicate the possibility that an unobserved variable might have influenced the model (cf. Bullock et al., 2010, p. 551). A summary of our findings appears in Table 4.

Table 4: Summary of Findings

Hypothesis		Testing Result
H_{1a}	A micro-targeted political Facebook message with a targeting disclosure label (vs. a regular disclosure label or no disclosure label) will lower the perceived credibility of both the source and the message.	Rejected
H_{1b}	A micro-targeted political Facebook message with a targeting disclosure label (vs. a regular disclosure label or no disclosure label) will lower the perceived trustworthiness of the source.	Rejected
H_{2}	A micro-targeted political Facebook advertisement with a targeting disclosure label (vs. a regular disclosure label or no disclosure label) will activate stronger persuasion knowledge.	Rejected
H_{3a}	A micro-targeted political Facebook advertisement with a targeting disclosure label (vs. a regular disclosure label or no disclosure label) will activate stronger persuasion knowledge, negatively impacting source and message credibility.	Rejected
H_{3b}	A micro-targeted political Facebook advertisement with a targeting disclosure label (vs. a regular disclosure label or no disclosure label) will activate stronger persuasion knowledge, negatively impacting source trustworthiness.	Rejected

4.1 Additional Analyses

After conducting the pre-registered analyses, our results demonstrated that there were no significant effects and that none of our hypotheses could be supported. To further investigate our findings, we performed additional Bayesian analyses using JASP (JASP Team, 2021). By adding Bayesian hypothesis testing, the probability of the observed data given the null hypothesis (H_0) is compared to the probability of the observed data given the alternative hypothesis (H_1) (Wagenmakers, 2007). The Bayes Factor (BF_{01}) is a ratio of these probabilities and is commonly interpreted as the weight of evidence in support of the null versus the alternative hypothesis. Our interpretation of Bayes Factors will rely on Wagenmakers et al. (2011).

For our first hypothesis, we conducted three Bayesian ANOVAs for our dependent variables with exposure to the targeting label as a fixed factor. Concerning source credibility, the analysis showed substantial evidence for H_0 compared to H_1 (BF01 = 4.875), favoring the absence of the effect of the targeting disclosure label on source credibility. Next, regarding message credibility, the analysis showed substantial evidence for H_0 compared to H_1 (BF_{01} = 5.184), favoring the absence of any effect of the targeting disclosure label on message credibility. Next, concerning source trustworthiness, the analysis showed substantial evidence for H_0 compared to H_1 (BF_{01} = 6.529), favoring the absence of the effect of the targeting disclosure label on source trustworthiness. For our second hypothesis, we also conducted a Bayesian ANOVA for our proposed mediator (PK) with exposure to the targeting label as a fixed factor. For PK, this ANOVA showed substantial evidence for H_0 compared to H_1 (BF_{01} = 6.532), favoring the absence of any effect of the targeting disclosure label on message credibility.

Our third hypothesis proposed a stronger mediation effect for exposure to a targeting disclosure label compared to exposure to a regular disclosure label or no exposure to any disclosure label. However, because there is substantial evidence for the absence of any effect of the targeting label on persuasion knowledge, we cannot investigate a possible mediation effect using Bayesian analysis.

5 Discussion

This study aimed to investigate how targeting disclosure labels on micro-targeted political advertisements on Facebook impact source credibility, message credibility, and source trustworthiness. Furthermore, we investigated the mediating role of PK on these direct effects. To contribute to the existing body of research, we incorporated source and message credibility as well as source trustworthiness as dependent variables. Additionally, we exposed participants to statements that aligned with their views in an attempt to simulate micro-targeting.

When considering the direct effects of our manipulation on our dependent variables, we found minor and non-significant differences between the three experimental groups for our dependent variables. Exposure to the targeting disclosure label led to minor increases in source credibility. As assumed, exposure to a targeting disclosure label led to small decreases in message credibility. Contradicting our expectations regarding source trustworthiness, the lowest mean was observed for the no disclosure label group, with the means for the other two groups almost identical. Ultimately, however, we rejected our first hypothesis because no significant differences were found. For our second hypothesis, supporting the findings of Kruikemeier et al. (2016), we observed no differences in the activation of PK between our experimental conditions. Additionally, our third hypothesis did not hold true: We observed no mediating effect of PK on the impact of our manipulations on source credibility, message credibility, and source trustworthiness. This can be explained by the fact that we did not observe any direct effect of those manipulations on these dependent variables. Additionally, we investigated our data using Bayesian analysis to further investigate our initial findings (Wagenmakers, 2007). In conclusion, all our analyses regarding our first two hypotheses yielded results favoring the absence of any of our proposed effects. We were not able to test our third hypothesis due to the evidence favoring the absence of any effect of our manipulations on either the proposed mediating variable or the dependent variables.

The most consistent observation throughout our findings is the fact that participants seemed not to notice the disclosure labels. This aligns with other research on disclosures that has similarly noted that not all participants recalled the disclosure messages (Evans et al., 2017; Kruikemeier et al., 2016; van Reijmersdal et al., 2021; Wojdynski & Evans, 2016). However, much of the previous work recorded higher levels of disclosure recollection than observed in the current study. Although we tried to manipulate the disclosure labels in a manner resembling Facebook's approach to achieve ecological validity, we recognize that, in hindsight, this meant we developed seemingly too-subtle disclosures. As discussed, selling advertising space on the platform is a large part of Facebook's business model, and we wanted to make sure we at least tried to design a disclosure label that could be implemented without disturbing the layout of the advertisements or posts. Our findings suggest that our targeting disclosure labels were too subtle to be recognized by participants, precluding their contribution to the recognition of advertisements or (in turn) to the transparency of targeting procedures. Furthermore, participants were exposed to the disclosure labels in an experimental setting in which they saw only one advertisement. In a real-world setting, they would be exposed to multiple posts and advertisements when browsing their Facebook timeline, making the recall of disclosure labels potentially even harder. Whatever the case, our disclosure labels were ultimately too subtle to even be recognized by participants.

5.1 Implications

Platforms should investigate the disclosures they use because disclosure messages are sometimes not recalled correctly or even noticed. Although our goal was not to evaluate the effectiveness of Facebook's disclosure practices, we now clearly see that it uses disclosure labels that are too subtle (cf. Boerman & Kruikemeier, 2016; Kruikemeier et al., 2016) and that do not get noticed substantially enough to be encoded in a manner that increases transparency (cf. Binford et al., 2021). Although other researchers have demonstrated promising results concerning the recall of information from disclosures, the most significant prerequisite for this is that disclosures are actually noticed in the first place (Binford et al., 2021; van Reijmersdal et al., 2016). Furthermore, the prominence of the disclosure seems important (Amazeen & Wojdynski, 2020; Boerman et al., 2015), which could imply that although platforms try to keep everything within their own corporate layout, they do not yet do enough to inform their audiences.

One of the implications of our findings concerns the processing of cues using a heuristic route, as explained in the Elaboration Likelihood Model (Petty & Cacioppo, 1986). When people notice a disclosure label, they are more likely to know that the message has been designed to persuade them, meaning that the central processing of arguments might be activated. This, in turn, can lead to a more critical evaluation of both the sender and the message. In the current setting, a user's focus might be on the Facebook post itself instead of the disclosure label used, which would also explain the low number of correct recalls observed in our study, even after using a timer to ensure participants viewed the manipulation for at least 30 seconds. Although it makes sense for users to focus on the content itself, disclosure labels are used to increase transparency and inform them about the nature of the content they consume, making it, again, important for platforms to ensure disclosures are more prominent and (therefore) more likely to be seen. Disclosures should be used to inform users instead of being used for the sake of being used.

Users are exposed to large amounts of content on SNSs. Some of this content is posts or pictures of friends and messages from groups they follow. Meanwhile, some content is advertising from companies and political parties. Although most SNS users have been exposed to persuasive messages on these platforms in the past, it can still be hard to distinguish between ads and actual user content (Wojdynski & Evans, 2016). PK helps users manage these persuasive attempts and we maintain – partly based on the current body of work regarding disclosures and their effectiveness – that using disclosure labels can give users a better chance of recognizing these attempts and better distinguishing between tailored and non-tailored advertising.

5.2 Limitations and Future Work

The current study's biggest limitation was participants' lack of recall of our manipulation disclosures. Although we aimed to achieve high levels of ecological validity by using Facebook's current sponsored disclosure format and developing targeting disclosure messages to resemble it, we recognize that we might have overamplified that idea to the point that our disclosures were too subtle to be noticed by our participants. Therefore, the current body of work would benefit substantially from more research concerning the design of disclosures, especially the factors that can improve recognition of disclosures. At the same time, we recognize that most studies measure recall or perform manipulation checks using self-report approaches rather than including a measurement directly after exposure to the stimuli, potentially causing users to check the wrong option. This is possible for participants in all experimental conditions. One solution would be to ask participants about the disclosures on the post directly after exposure and treat the recall question as a measure and not just a manipulation check. Another option would be to use eye-tracking. Using eye-tracking would enable researchers to not only measure the visual fixation on the disclosures but – if combined with the recall question – also assess potential differences in gazing behavior and recall. This is because looking at a disclosure might not be the same as internalizing it and understanding its meaning.

Additionally, we also recognize that we did not include the information that the targeted post was an advertisement in the disclosure for our targeting condition. We wanted to keep the disclosure short, so we did not include this information. In hindsight, it is possible that there was no difference between our experimental groups in terms of PK because we did not provide that information to our participants. Furthermore, we recognize that content being targeted does not also necessarily mean that it is sponsored. For instance, Facebook also shows users content that they are more likely to engage with based on their past interests and interactions. However, we did state that the content was sponsored in the "sponsored" condition, but we did not find any effects in that group compared to the control condition without a disclosure.

Meanwhile, although we measured the credibility of the message and the source, we only considered the trustworthiness of the source, ignoring the trustworthiness of the message. This might represent an opportunity for future work. In hindsight, if we had included a measure for this, we would have been able to provide a clearer overview of potential differences in perceptions between the message and the source based on different disclosure labels.

This study was conducted around the time of the 2021 German general elections. As such, the opinions of participants on the statements used for the micro-targeting simulation might have been more explicit than usual. Nonetheless, because the sample comprised only German citizens, these potential effects would have been identical across the whole sample. Within the manipulation we constructed, we made use of a made-up German political party

("Deutschland zusammen"). We did this to ensure that no participants were influenced by previous exposure to an advertisement for a real political party during the period around the general elections. We did this to protect our participants from any possible persuasion during our experiment, but we also recognize that this possibly influenced our findings. Although participants viewed the Facebook post for a minimum of 30 seconds, a novelty effect might have taken place. We tried to match the post to the views of the participant, which might have interested them in the political party and made them more focused on the name and profile picture of the party, undermining the possibility of them paying attention to the manipulated disclosure labels. When considering the mean score for source credibility, it is overall lower than the other measured constructs – this might imply that people were thinking too much about the fake political party and questioning its credibility, instead of paying attention to the whole Facebook post.

Because PMT in its current form is a rather new phenomenon, the actual practices of consulting firms, political parties, and SNSs remain subject to debate. Furthermore, the influence of PMT on specific political instances (e.g., Trump's campaign, the Brexit campaign) continues to be questioned. These inquiries are indeed valid, and it seems critical to try to inform the public about the practices that allegedly take place on the SNSs where they consume a considerable volume of information, especially given the potential threat to democracy that PMT might represent. Additionally, we maintain that – because they have been found to predict voting behavior – credibility and trustworthiness are important constructs for investigations into PMT. That is, although this study's findings might differ from our expectations, we remain convinced that it is necessary to develop countermeasures that can be implemented in ways that are acceptable for SNSs, and we encourage future researchers to keep that in mind while exploring novel implementations.

6 Conclusion

Considering as a mediating factor PK, a mechanism that has been studied in depth in the context of consumer research, this work has studied the effects on credibility and trustworthiness of informing users about the targeting practices taking place on SNSs. Contrary to expectations, we did not observe any effects of our manipulations on credibility or trustworthiness. Supporting extant research, a large part of our sample did not recall seeing any disclosure label, regardless of the manipulation they were exposed to. Although this is an interesting result, it introduced complications around measuring the actual effects of the disclosure labels. This does represent a problem, but we would like to emphasize that it is also the largest takeaway from this study, especially given that one disclosure label resembled Facebook's standard practice and another disclosure label moved beyond this by including additional information and making the disclosure more salient. Platforms use disclosures as a transparency measure to inform users about advertising practices or to at least try to or appear to be trying to. However, if users do not notice disclosure labels in the current format, this might be perceived to be an insufficient effort, window-dressing, or even empty transparency.

References

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing*, 91(1), 34–49. https://doi.org/10.1016/j.jretai.2014.09.005

Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. (2020). Chi-square for model fit in confirmatory factor analysis. *Journal of Advanced Nursing*, 76(9), 2209–2211. https://doi.org/10.1111/jan.14399

Amazeen, M. A., & Wojdynski, B. W. (2020). The effects of disclosure format on native advertising recognition and audience perceptions of legacy and online news publishers. *Journalism*, 21(12), 1965–1984. https://doi.org/10.1177/1464884918754829

Appelman, A., & Sundar, S. S. (2016). Measuring message credibility: Construction and validation of an exclusive scale. *Journalism and Mass Communication Quarterly*, 93(1), 59–79. https://doi.org/10.1177/1077699015606057

Barbu, O. (2014). Advertising, Microtargeting and Social Media. *Procedia – Social and Behavioral Sciences*, 163, 44–49. https://doi.org/10.1016/j.sbspro.2014.12.284

Barocas, S. (2012). The price of precision: Voter microtargeting and its potential harms to the democratic process. *Proceedings of the First Edition Workshop on Politics, Elections and Data – PLEAD '12*, 31. https://doi.org/10.1145/2389661.2389671

Ben-David, A. (2020). Counter-archiving Facebook. *European Journal of Communication*, 35(3), 249–264. https://doi.org/10.1177/0267323120922069

Binford, M. T., Wojdynski, B. W., Lee, Y.-I., Sun, S., & Briscoe, A. (2021). Invisible transparency: Visual attention to disclosures and source recognition in Facebook political advertising. *Journal of Information Technology & Politics*, 18(1), 70–83. https://doi.org/10.1080/19331681.2020.1805388

Bodó, B., Helberger, N., & De Vreese, C. H. (2017). Political micro-targeting: A manchurian candidate or just a dark horse? *Internet Policy Review*, 6(4). https://doi.org/10.14763/2017.4.776

Boerman, S. C., & Kruikemeier, S. (2016). Consumer responses to promoted tweets sent by brands and political parties. *Computers in Human Behavior*, 65, 285–294. https://doi.org/10.1016/j.chb.2016.08.033

Boerman, S. C., & van Reijmersdal, E. A. (2016). Informing Consumers about "Hidden" Advertising: A Literature Review of the Effects of Disclosing Sponsored Content. In P. De Pelsmacker (Ed.), *Advertising in New Formats and Media* (pp. 115–146). Emerald Group Publishing Limited. https://doi.org/10.1108/978-1-78560-313-620151005

Boerman, S. C., van Reijmersdal, E. A., & Neijens, P. C. (2012). Sponsorship Disclosure: Effects of Duration on Persuasion Knowledge and Brand Responses. *Journal of Communication*, 62(6), 1047–1064. https://doi.org/10.1111/j.1460-2466.2012.01677.x

Boerman, S. C., van Reijmersdal, E. A., & Neijens, P. C. (2015). Using Eye Tracking to Understand the Effects of Brand Placement Disclosure Types in Television Programs. *Journal of Advertising*, 44(3), 196–207. https://doi.org/10.1080/00913367.2014.967423

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology*, 98(4), 550–558. https://doi.org/10.1037/a0018933

Cadwalladr, C. (2018, March 18). 'I made Steve Bannon's psychological warfare tool': Meet the data war whistleblower. *The Guardian*. https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump

Campbell, M. (1995). When Attention-Getting Advertising Tactics Elicit Consumer Inferences of Manipulative Intent: The Importance of Balancing Benefits and Investments. *Journal of Consumer Psychology*, 4(3), 225–254. https://doi.org/10.1207/s15327663jcp0403_02

Carr, C. T., & Hayes, R. A. (2014). The Effect of Disclosure of Third-Party Influence on an Opinion Leader's Credibility and Electronic Word of Mouth in Two-Step Flow. *Journal of Interactive Advertising*, 14(1), 38–50. https://doi.org/10.1080/15252019.2014.909296

Coppock, A., Green, D. P., & Porter, E. (2022). Does digital advertising affect vote choice? Evidence from a randomized field experiment. *Research & Politics*, 9(1), 205316802210769. https://doi.org/10.1177/20531680221076901

De Keyzer, F., van Noort, G., & Kruikemeier, S. (2022). GOING TOO FAR? HOW CONSUMERS RESPOND TO PERSONALIZED ADVERTISING FROM DIFFERENT SOURCES. *Journal of Electronic Commerce Research*, 23(3), 22.

Deng, X., Li, M., & Suh, A. (2020). Recommendation or advertisement? The influence of advertising-disclosure language with pictorial types on influencer credibility and consumers' brand attitudes. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12427 LNCS, 234–248. https://doi.org/10.1007/978-3-030-60152-2_19

Dobber, T., Trilling, D., Helberger, N., & de Vreese, C. H. (2017). Two crates of beer and 40 pizzas: The adoption of innovative political behavioural targeting techniques. *Internet Policy Review*, 6(4). https://doi.org/10.14763/2017.4.777

Dommett, K. (2020). Regulating Digital Campaigning: The Need for Precision in Calls for Transparency. *Policy and Internet*, 12(4), 432–449. https://doi.org/10.1002/poi3.234

Elswah, M., & Howard, P. N. (2020). *Tunisia-memo-English.pdf* [PDF]. The Challenges of Monitoring Social Media in the Arab World: The Case of the 2019 Tunisian Elections. https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2020/03/Tunisia-memo-English.pdf

Endres, K., & Kelly, K. J. (2018). Does microtargeting matter? Campaign contact strategies and young voters. *Journal of Elections*, Public Opinion and Parties, 28(1), 1–18. https://doi.org/10.1080/17457289.2017.1378222

European Commission. (2022). *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL* on a Single Market For Digital Services (Digital Services Act) and amending Directive. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN

Evans, N. J., Phua, J., Lim, J., & Jun, H. (2017). Disclosing Instagram Influencer Advertising: The Effects of Disclosure Language on Advertising Recognition, Attitudes, and Behavioral Intent. *Journal of Interactive Advertising*, 17(2), 138–149. https://doi.org/10.1080/15252019.2017.1366885

Friestad, M., & Wright, P. (1994). *The Persuasion Knowledge Model: How People Cope with Persuasion Attempts*. https://academic.oup.com/jcr/article/21/1/1/1853712

Gandy, O. H. (2000). Dividing Practices: Segmentation and Targeting in the Emerging Public Sphere. In W. L. Bennett & R. M. Entman (Eds.), *Mediated Politics* (1st ed., pp. 141–159). Cambridge University Press. https://doi.org/10.1017/CBO9780511613852.008

Ghosh, D. (2018, October 4). *What is microtargeting and what is it doing in our politics?* Mozilla. https://blog.mozilla.org/internetcitizen/2018/10/04/microtargeting-dipayan-ghosh/

Giasson, T., Le Bars, G., & Dubois, P. (2019). Is Social Media Transforming Canadian Electioneering? Hybridity and Online Partisan Strategies in the 2012 Quebec Election. *Canadian Journal of Political Science*, 52(2), 323–341. https://doi.org/10.1017/S0008423918000902

Gorton, W. A. (2016). Manipulating citizens: How political campaigns' use of behavioral social science harms democracy. *New Political Science*, 38(1), 61–80. https://doi.org/10.1080/07393148.2015.1125119

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7. ed., Pearson new internat. ed). Pearson.

Ham, C. D., Nelson, M. R., & Das, S. (2015). *How to measure persuasion knowledge. International Journal of Advertising*, 34(1), 17–53. https://doi.org/10.1080/02650487.2014.994730

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium. *Communication Monographs*, 76(4), 408–420. https://doi.org/10.1080/03637750903310360

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (Second edition). Guilford Press.

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67(3), 451–470. https://doi.org/10.1111/bmsp.12028

Hetherington, M. J. (1999). The Effect of Political Trust on the Presidential Vote, 1968–96. *American Political Science Review*, 93(2), 311–326. https://doi.org/10.2307/2585398

Housholder, E. E., & LaMarre, H. L. (2014). Facebook Politics: Toward a Process Model for Achieving Political Source Credibility Through Social Media. *Journal of Information Technology & Politics*, 11(4), 368–382. https://doi.org/10.1080/19331681.2014.951753

Jamieson, K. H. (2013). Messages, micro-targeting, and new media technologies. Forum (Germany), 11(3), 429–435. https://doi.org/10.1515/for-2013-0052

JASP Team. (2021). *JASP* (0.14.1).

Jung, A. R., & Heo, J. (2019). Ad Disclosure vs. Ad Recognition: How Persuasion Knowledge Influences Native Advertising Evaluation. *Journal of Interactive Advertising*, 19(1), 1–14. https://doi.org/10.1080/15252019.2018.1520661

Kaiser, B. (2019). *Targeted: The Cambridge Analytica whistleblower's inside story of how big data, Trump, and Facebook broke democracy and how it can happen again.*

Kenny, D. A. (2017). *MedPower: An interactive tool for the estimation of power in tests of mediation*. https://davidakenny.shinyapps.io/MedPower/

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Kruikemeier, S., Sezgin, M., & Boerman, S. C. (2016). Political Microtargeting: Relationship between Personalized Advertising on Facebook and Voters' Responses. *Cyberpsychology, Behavior, and Social Networking*, 19(6), 367–372. https://doi.org/10.1089/cyber.2015.0652

Leerssen, P., Ausloos, J., Zarouali, B., Helberger, N., & de Vreese, C. H. (2019). Platform ad archives: Promises and pitfalls. *Internet Policy Review*, 8(4), 1–21. https://doi.org/10.14763/2019.4.1421

Liberini, F., Redoano, M., Russo, A., Cuevas, Á., & Cuevas, R. (2020). Politics in the Facebook Era – Evidence from the 2016 US Presidential Elections. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3584086

Madsen, J. K. (2019). *The psychology of micro-targeted election campaigns*. Palgrave Macmillan.

Main, K., Dahl, D., & Darke, P. (2007). Deliberative and Automatic Bases of Suspicion: Empirical Evidence of the Sinister Attribution Error. *Journal of Consumer Psychology*, 17(1), 59–69. https://doi.org/10.1207/s15327663jcp1701_9

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences of the United States of America*, 114(48), 12714–12719. https://doi.org/10.1073/pnas.1710966114

Murray, G. R., & Scime, A. (2010). Microtargeting and electorate segmentation: Data mining the American National Election Studies. *Journal of Political Marketing*, 9(3), 143–166. https://doi.org/10.1080/15377857.2010.497732

Ohanian, R. (1990). Construction and Validation of a Scale to Measure Celebrity Endorsers' Perceived Expertise, Trustworthiness, and Attractiveness. *Journal of Advertising*, 19(3), 39–52.

Ortega, A. L. (n.d.). *Are microtargeted campaign messages more negative and diverse? An analysis of Facebook ads in European election campaigns*.

Papakyriakopoulos, O., Hegelich, S., Shahrezaye, M., & Serrano, J. C. M. (2018). Social media and microtargeting: Political data processing and the consequences for Germany. *Big Data & Society*, 5(2), 2053951718811844. https://doi.org/10.1177/2053951718811844

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. *Advances in Experimental Social Psychology*, 19, 123–205. https://doi.org/10.1016/S0065-2601(08)60214-2

Segijn, C. M., & van Ooijen, I. (2022). Differences in consumer knowledge and perceptions of personalized advertising: Comparing online behavioural advertising and synced advertising. *Journal of Marketing Communications*, 28(2), 207–226. https://doi.org/10.1080/13527266.2020.1857297

Stubb, C., & Colliander, J. (2019). "This is not sponsored content" – The effects of impartiality disclosure and e-commerce landing pages on consumer responses to social media influencer posts. *Computers in Human Behavior*, 98, 210–222. https://doi.org/10.1016/j.chb.2019.04.024

Van Der Goot, M. J., Van Reijmersdal, E. A., & Zandbergen, S. K. P. (2021). Sponsorship Disclosures in Online Sponsored Content: Practitioners' Considerations. *Journal of Media Ethics*, 36(3), 154–169. https://doi.org/10.1080/23736992.2021.1935962

van Reijmersdal, E. A., Fransen, M. L., van Noort, G., Opree, S. J., Vandeberg, L., Reusch, S., van Lieshout, F., & Boerman, S. C. (2016). Effects of Disclosing Sponsored Content in Blogs: How the Use of Resistance Strategies Mediates Effects on Persuasion. *American Behavioral Scientist*, 60(12), 1458–1474. https://doi.org/10.1177/0002764216660141

van Reijmersdal, E. A., Rozendaal, E., Hudders, L., Vanwesenbeeck, I., Cauberghe, V., & van Berlo, Z. M. C. (2020). Effects of Disclosing Influencer Marketing in Videos: An Eye Tracking Study Among Children in Early Adolescence. *Journal of Interactive Marketing*, 49, 94–106. https://doi.org/10.1016/j.intmar.2019.09.001

van Reijmersdal, E. A. van, Opree, S. J., & Cartwright, R. F. (2021). Brand in focus: Activating adolescents' persuasion knowledge using disclosures for embedded advertising in music videos. *Communications*. https://doi.org/10.1515/commun-2019-0168

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems ofp values. *Psychonomic Bulletin & Review*, 14(5), 779–804. https://doi.org/10.3758/BF03194105

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, 100(3), 426–432. https://doi.org/10.1037/a0022790

Wentzel, D., Tomczak, T., & Herrmann, A. (2010). The moderating effect of manipulative intent and cognitive resources on the evaluation of narrative ads. *Psychology and Marketing*, 27(5), 510–530. https://doi.org/10.1002/mar.20341

Wilson, D. G. (2017). The ethics of automated behavioral microtargeting. *AI Matters*, 3(3), 56–64. https://doi.org/10.1145/3137574.3139451

Winter, S., & Krämer, N. C. (2014). A question of credibility – Effects of source cues and recommendations on information selection on news sites and blogs. Communications, 39(4), 435–456. https://doi.org/10.1515/commun-2014-0020

Winter, S., Maslowska, E., & Vos, A. L. (2021). The effects of trait-based personalization in social media advertising. *Computers in Human Behavior*, 114. https://doi.org/10.1016/j.chb.2020.106525

Wojdynski, B. W., & Evans, N. J. (2016). Going Native: Effects of Disclosure Position and Language on the Recognition and Evaluation of Onlin. *Journal of Advertising*, 45(2), 157–168. https://doi.org/10.1080/00913367.2015.1115380

Wylie, C. (2019). *Mindf*ck: Cambridge Analytica and the plot to break America* (First edition). Random House.

Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y., & Chen, Z. (2009). How much can behavioral targeting help online advertising? *Proceedings of the 18th International Conference on World Wide Web – WWW '09*, 261. https://doi.org/10.1145/1526709.1526745

Zarouali, B., Dobber, T., De Pauw, G., & de Vreese, C. (2020). Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media. *Communication Research*. https://doi.org/10.1177/0093650220961965

Zuiderveen Borgesius, F. J., Möller, J., Kruikemeier, S., Fathaigh, R., Irion, K., Dobber, T., Bodo, B., & de Vreese, C. (2018). Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, 14(1), 82–96. https://doi.org/10.18352/ulr.420

Appendix

Definitions of the variables

Variable	Definition	References
Source credibility	Indication if the sender is able to provide valid statements on a topic, according to the receiver.	Winter and Krämer, 2014
Message credibility	Indication of an individual's judgement of the veracity of the content of communication.	Appelman and Sundar, 2016
Source trustworthiness	Indication of the receivers perception of honesty of the communicator.	Kruikemeier et al., 2016
Persuasion knowledge	Indicator of receivers' personal knowledge and beliefs about the motives and tactics of an advertisement and the sender of this advertisement.	Friestad and Wright, 1994

Operationalization of the measures

Variable	Items	Scale	References
Source credibility	competent – incompetent experienced – not experienced qualified – non-qualified	5-point semantic	Winter and Krämer, 2014
Message credibility	accurate authentic believable	7-point Likert	Appelman and Sundar, 2016
Source trustworthiness	undependable – dependable [a] dishonest – honest a selfish – unselfish [a] unreliable – reliable [a] untrustworthy – trustworthy [a] sincere – insincere [b]	7-point semantic	[a] Kruikemeier et al., 2016 [b] Winter and Krämer, 2014
Persuasion knowledge	The post feels like an ad The post promotes the sender Sender paid to post this message The post of the sender is an ad The post is sponsored by sender	7-point Likert	Kruikemeier et al., 2016

Means and standard deviations for correct recallment

Measure	No disclosure ($n = 54$)	Sponsored disclosure ($n = 18$)	Targeting disclosure ($n = 13$)
Persuasion Knowledge	3.88 (1.06)	4.74 (1.01)	4.18 (1.13)
Source Credibility	2.93 (1.01)	3.19 (0.97)	3.23 (0.80)
Message Credibility	4.15 (1.55)	4.39 (1.67)	4.26 (0.64)
Source Trustworthiness	4.14 (1.36)	4.22 (1.51)	4.17 (0.57)

Note. $N = 151$. All constructs were measured on a 7-point Likert or semantic scale.

Results of the mediation analyses: The indirect effects of targeting disclosure labels on the credibility of the source and message and source trustworthiness through persuasion knowledge

			95 % BCBCI	
Variable	**Indirect effect**	**SE**	**LL**	**UL**
Targeting disclosure vs. regular disclosure				
Source credibility through PK	0.00	.02	-0.06	0.03
Message credibility through PK	-0.01	.03	-0.08	0.04
Source trustworthiness through PK	-0.01	.03	-0.08	0.04
Targeting disclosure vs. no-disclosure				
Source credibility through PK	0.00	.02	-0.03	0.05
Message credibility through PK	0.01	.03	-0.04	0.08
Source trustworthiness through PK	0.01	.03	-0.04	0.07

Note. BCBCI, bias-corrected bootstrap confidence interval; PK, persuasion knowledge; SE, standard error; LL, lower limit; UL, upper limit

Acknowledgements

The authors thank Nur Efsan Cetinkaya for helping with the questionnaire and translations.

Date received: June 2022
Date accepted: June 2023

Article II

Jansen, M.-P., Meier, Y., & Krämer, N.C. (In revision) Time for transparent targeting: an investigation of targeting disclosures, coping mechanisms, credibility, and political attitude. *Behaviour and Information Technology.*

Time for transparent targeting: an investigation of targeting disclosures, coping mechanisms, credibility, and political attitude

Martin-Pieter Jansen[1], Yannic Meier[1], and Nicole C. Krämer[1, 2]

[1]Social Psychology: Media and Communication, University of Duisburg-Essen, 47057, Duisburg, Germany

[2]Research Center Trustworthy Data Science and Security, 44227, Dortmund, Germany

Author Note

Martin-Pieter Jansen [https://orcid.org/0000-0003-2876-8073]

Yannic Meier [https://orcid.org/0000-0002-5726-3229]

Nicole C. Krämer [https://orcid.org/0000-0001-7535-870X]

The authors report there are no competing interests to declare.

Correspondence concerning this article should be addressed to Martin-Pieter Jansen, Social Psychology: Media and Communication, University of Duisburg-Essen, Forsthausweg 2, 47057 Duisburg, Germany. E-Mail: martin-pieter.jansen@uni-due.de

Abstract

A major problem with political microtargeting (PMT) is that users are often unaware that they are being targeted, whereas current transparency approaches fail to inform users. Therefore, the current work investigates two types of disclosures on a political advertisement, one that is currently used by platforms: a light gray sentence stating that content is sponsored, and a salient and dynamic disclosure that informs users about the parameters and data points used to target them. This work investigates whether these disclosures, as a means of transparency to counter the potential negative effects of PMT, affect users in terms of persuasion knowledge, resistance, source credibility, and attitude towards the advertised politician. In a preregistered online one-factorial between-subjects experiment ($N = 547$), we found no evidence for the claim that a more elaborate, salient targeting disclosure increases users' persuasion knowledge. However, our results show that persuasion knowledge is related to cognitive resistance, which, in turn, is related to both source credibility and users' attitude towards the advertised politician. Our results further indicate that this process is cognitive, rather than affective. Finally, most participants seemed to recall a disclosure, although it was not always the correct one.

Keywords: microtargeting, disclosures, resistance, credibility, persuasion knowledge

Time for transparent targeting: an investigation of targeting disclosures, coping mechanisms, credibility, and political attitude

As people become more active online, the landscape of political communication shifts with their target audiences. Large parts of communication between (potential) voters and politicians occur online, specifically on social networking sites (SNS) (Giasson et al., 2019). Online behavior leaves digital breadcrumbs, especially on SNS, where behavior is interactive, as opposed to websites where most people consume information and a more one-directional sender-receiver relationship exists. The digital traces that users leave behind are very valuable to political consultancy firms, which use these data points to segment small groups of users and narrowly target them with political advertisements (Gandy, 2000; Murray and Scime, 2010). The allocation of users into groups and adapting messages that are supposed to resonate effectively with a specific audience is called *microtargeting* or, in the political landscape, *political microtargeting* (PMT) (Kosinski et al., 2013; Kruikemeier et al., 2016; Matz et al., 2017; Zarouali et al., 2020; Zuiderveen Borgesius, 2016). Two prime examples of the use of PMT are the 2016 United States elections and the Brexit referendum in the United Kingdom. In both examples, political consultancy firm *Cambridge Analytica* presumably used PMT through the gathering and connection of data from more than 50 million Facebook users, building psychological profiles, and targeting those users with messages made to resonate with them (Cadwalladr, 2018). In addition, PMT can be used not only to persuade voters, but also to discourage political participation, including electoral turnout (Bodó et al., 2017). One of the problems with PMT is that people do not recognize persuasive attempts that senders use to affect them. One solution to this could be the use of disclosures. These labels inform users directly about the content they read or engage with, showing that content is sponsored or paid for by certain parties or companies (e.g., a collaboration between an influencer and a brand). In the case of PMT, disclosures provide platforms with the

possibility of informing users that they are targeted by the sender of a message through their data.

Although research on marketing and advertising has already focused on disclosures (Amazeen and Wojdynski, 2020; Boerman et al., 2015; Cain, 2011), research on disclosures for political microtargeted advertisements is still scarce (Binford et al., 2021; Kruikemeier et al., 2016). In addition, the effects of seeing disclosures could go beyond simply informing receivers and prompt psychological mechanisms, such as counterarguing and contesting the message, but can also negatively affect users' image of the party sending the message or the politician being advertised (Kruikemeier et al., 2016; Van Reijmersdal et al., 2016). However, currently used disclosures do not always seem to be correctly recalled (Binder et al., 2022; Binford et al., 2021; Kruikemeier et al., 2016).

Therefore, the current work investigates two different types of disclosures: a regular sponsorship disclosure and a dynamic targeting disclosure containing more information. In this disclosure, users can hover over a targeting measure to show the parameters used to target them (i.e., age, gender, and behavior). Building on and expanding existing work, this study has an experimental design and focuses on the two different types of disclosures. Consequently, we aim to examine how users cope with messages when it is disclosed that they are targeted by the sender of the message, and the mechanisms that are activated after exposure to the disclosure, being persuasion knowledge, and cognitive and affective resistance. Moreover, we investigate the different outcomes of these mechanisms with regard to source credibility and attitude towards an advertised politician.

Theoretical background

One of the problems regarding microtargeting is that it is a black box, and neither laypersons, journalists, nor scientists know exactly what is done by political consultancy firms. This black box makes it difficult to research and practically impossible to estimate the

effects of PMT (e.g., Cambridge Analytica's practices). Since the lack of transparency and information about the algorithms and models of political consultancy firms combined with the lack of a regulatory body makes it hard to focus on the 'sender' side of PMT, the solution could be to inform users by implementing transparency measures.

Disclosures

One approach concerning transparency and informing users could be the use of disclosures, specifically targeting disclosures (Binder et al., 2022; Binford et al., 2021; Jansen and Krämer, 2023; Kruikemeier et al., 2016). Disclosures on posts and advertisements on SNS were initially used to inform users of the messages being sponsored, labeling them as such. Within research on marketing and advertising, work already exists on disclosures, showing that they facilitate the recognition of advertisements (Amazeen and Wojdynski, 2020; Boerman et al., 2012, 2015; Jung and Heo, 2019).

Research on disclosures regarding PMT seems to be emerging but is still scarce and shows mixed results concerning recognition (Binder et al., 2022; Binford et al., 2021; Jansen and Krämer, 2023; Kruikemeier et al., 2016). Recent work acknowledges the opportunities that disclosures bring to inform receivers of the targeting practices that occur. Following the EU Digital Services Act (DSA), targeting disclosures should overlap with sponsored disclosures in containing information on when and on whose behalf content is displayed but simultaneously go beyond that and include individualized data used to target users and parameters used for those data points (European Commission, 2022).

Regarding the design of disclosures, the ones that SNS use nowadays, light grey embedded around the content (e.g., 'Sponsored, paid for by …'), seem to be insufficient to inform users and contribute to transparency (Binder et al., 2022; Binford et al., 2021; Kruikemeier et al., 2016). However, more prominent disclosures seem to be effective in informing users of the persuasive or targeted nature of a message. Prominent disclosures are

those that are more easily seen by users and therefore lead to higher levels of recognition that an advertisement is an advertisement (Amazeen and Wojdynski, 2020) For instance, research shows higher levels of advertising recognition with a highlighting square around the ad and the disclosure (Jost et al., 2022). Other work suggests that disclosures that include information that aligns with the regulations in the DSA are noticed more easily (compared to disclosures investigated in prior advertising research) (Dobber et al., 2023). In line with this, the current work investigates a more extensive disclosure containing information about the advertisement being targeted at users. As an exploratory feature of the disclosure, we implement a dynamic function showing participants that the advertisement is targeted to them based on their information. This makes the disclosure used in the experimental condition align more with the DSA, which states that the parameters and data points used to fill in those parameters should be disclosed (Digital Services Act, 2022; European Commission, 2022). This type of disclosure has been shows to lead to better levels of disclosure recall (Dobber et al., 2023). Besides providing participants with information that fits their person, the dynamic feature might also lead to better cognitive absorption and a more positive attitude towards the content (Oh and Sundar, 2015). In addition, the usage of red and underlined text for the parameters within the disclosure not only shows participants that extra information lies behind the word (e.g., as is the case with a hyperlink on a website) but the color red also attracts more visual attention (Baik et al., 2013), while making the disclosure more prominent since it stands out more in a user's timeline on a platform (Amazeen and Wojdynski, 2020). Moreover, the hover feature can show participants information in a more fluent manner than Facebook's 'Why am I seeing this ad' because the popup will (dis)appear based on the mouse movements and is therefore less interrupting and more readily available, while still being applicable in the platforms corporate design, compared to large boxes drawn around an advertisement (Jost et al., 2022). While existing work underlines the importance of

investigating new types of disclosures that adhere to the regulations of the DSA (Dobber et al., 2023; Jost et al., 2022), we examine a newly developed targeting disclosure with red and underlined text, as well as the option to see the actual datapoints that were used regarding the targeting parameters. To the best of our knowledge, targeting disclosures such as these have not been studied with regard to PMT.

Persuasion knowledge

Nowadays, users of social networking sites (SNSs) are being exposed to an increasing number of persuasive messages. To cope with these messages and their potential influence, receivers develop beliefs and knowledge regarding these attempts through previous exposure and experience. The overarching term most commonly used to describe these mechanisms is *persuasion knowledge* (Friestad and Wright, 1994). Research shows that if messages contain a disclosure, persuasion knowledge is more likely to be activated, whereas if there are no disclosures on a message, persuasion knowledge is less likely to be activated (Boerman et al., 2012; Kruikemeier et al., 2016; Van Reijmersdal et al., 2016). However, recent work emphasizes the importance of users actually seeing and perceiving the disclosure, while only then can the disclosure inform them and activate persuasion knowledge (van Reijmersdal et al., 2023). Moreover, seeing a disclosure can also lead to a change of meaning of the message (i.e., being aware of persuasive intent), which in political advertising could mean that people think the message might be less sincere (Friestad and Wright, 1994; Wojdynski and Evans, 2020). A previous study found a positive relationship between a disclosure and persuasion knowledge in the PMT context (Kruikemeier et al., 2016). However, this study used a sponsored disclosure and training condition, but not a dynamic disclosure. Moreover, other work shows that including parameters in the disclosure, as is obligated by the DSA, could be promising regarding recall of the disclosure as well as persuasion knowledge (Dobber et al., 2023).

As the practices of PMT differ from those of regular political advertising, we consider it important to examine whether the use of PMT, when recognized, generates differences in effects compared to regular political advertising. Furthermore, we propose that activating persuasion knowledge is the first step for users to cope with persuasive messages. Persuasion knowledge has been found to be a valid measure to scrutinize users' knowledge of the persuasive tactics occurring. However, the extension of this measurement applied to a microtargeting context developed by Binder and colleagues (2022), *targeting knowledge*, has not yet been investigated in other research. Therefore, as an exploratory research question, we compare our measured constructs between our two experimental groups, and also compare the level of targeting knowledge between those groups. Thus, we can compare persuasion and targeting knowledge between our groups, and scrutinize whether these constructs differ between and within groups.

Current research acknowledges the activation of persuasion knowledge through the implementation of disclosures (Binder et al., 2022; Dobber et al., 2023). However, disclosures such as those in the current study have not been investigated yet since the disclosures are newly developed. We expect that our targeting disclosure, including interactive parts that show the actual parameters used to target the user, activates higher levels of persuasion knowledge than a regular disclosure stating that the message is sponsored:

H_1: Exposure to a targeting disclosure leads to higher levels of persuasion knowledge than exposure to a sponsored disclosure.

RQ_1: What are the differences between users who receive a sponsored disclosure and users who receive a targeting disclosure regarding persuasion knowledge and targeting knowledge?

Resistance

Youn and Kim (2019) found that when an advertisement is recognized as such, persuasion knowledge is activated, and in turn, resistance could be activated as a coping mechanism, which negatively affects users' attitudes towards the content. Furthermore, research shows that when persuasion knowledge is activated as a coping mechanism for persuasive attempts, people use their existing knowledge to either be persuaded or resist the message (Sagarin et al., 2002; Tormala and Petty, 2004; Van Reijmersdal et al., 2016). According to Quinn and Wood (2004), resistance is the process through which one rejects influence and retains one's views. We propose that the activation of persuasion knowledge leads to resistance because receivers feel that a message is a perceived threat to freedom (Dillard and Shen, 2005). This potential threat of freedom can be countered through the activation of resistance, increasing users' certainty of their initial attitudes (Tormala and Petty, 2002). Furthermore, studies on online behavioral advertising, which can be seen as the predecessor of PMT, have found that the technique has been perceived as intrusive (Segijn and Van Ooijen, 2022) and even creepy (Segijn et al., 2022). This could mean that if users are made aware of the targeting practices taking place through disclosures and recognize this through persuasion knowledge, they tend to resist the message more. However, to the best of our knowledge, the concept of resistance to persuasion has not been investigated in the context of microtargeting.

People's attitudes contain distinctive cognitive and affective components. These components have valence dimensions that range from negative to positive (Ostrom, 1969; Rosenberg et al., 1960). Persuasion, which can be considered an attempt to change existing attitudes towards a brand or, in the case of PMT, a politician, has also been found to have different cognitive and affective dimensions (See, 2018), which makes receivers react differently to stimuli depending on which dimension is activated more (Di Plinio et al.,

2022). In addition, attitude change can be a function of one's favorable cognitive and affective reactions to a message simultaneously (Petty and Cacioppo, 1986; Rosselli et al., 1995). Furthermore, cognitive responses can mediate the effect of exposure to a message on a persuasive outcome more than affective responses can (Banerjee and Greene, 2012).

Users' resistance to a message can also be both cognitive ("I do not believe it!") and affective ("I do not like it!"), also at the same time (Knowles and Linn, 2004, p. 7; Sagarin and Cialdini, 2004). Previous studies on sponsored content and advertising show that affective and cognitive resistance can be activated as a result of persuasion knowledge activation (Fransen, Smit, et al., 2015; Fransen, Verlegh, et al., 2015). We deem it important to investigate both routes simultaneously and scrutinize potential differences in resistance and differences in psychological effects on users. Cognitive resistance towards a persuasive message means that users have a negative reaction towards a persuasive message based on their preexisting thoughts and various (message) factors (e.g., disclosure or through recognizing it themselves). Previous work has found that counterarguing is an effective and often-used strategy for resisting messages or advertisements (van Reijmersdal et al., 2016; Zuwerink Jacks and Cameron, 2003). Additionally, disclosures could increase thinking about advertising and decrease positive or neutral cognitions about brands or advertisers, thereby activating cognitive resistance (Boerman et al., 2012). To the best of our knowledge, resistance in these forms has not yet been investigated in the case of microtargeting. Building on the work of Van Reijmersdal and colleagues (2016), which distinguished between cognitive and affective resistance with regard to disclosing sponsored content in blogs, we adapt this to the context of political microtargeting and expect that persuasion knowledge has a positive relationship with the cognitive dimension of resistance:

H_2: Persuasion knowledge is positively related to cognitive resistance.

Although different kinds of cognitions as a response to persuasive messages are important, users also tend to have affective responses to these messages. An affective response can be defined as a change in a user's mood or feelings after exposure to a message (e.g., getting annoyed or irritated by commercials) (Edell and Burke, 1987; Schwarz et al., 1991). A positive mood reduces the motivation to process content and contextual signals systematically. Equivalently, a negative mood induces more effortful, detail-oriented, analytical, and critical processing, which could lead to skepticism towards the persuasive message (Boerman et al., 2012; Bohner et al., 1992). Other studies have shown that persuasion knowledge activation is positively related to affective resistance with regard to disclosing sponsored content in blogs (van Reijmersdal et al., 2016) and can even prompt feelings of anger (Beckert et al., 2021). Examining the second route of resistance following Van Reijmersdal and colleagues (2016), we expect that in the case of microtargeting, persuasion knowledge has a positive relationship with the affective dimension of resistance:

H_3: Persuasion knowledge is positively related to affective resistance.

Source Credibility

Research on advertisement disclosures on SNS shows that disclosures are likely to have a negative effect on the credibility of the source of a message and the message itself (Carr and Hayes, 2014; Deng et al., 2020; Stubb and Colliander, 2019; Wojdynski and Evans, 2016). Source credibility is mostly conceptualized as a set of positive characteristics associated with the source of a message that could potentially increase the acceptance of the message by receivers (Housholder and LaMarre, 2014; Ohanian, 1990). Receivers' perceived credibility influences how they use and process the information that they receive from a source (Madsen, 2019).

Moreover, credibility has been found to lead to more positive responses to persuasive attempts, as described in the Elaboration Likelihood Model (ELM) (Petty and Cacioppo, 1984). In addition, credibility is an important factor when the subject of discussion is something that the receiver of a message has less knowledge about, which is often the case for political topics (Yoon et al., 2005). Source credibility is an important predictor of voting intention, and a decreased perceived credibility tends to decrease the willingness to vote for a certain party or candidate, possibly making it counterproductive for political actors or parties to take actions that potentially lower credibility (Hetherington, 1999; Housholder and LaMarre, 2014; Madsen, 2019). Although disclosures are usually obligatory by regulations or laws, and not something political actors and parties choose to implement, they could still benefit from the awareness of potential negative effects (European Commission, 2022).

In line with Van Reijmersdal and colleagues (2016), we expect that if people are informed about a message being sponsored or targeted at them through a disclosure, they activate resistance through persuasion knowledge. Subsequently, we propose that higher levels of resistance decrease the perceived credibility of the source (Carr and Hayes, 2014). In this study, we use the construct of source credibility as a measure to research the credibility perceptions that users have towards the political party that sends the message. While the politician who is advertised might sometimes be the source of the message, we used the party, which is also shown to be the sender of the ad in the stimulus material. We propose that being aware of the targeting that a source employs to persuade users has the potential to increase resistance and lower users' perception of credibility regarding the source:

H_4: Cognitive resistance is negatively related to source credibility.

In line with our other hypotheses and earlier work, we investigate the two separate routes of cognitive and affective resistance, as they are two separate psychological processes. The rationale behind measuring the routes as separate is a potential difference between counterarguing with a message or sender or actually getting annoyed or irritated. There are mixed results on the activation of cognitive and affective resistance at the same time (Fransen, Smit, et al., 2015; Petty & Cacioppo, 1986; Rosselli et al., 1995) and separately (Banerjee and Greene, 2012; Di Plinio et al., 2022; See, 2018), and their effects on persuasive outcomes. Therefore, we scrutinize both types of resistance and propose that, like cognitive resistance, affective resistance is negatively related to source credibility and investigate the potential differences between the types of resistance:

H_5: Affective resistance is negatively related to source credibility.

RQ_2: What are the differences in the relationships between cognitive and affective resistance, and source credibility?

Attitude towards the politician

When receivers are exposed to content on SNS, they form attitudes towards both the content and the sender. However, the sender of the content (source) does not always have to be the person displayed (e.g., a post from a political party, advertising a specific politician). Users' perceived credibility of a source has been found to be one of the factors influencing their attitudes towards that source (Kumkale et al., 2010). Furthermore, since source credibility has been shown to predict voting behavior and choice for a candidate (Madsen, 2019, p. 107), and high levels of source credibility have been shown to predict more positive candidate evaluations, we deem it important to research a possible relationship between the perceived credibility of the source and receivers' attitudes towards this source (Funk, 1999; Markus, 1982; Miller et al., 1986). Moreover, in earlier work on disclosures and brand

attitudes, Van Reijmersdal and colleagues (2016) found that cognitive and affective resistance have negative relations to attitudes, in line with work that found that counterarguing a persuasive message leads to lower persuasion in terms of attitudes (Zuwerink Jacks and Cameron, 2003). Furthermore, reactance to a message has been found to have a negative effect on brand attitudes (Beckert and Koch, 2022). In line with this, we propose that users could resist a message containing a disclosure because it feels intrusive to be targeted, which lowers users' attitude towards the politician:

H_6: Cognitive resistance is negatively related to attitude towards the politician.

H_7: Affective resistance is negatively related to attitude towards the politician.

RQ_3: What are the differences in the relationships between cognitive and affective resistance and attitude towards the politician?

Furthermore, higher levels of source credibility lead to more favorable attitudes than lower levels of source credibility (Tormala et al., 2006). Accordingly, we propose a positive effect of source credibility on users' attitude towards a politician:

H_8: Source credibility is positively related to attitude towards the politician.

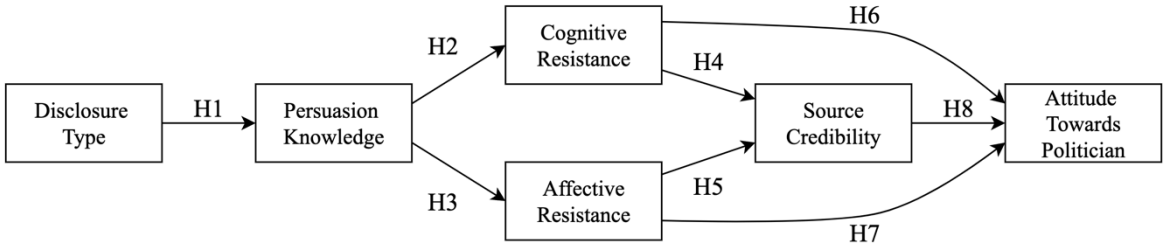
Finally, as an exploratory research question, we investigate the potential differences between our two different types of disclosures on cognitive and affective resistance, source credibility, and attitude towards the politician. This enables us to at least speculate about the most fitting designs concerning disclosures on microtargeted advertisements on SNS, which is important not only to inform users but also for platforms to oblige to the new guidelines from the European Union.

RQ4: What, based on exposure to the different types of disclosures, are the differences in cognitive and affective resistance, source credibility, and attitude towards the politician?

An overview of our proposed hypotheses and path model is shown in Figure 1.

Figure 7

Proposed path diagram



Method

The online experiment was approved by the ethics committee of the University of Duisburg-Essen We preregistered this study before collecting data: https://osf.io/sftbe/?view_only=10057e05315342cbbae9e95e91c2da20. All participants provided informed consent before participation. Supplementary materials, code, and measures in both English and German are publicly accessible on OSF: https://osf.io/dwygm/?view_only=9e486925981f4b72ab9c16e59369687c

Design

To test our hypotheses, we conducted an online experiment using a factorial between-subjects design containing two groups. Participants were exposed to a Facebook advertisement containing either a sponsored disclosure, in line with the disclosures Facebook uses ("Sponsored, paid for by …") in the control condition, or a Facebook advertisement with

a targeting disclosure ("This message is targeted on you based on your: age, gender, and online behavior"). The reason that there is a sponsored disclosure on the advertisement in our control condition is that an advertisement on Facebook without a disclosure would be an organic post and not an advertisement because these disclosures are regulated and mandatory.

When hovering over age, gender, and online behavior on the targeting disclosure, more information about the targeting parameters was provided. These parameters were based on the demographic information we asked the participants before the start of our microtargeting simulation and were thus in line with their answers. Participants' registered age was converted into groups (e.g., a participant that filled in their age as 35, fell into the 31-40 years group) and the following message was given: *This message is targeted on people between the ages of [age group]*. The gender parameter was a direct insert of the gender the participant answered that they identified with (female, male, or diverse), and the following message was given: *This message is targeted on [gender] users.* For online behavior, we were unable to use information, as we did for demographics. Therefore, the following message was given: *This message is targeted on you through your online behavior and interactions on Facebook*. We were able to measure how many times participants hovered over each term individually, and thus, were exposed to the targeting parameters we included. We recognize that the hover option was possible because participants took part in the experiment on a computer. On a smartphone or tablet, participants would need to click on the term, which we were unable to program because of limitations in the platform we used to conduct the experiment.

We created a total of four different Facebook advertisements through the HTML and PHP programming options in the platform we used to conduct the experiment (SoSci Survey), meaning we changed input fields for the stimuli, which led to four different versions (i.e., pro- or anti-climate change regulations for both conditions; see procedure). For each of

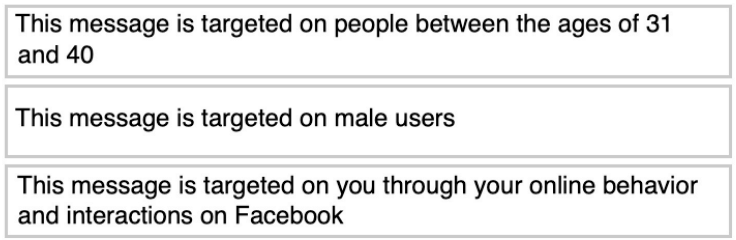
the two experimental conditions, the picture containing the politician was the same, the profile picture and the name of the page that posted the post were the same, as were the number of likes and comments (see Figure 2). The statement above the post differed for the pro and anti-climate change regulations groups, leaving both groups with a statement supposedly in line with their own beliefs. The only other differences were the two disclosures that were manipulated. Our manipulated disclosures looked like the disclosures Meta/Facebook uses nowadays, although we did change the font size and color to be able to use the dynamic functions we applied (i.e., hovering and using parameters that participants reported; see Figure 3).

Figure 8

Stimuli

Figure 9

Dynamic boxes including example parameters



Sample

Our sample, consisting of 577 German Facebook users, was recruited through panel provider Respondi AG. From our total sample, 30 people completed the experiment in a time of less than one and a half minutes, and therefore they were excluded from the dataset, leaving a total sample size of 547 participants for the analyses. Within our sample, the age ranged from 18 to 74 years (M = 46.8, SD = 15.6). Of these participants, 199 identified as female, and 348 as male. Regarding education, most of the participants have an intermediate high school diploma (n = 163) and 123 have a university entrance qualification. 72 participants have a master's degree and 67 have a bachelor's degree. Furthermore, 57 participants had a qualifying middle school diploma and 44 had an advanced technical college entrance diploma. Finally, four participants have a doctorate and 17 participants responded that they have a different educational background.

Randomization checks showed no differences between our experimental groups concerning age ($F(1, 545) = 0.15$, $p = .736$), gender ($\chi^2(1, N = 547) = 0.14$, $p = .704$) or level of education ($\chi^2(7, N = 547) = 11.74$, $p = .109$).

Power

Due to our available resources, we aimed to collect responses from 500 participants (Lakens, 2022). However, we were able to collect 547 responses. Statistical power was calculated using R (version 4.1.2) (R Core Team, 2021) and the simsem package (Version 0.5-13) (Pornprasertmanit et al., 2021). The power analysis aimed to find the smallest effect size that could be found with a recommended minimum power of 90%, given our current

sample size as well as the statistical method. The smallest effect size of interest (SESOI; Lakens et al., 2018) is complementary to and comparable to the alpha-level, i.e., a threshold for rejecting results. To determine our SESOI, we calculated different simulation analyses with the structure of the proposed structural equation model (SEM). While including different effect sizes in our simulation and keeping the α level (α = .05), power (90%), and sample size (N = 547) constant, we ran 1,000 replications for our model. Consequently, we set our SESOI at β = |.14| and did not interpret results with effect sizes below that threshold.

Pre-tests

To at least simulate microtargeting, we intended to expose participants to advertisements in line with their beliefs. To accomplish this, we asked participants if they agreed or disagreed with a set of 10 statements regarding politically loaded topics. Four of these statements were chosen from a set of 10 pre-tested statements (N = 100), investigating whether they clearly described the views of someone who was *pro-* or *anti-climate* change regulations (i.e., someone who is *pro-climate change,* is willing to counter climate change by supporting the regulations; the opposite is true for *anti*). We showed participants a total of 10 statements on various political topics to counter test effects. The answers to the four statements regarding climate change regulations were registered and later used to show a statement that was either pro- or anti-climate change regulations on the Facebook post itself.

Moreover, in another pre-test, we tested the design and layout of our dynamic disclosure. We pre-tested four types of font styles and colors, representing different levels of prominence (N = 36), to find the one perceived as most prominent, which was finally used in our experiment.

Procedure

After providing informed consent, the participants were asked about their demographic information because we needed that information to fill the parameters for the

targeting disclosure group. Thereafter, participants were asked if they agreed or disagreed with 10 statements on politically loaded topics. After the statements, participants were shortly briefed again and told that the next page would contain a Facebook post and that the 'next-button' would appear after 15 seconds, to ensure that they had a close look at the post. Thereafter, we exposed the participants to our stimuli and asked them questions regarding our variables and our manipulation check. Finally, participants were debriefed and thanked. The average completion time was 5 minutes.

Measures

All measured constructs were tested for factor validity using confirmatory factor analyses (CFAs). The results of the CFAs and the measures of reliability and internal consistency are shown in Table 1. We measured *persuasion knowledge* through five items validated in earlier research on PMT by Kruikemeier et al. (2016) (based on Ham et al. (2015)) on a Likert scale ranging from 1 (= strongly disagree) to 7 (= strongly agree) (e.g., "The post feels like an ad"). The average variance extracted for this measure was just below the threshold of .50, but we kept all items because other reliability measures had a sufficient outcome, and the scale was validated in other research regarding PMT. We measured the *cognitive resistance* strategy 'counterarguing', and the *affective resistance* strategy 'negative affect' by asking participants to what extent they agreed with two times four statements based on Van Reijmersdal et al. (2016) on a Likert scale ranging from 1 (= strongly disagree) to 7 (= strongly agree) (e.g., cognitive: "While reading I contested the information of the message" & affective: "While reading the message I felt angry"). *Source credibility* was measured using the scale by Ohanian (1990), which is often used in persuasion research and contains three dimensions (attractiveness, trustworthiness, and expertise). This scale is measured on five-point semantic scales. Asking participants which term best applies to the source of the message (i.e., the political party; e.g., "reliable – unreliable"). In the

introductory text, we explicitly told the participants that these questions were about the political party. Due to an overlap in translation into German, the item for dependable – undependable for the dimension of trustworthiness was not measured in this study. The scores for the three dimensions were summed to obtain the total average score for the analyses. We measured *attitude towards politician* using a nine-item, seven-point semantic scale that has recently been used in microtargeting research by Dobber et al. (2021). We explicitly asked participants to assess the politician portrayed in the advertisement according to the given characteristics (e.g., "corrupt – upright").

Table 1

Results of the Confirmatory Factor Analyses

Measured constructs	χ^2	*df*	*p*	**CFI**	**TLI**	**RMSEA**	**SRMR**	α	ω	**AVE**
Persuasion Knowledge	81.44	5	<.001	.92	.84	.17	.04	.81	.82	.48
Cognitive Resistance	71.82	2	<.001	.95	.85	.25	.03	.89	.90	.69
Affective Resistance	257.20	2	<.001	.92	.76	.48	.02	.97	.97	.89
Source Credibility	1252.00	77	<.001	.86	.83	.17	.06	.97	.97	.67
Attitude Towards Politician	315.20	27	<.001	.95	.93	.14	.03	.96	.96	.74
Targeting Knowledge	172.80	5	<.001	.92	.83	.25	.06	.89	.89	.65

Note. Reliability measures are Cronbach's α, McDonald's ω, and average variance extracted. CFI = Comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual; AVE = average variance extracted.

Exploratory variable

As an exploratory behavioral variable, we were able to measure the number of times participants in the targeting disclosure group hovered over the dynamic functions in our stimuli and thus were exposed to the actual targeting parameters we developed, which can be seen as behavioral data of participants interacting with the stimulus. We measured each specific term (i.e., age, gender, and online behavior) and the number of times the participants hovered over it to see the specific information. Because the control condition did not have any dynamic functions, we were unable to measure this for that condition. As a second exploratory variable, we asked participants the extent to which they found the posts they received targeted at them. We did this by using the *targeting knowledge* scale developed in recent microtargeting work by Binder et al. (2022), consisting of five items measured on a seven-point Likert scale ranging from 1 (= strongly disagree) to 7 (= strongly agree) (e.g., "The post is based on my personal preferences"). Measuring this has implications for our simulation of microtargeting and its feasibility for future research.

As a manipulation check, we asked participants if they recalled a sponsored disclosure, a targeting disclosure, or no disclosure on the Facebook post. We also asked participants for their demographic information.

Results

Statistical analyses were performed using R (version 4.1.2) (R Core Team, 2021) and Jamovi (version 2.0.0.0) (The jamovi project, 2022). To investigate the differences between the control condition and the targeting disclosure condition, we analyzed our data using structural equation modeling with the variable mean scores as observed variables. Bivariate correlations, means, and standard deviations for the measured variables are shown in Table 2. The code of the conducted analyses is available on the OSF (https://osf.io/dwygm/?view_only=9e486925981f4b72ab9c16e59369687c).

Regarding our manipulation check, of the participants that were in the targeting disclosure condition, 41 recalled targeting information on the disclosure (15 of those participants used the hover function), while 124 wrongly recalled sponsored information on the disclosure, and 115 recalled no disclosure. Interestingly, of the participants who were in the control condition with the sponsored disclosure, 63 recalled seeing targeting information in the disclosure, while 187 correctly recalled seeing a sponsored disclosure, and 17 participants did not recall the disclosure.

Table 2

Means, Standard Deviations, and Bivariate Correlations of the Measured Constructs

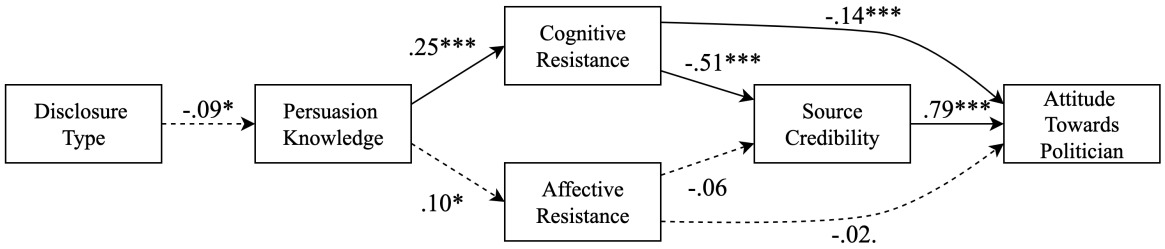
Measured construct	*M* (*SD*)	1	2	3	4	5
1 Persuasion Knowledge	4.6 (1.3)	-				
2 Cognitive Resistance	3.5 (1.7)	.25***	-			
3 Affective Resistance	2.3 (1.7)	.10*	.61***	-		
4 Source Credibility	2.6 (0.8)	-.12**	-.54***	-.37***	-	
5 Attitude Towards Politician	4.3 (1.4)	-.15***	-.57***	-.39***	.87***	-
6 Targeting Knowledge	3.3 (1.7)	-.09*	-.45**	-.22***	.67***	.65***

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Tests to see if our data met the assumption of collinearity for source credibility and attitude towards the politician showed that multicollinearity was not a concern (Attitude towards the politician, Tolerance = .703, VIF = 1.42) (Hair et al., 2014).

The path model is shown in Figure 4. The first model did not fit well. While investigating the model fit through the modification indices, we found that cognitive resistance and affective resistance correlated, after including this correlation, our model showed an adequate fit: $\chi^2(6) = 5.99$, $p = .424$, $\chi^2/df = 0.99$, CFI = 1.00, TLI = 1.00, RMSEA = .00, 90% CI [.00, .06], SRMR = .01.

Figure 10

Path Model



Note. Numbers represent standardized regression coefficients. Dashed lines indicate a path is either statistically meaningless (not significant) or theoretically meaningless (below β = |.14|). * $p < .05$, ** $p < .01$, *** $p < .001$.

The first hypothesis (H_1) predicted that exposure to a targeting disclosure would lead to higher levels of persuasion knowledge compared to our control condition. Contrary to our hypothesis, our results revealed a small negative effect (β = -.09, $p = .022$), which is also below our SESOI (β = |.14|). Thus, we cannot interpret this effect, leading us to reject H_1. For the second hypothesis (H_2), which predicted a positive relationship between persuasion knowledge and cognitive resistance, we found a significant positive relationship (β = .25, $p < .001$), supporting this hypothesis. Furthermore, we hypothesized that persuasion knowledge would also be positively related to affective resistance (H_3). Our results showed a significant positive relationship (β = .10, $p = .031$), but the effect size is below our SESOI; therefore, we reject this hypothesis.

Our fourth hypothesis (H_4) predicted a negative relationship between cognitive resistance and source credibility. Our results showed a significant negative relationship (β = -.51, $p < .001$), supporting this hypothesis. With the fifth hypothesis (H_5), we proposed a negative relationship between affective resistance and source credibility. The results show

that there is a non-significant negative relationship (β = -.06, p = .223). Therefore, we reject this hypothesis. For the sixth hypothesis (H_6), a negative relationship was hypothesized between cognitive resistance and attitude towards the politician. The results show a significant negative relationship (β = -.14, p = .002), supporting this hypothesis. Furthermore, for our seventh hypothesis (H_7), a negative relationship between affective resistance and attitude towards the politician was predicted. Our results showed a non-significant negative relationship (β = -.02, p = .571), rejecting the hypothesis. Finally, our eighth hypothesis (H_8) predicted a positive relationship between source credibility and attitude towards the politician. Our results showed a significant positive relationship (β = 79, p < .001), supporting this hypothesis. In addition, we found a significant covariance between cognitive and affective resistance (β = .61, p < .001). Even though the relationships of both types of resistance on source credibility and attitude towards the politician differ, this shows that the two variables are also positively related.

As exploratory analyses (RQ1), we first investigated the differences between our measured variables based on exposure to the two different types of disclosures. We conducted ANOVAs for all measured variables with experimental groups (targeting disclosure vs. control) as factors and found significant differences for *persuasion knowledge* (F(1,545) = 5.23, p = .023), ($M_{control}$ = 4.8, SD = 1.3) ($M_{targeting}$ = 4.5, SD = 1.2) and *targeting knowledge* (F(1,545) = 9.72, p = .002), ($M_{control}$ = 3.1, SD = 1.7) ($M_{targeting}$ = 3.6, SD = 1.6). This means that for the control condition, the average level of persuasion knowledge was higher than that in the targeting condition, while for targeting knowledge, the level in the targeting condition was higher than that in the control condition. Second, we investigated whether the option to hover over terms (i.e., age, gender, online behavior) in the targeting disclosure condition was used by participants, and how this could correlate with persuasion knowledge or targeting knowledge (note: this hovering was only possible for participants in

the targeting disclosure condition). Regarding the hovering, on average this possibility was used less than one time per participant (N = 280, M = 0.9, SD = 1.8). Nevertheless, only 92 participants in the experimental condition actually used the hovering option, and on average they did so between two and three times (N = 92, M = 2.7, SD = 2.2). We investigated differences in our measured constructs between participants that did or did not use the hovering option, we only found significant differences in *source credibility* (F(1,278) = 8.57, p = .004), ($M_{nohover}$ = 2.7, SD = 0.7) (M_{hover} = 2.4, SD = 0.7) and *attitude towards the politician* (F(1,278) = 5.04, p = .026), ($M_{nohover}$ = 4.4, SD = 1.3) (M_{hover} = 4.0, SD = 1.4). This means that for participants in the experimental condition who used the hovering, compared to participants in the experimental condition who did not, both source credibility and attitude towards the advertised politician were lower on average. Moreover, we investigated differences in our measured constructs between participants who used the hovering option and the control group but did not find any significant differences. Results of these analyses can be found online: (https://osf.io/dwygm/?view_only=9e486925981f4b72ab9c16e59369687c).

Discussion

The current study aimed to investigate whether targeting disclosures for political microtargeted advertisements can be beneficial for users in terms of raising their awareness that they are being targeted and how this relates to the cognitive and affective processing of advertisements as well as attitudes. To do so, we tested an integrative path model that portrays how social media users build their attitudes towards microtargeted content.

We assumed that exposure to a dynamic targeting disclosure (compared to a sponsorship disclosure), showing data used to target users on an advertisement, activates higher levels of persuasion knowledge, that is, the recognition of the message as an advertisement, in contrast to a sponsored label (Binder et al., 2022; Kruikemeier et al., 2016).

Although persuasion knowledge is described as a coping mechanism regarding persuasive attempts, we found a small negative effect within our sample, which is contrary to our expectations (Friestad and Wright, 1994; Van Reijmersdal et al., 2016). However, the found relationship was below our set threshold for the SESOI; thus, the current study does not have enough power to interpret this result as meaningful. We want to emphasize that this does not have to be the case for future studies, and we encourage other researchers to investigate these potential effects. Nevertheless, we would elaborate on the negative effect on a speculative basis since it is opposite to what we predicted. Looking at the stimuli used in this study, the disclosure in the control condition clearly states that the message is sponsored and paid for by the political party. In the targeting disclosure condition, the disclosure only states that the message is targeted at the participant based on the parameters. The label does not disclose the sponsored or persuasive nature of the Facebook post. Looking at our first research question, we see that regarding persuasion knowledge, the mean in the control condition was higher than that in the targeting condition.

Consistent with previous studies, not all participants recalled the correct disclosure (Binder et al., 2022; Evans et al., 2017; Jansen and Krämer, 2023; Kruikemeier et al., 2016; Van Reijmersdal et al., 2021). Only 17% of participants in the condition who saw a post with a targeting disclosure recalled the correct disclosure, while in the control condition, 70% of participants recalled the correct disclosure. Interestingly, users recall a sponsored disclosure more easily. The sponsored disclosure used in the control condition was designed to look like the one Facebook uses. Initially, we used this label to improve ecological validity while acknowledging that it has been found to not be a sufficient transparency measure; therefore, we used it as our control group (Binder et al., 2022; Binford et al., 2021). However, our findings show that the disclosure users are familiar with is recalled more easily, which means that a repeated exposure effect, increasing recognition, and potential familiarity might occur

(Montoya et al., 2017). Nevertheless, we found differences between our two conditions and targeting knowledge, and argue that disclosures could improve transparency, but users might need to become aware of targeting processes as a whole before they can recognize these processes through a disclosure. Moreover, we included hovering options in the targeting disclosure, which only 33% of participants used. However, participants who used this option used it three times on average, which shows the possibilities of more elaborate and interactive disclosures for SNS, which could take time for users to get accustomed to (Amazeen and Wojdynski, 2020; Binder et al., 2022; Boerman et al., 2015). The interactive part of the disclosure was implemented for use on a computer, but a different format is needed to be feasible for users on mobile devices (e.g., an 'I' button, providing additional targeting information if pressed on a mobile phone, or a small box with additional information if the screen is larger, such as a tablet).

Besides testing the potential effects of the targeting disclosure on persuasion knowledge, we expected that persuasion knowledge activation would relate to users resisting the message more, both cognitively and affectively (Beckert et al., 2021; Van Reijmersdal et al., 2016). We observed this relationship only for cognitive resistance. Although users' persuasion knowledge did not appear to be activated by exposure to our disclosures, the relationship between persuasion knowledge and cognitive resistance implies that when users are aware that they are being persuaded, they are more likely to resist the message by counterarguing with it (Fransen, Smit, et al., 2015; Fransen, Verlegh, et al., 2015; Youn and Kim, 2019). Hence, individuals who are aware that a political message is an advertisement could have more doubts about the content of the message and rather reject it. For instance, they might think that the message is not honest and is only aimed at attracting voters. Moreover, political advertising can be considered as a high-involvement process. This is in contrast to the work by Van Reijmersdal and colleagues (2016), where the product used in the

stimulus material was a rather low-involvement product (i.e., a casserole mix). Political advertising could be seen as an issue in which users experience more involvement, which might explain why they actively counterargue with the message.

The proposed relationship between persuasion knowledge and affective resistance was significant, but lower than our threshold for meaningful interpretation. This implies that users counterargue with the message, but in our sample, they might not have experienced feelings like anger to the same extent. At the same time, we argue that this finding can also be a result of our operationalization. We operationalized cognitive resistance as counterarguing, while operationalizing affective resistance as negative affect (van Reijmersdal et al., 2016). Users may not feel angry because they are targeted. While counterarguing as a psychological reaction to a message is something that users do daily, actually getting angry or annoyed might be a stronger reaction to an advertisement on a platform you receive rather passively.

Regarding users' attitudes towards the political party, we proposed that both dimensions of resistance are negatively related to users' perceived credibility of the source (Housholder and LaMarre, 2014; Van Reijmersdal et al., 2016). Our analyses showed that cognitive resistance is related to a decrease in the perceived credibility of the source, implying that users who counterargue with the message perceive the source as less credible (Carr and Hayes, 2014; Deng et al., 2020; Wojdynski and Evans, 2016). However, there was no evidence regarding affective resistance; it should be noted that users may still experience affective resistance and that it could be part of other psychological processes, as it was positively correlated with cognitive resistance. An explanation for this could be that when users counterargue with a message, they actively process the information, which might lead them to engage in more critical processing of the message, a process that is more likely to be activated through disclosures (Boerman et al., 2014). In turn, this systematic and biased way of processing a message could project on users' perceptions of the party or source. For

instance, users could believe that the source is less sincere because they sent a persuasive message, potentially decreasing their credibility perceptions. In contrast, it is possible that, since there is no activation of affective resistance through persuasion knowledge, users do not have affective reactions towards the message. If this is the case, this will also not project their perceptions of the source and, thus, does not affect their perceived credibility.

In addition, we proposed that both types of resistance would have a negative relationship with attitude towards the politician, which we only found to be true for cognitive resistance. The same reasoning used to explain the relationship between both types of resistance and source credibility might explain these results. However, looking at the strengths of the relationships, we see that the direct relationship between cognitive resistance and source credibility is stronger than the direct relationship with attitude towards the politician, while the relations of affective resistance on both credibility and attitude towards the politician are low and non-significant. A reason for the difference in the strength of the relationship between cognitive resistance and credibility, as well as cognitive resistance and attitude towards the politician, could be that the politician is only advertised in the image of the post, even though the politician is fake. The source of the message, or political party, is named at the top of the post as the sender of the message and might be more visually present to some users because the politician is non-existent and therefore does not lead to the activation of prior attitudes. Moreover, the user may perceive the source of the message as the one presenting the persuasive message, while the politician is 'just' advertised in it because they are a part of that party. This might lead to negative perceptions being activated more in reaction to the party, but less towards the advertised politician.

Additionally, as expected, we found a relationship between users' perceived credibility of the source and their attitude towards the advertised politician (Funk, 1999; Madsen, 2019, p. 107; Markus, 1982). This relation could be explained as an extension of the

rationale used above, while the negative perceptions towards the source might not project on the advertised politician as much, possibly because the opposite is true as well. In that case, it is possible that when users have more favorable credibility perceptions towards a party, the politician in the advertisement benefits from that. However, this interpretation might be possible but is on a speculative basis, since we used a fake politician and a fake party, and this might have influenced our results. While the use of a fake party and politician allows research to isolate the effects of manipulations in experiments and prevent confounding of existing attitudes, we recognize that this also makes it harder to translate results to a real-world setting because of a decrease in ecological validity. For instance, the relationship between resistance and source credibility might be less strong than it could have been with an existing political party and political actor, as there would be existing attitudes that form over time (Madsen, 2019; Markus, 1982). On the contrary, it is also possible that since the political party was unknown to participants before this experiment, this led to an overestimation of our results, as participants might have had more systematic processing of the advertisement through a novelty effect, which could have led to participants thinking too much about the political party and disregarding the disclosures.

Finally, we found that cognitive and affective resistance were positively correlated, although affective resistance was not related to persuasion knowledge or source credibility. We recognize that participants saw our manipulation in an experimental setting (i.e., a timer before they could click through to the next page, forcing longer exposure), which might have led them to use more cognitive resources to process the message, including the disclosure, than they would have used in a less superficial setting. In a real-world setting, SNS users are exposed to many messages in a short amount of time, which could lead to less central processing of the messages, including the disclosures on these messages, making it harder to process transparency measures if platforms provide them (Petty and Cacioppo, 1986).

Concerning disclosures as possible measures, platforms, as well as researchers, and eventually users would benefit from a better understanding of the most effective designs concerning transparency and information (Amazeen and Wojdynski, 2020). This study shows mixed results concerning these disclosures. On the one hand, disclosures are not always recalled correctly; on the other hand, disclosures are recalled, which highlights the potential of disclosures as a transparency measure, in line with other research (Binford et al., 2021; Kruikemeier et al., 2016; Van Reijmersdal et al., 2016). Furthermore, we notice that while disclosures are not always perceived and recalled in experimental studies, this might match a real-world setting. People focus more on content while scrolling Facebook timelines and perhaps less on disclosures on this content. The question arises if users consciously or unconsciously process content on their timelines, as described in the Elaboration Likelihood Model (Petty and Cacioppo, 1984). From a usage perspective, it would make sense to focus more on the actual content; for instance, users look at a video or image in their timeline before anything else (Vergara et al., 2021), which makes it even more important to ascertain that disclosures are designed fittingly: informative and salient, but not too invasive.

Altogether, our findings imply that resisting messages, as a possible result of high persuasion knowledge, is a cognitively rather than affectively shaped process. Moreover, only the cognitive processes were negatively related to both source credibility and attitude towards the advertised politician. This means that resisting persuasive attempts and attitude formation appear to be rather conscious and controlled processes that are not associated with high negative affect. Our findings underline the importance of awareness concerning microtargeting, as this seems to be a precondition for challenging the content, source, and politician of a political advertisement. Microtargeted users might resist messages more if they are aware of the targeting that occurs. In addition, this resistance possibly not only decreases perceived credibility, but also users' attitude towards the advertised politician. For political

parties and advertisers working with them, this implies that using PMT should be a well-considered choice, as it might be counterproductive for affecting voters (Hetherington, 1999; Housholder & LaMarre, 2014).

Limitations and future work

In order to not influence the participants of this study beyond this experiment, we made use of a made-up political actor and party. As described, political behavior and actual voting, while being influenced by perceived credibility and attitudes towards politicians, consist of many factors and not just exposure to online political advertisements. While avoiding ethical risks, we recognize a potential novelty effect, as participants had never seen this party or politician before. Therefore, their focus could have been more on the politician and the party, recalling the disclosure worse than they possibly would in a real-world setting, even with our 15-second timer. Future research implementing objective measures such as eye tracking would describe users' exposure to the actual disclosure better, as we had no control over that besides the timer. Furthermore, the domain of disclosures from a policy and transparency perspective would benefit from investigating whether users do not recall disclosures due to a lack of interest or an actual lack of exposure or recognition. Regarding the measures for source credibility and attitude towards the politician, there is a possibility that, even though we explicitly stated that source credibility concerned the party and attitude towards the politician concerned the actual politician, there might be overlap in these measurements. An experiment that uses existing political advertisements might be able to elaborate on the potential differences between attitudes towards a party and attitudes towards a politician.

In this experiment, we tried to simulate microtargeting by showing participants statements that were close to their own beliefs. However, we recognize that our measurements of counterarguing as a measure of cognitive resistance possibly come from

users disagreeing with the statement despite our efforts to match the statement to their previous answers. Future research can benefit from other ways of tailoring stimuli to users' beliefs while investigating PMT, for instance, by asking users more about those beliefs and using a larger battery of statements for stimuli.

In this study, we choose to compare our targeting disclosure with the disclosure that Facebook currently uses as a control group. While there is no possibility of advertising on the platform without the advertisement containing this disclosure, we still recognize that future work could scrutinize potential differences between an organic post (i.e., without a disclosure), an advertisement with the current disclosure, and a targeting disclosure to isolate potential differences. Furthermore, we recognize that the process of political communication is not isolated to the measures we used in this study, and we encourage other researchers to investigate user behavior regarding transparency and advertising on SNS through other measurements like for instance, such as ad avoidance, blocking behavior, and engagement measures such as commenting and sharing.

Moreover, to increase ecological validity, research focusing on PMT would benefit from field experiments using existing political advertising. This could be possible by having users install a browser plugin that recognizes political advertisements on their Facebook timeline to better indicate exposure. This would also have the possibility to conduct a study with a longitudinal nature, which could control for party preferences at the beginning, so possible fluctuations in political attitudes could be investigated, and the advertisements might be targeted at users to fit their beliefs and attitudes from the start. In addition, a field experiment might be able to test users' attitudes and the potential fluctuations in those attitudes due to exposure to targeted advertisements. Simultaneously, a field experiment would be able to introduce other parameters in the dynamic aspect of the disclosure, ensuring

that users are aware of which different data points are used to target them, compared to our targeting parameters (age, gender, and online behavior).

Furthermore, in line with central and peripheral processing as described above, we forced participants to view the message including the disclosure for 15 seconds, while this was done to make sure they did not click through to the next page immediately, we recognize that this potentially leads to an overestimation of potential effects of exposure to a disclosure as a real-world setting might find less strong relations, due to a less forced and superficial setting. Finally, apart from the effect of the manipulation on persuasion knowledge, this study has a cross-sectional design; therefore, we are not able to investigate causal relations, meaning that the other paths in our model are bi-directional.

Conclusion

Online PMT is a technique that for most users occurs in the background. While there seems to be no consensus regarding the effects of the technique, research on transparency concerning it is also scarce. As calls for transparency regarding the use of private information for microtargeted ads have become louder, the present study investigated the effectiveness of a disclosure label to provide users with more information. We tested an integrative model that was meant to depict one possible way in which receivers of microtargeted ads might react to the content. Although the disclosure label did not have a meaningful effect on users' persuasion knowledge, persuasion knowledge was positively related to cognitive (but not affective) resistance to the ad. This shows that people begin to question the content of messages when they are aware that a political message is an advertisement. Moreover, people who challenged the content of the message attributed lower credibility to the source of the message and had a more negative attitude towards the politician. Together, these findings show that the reaction to microtargeted political content is a cognitive rather than an affective process. Although the exact role of targeting disclosures still seems ambiguous, it seems an

important first step in making people aware that they are being targeted based on their personal information. Therefore, the present study is an important contribution in showing that social media users react negatively to political microtargeting when they are aware of the use of their personal data.

References

Amazeen, M. A., and Wojdynski, B. W. (2020). The effects of disclosure format on native advertising recognition and audience perceptions of legacy and online news publishers. *Journalism*, *21*(12), 1965–1984. https://doi.org/10.1177/1464884918754829

Baik, M., Suk, H.-J., Lee, J., and Choi, K. (2013). *Investigation of eye-catching colors using eye tracking* (B. E. Rogowitz, T. N. Pappas, and H. De Ridder, Eds.; p. 86510W). https://doi.org/10.1117/12.2001141

Banerjee, S. C., and Greene, K. (2012). Role of Transportation in the Persuasion Process: Cognitive and Affective Responses to Antidrug Narratives. *Journal of Health Communication*, *17*(5), 564–581. https://doi.org/10.1080/10810730.2011.635779

Beckert, J., and Koch, T. (2022). Why context matters in a changing social media environment: The role of placement prominence as a moderator of disclosure effects in sponsored influencer content. *Studies in Communication Sciences*, *22*(1). https://doi.org/10.24434/j.scoms.2022.01.3061

Beckert, J., Koch, T., Viererbl, B., and Schulz-Knappe, C. (2021). The disclosure paradox: How persuasion knowledge mediates disclosure effects in sponsored media content. *International Journal of Advertising*, *40*(7), 1160–1186. https://doi.org/10.1080/02650487.2020.1859171

Binder, A., Stubenvoll, M., Hirsch, M., and Matthes, J. (2022). Why Am I Getting This Ad? How the Degree of Targeting Disclosures and Political Fit Affect Persuasion Knowledge, Party Evaluation, and Online Privacy Behaviors. *Journal of Advertising*, 1–17. https://doi.org/10.1080/00913367.2021.2015727

Binford, M. T., Wojdynski, B. W., Lee, Y.-I., Sun, S., and Briscoe, A. (2021). Invisible transparency: Visual attention to disclosures and source recognition in Facebook

political advertising. *Journal of Information Technology & Politics*, *18*(1), 70–83. https://doi.org/10.1080/19331681.2020.1805388

Bodó, B., Helberger, N., and De Vreese, C. H. (2017). Political micro-targeting: A manchurian candidate or just a dark horse? *Internet Policy Review*, *6*(4). https://doi.org/10.14763/2017.4.776

Boerman, S. C., Van Reijmersdal, E. A., and Neijens, P. C. (2012). Sponsorship Disclosure: Effects of Duration on Persuasion Knowledge and Brand Responses. *Journal of Communication*, *62*(6), 1047–1064. https://doi.org/10.1111/j.1460-2466.2012.01677.x

Boerman, S. C., Van Reijmersdal, E. A., and Neijens, P. C. (2014). Effects of Sponsorship Disclosure Timing on the Processing of Sponsored Content: A Study on the Effectiveness of European Disclosure Regulations: EFFECTS OF SPONSORSHIP DISCLOSURE TIMING. *Psychology & Marketing*, *31*(3), 214–224. https://doi.org/10.1002/mar.20688

Boerman, S. C., Van Reijmersdal, E. A., and Neijens, P. C. (2015). Using Eye Tracking to Understand the Effects of Brand Placement Disclosure Types in Television Programs. *Journal of Advertising*, *44*(3), 196–207. https://doi.org/10.1080/00913367.2014.967423

Bohner, G., Crow, K., Erb, H.-P., and Schwarz, N. (1992). Affect and persuasion: Mood effects on the processing of message content and context cues and on subsequent behaviour. *European Journal of Social Psychology*, *22*(6), 511–530. https://doi.org/10.1002/ejsp.2420220602

Cadwalladr, C. (2018, March 18). ‘I made Steve Bannon’s psychological warfare tool’: Meet the data war whistleblower. *The Guardian*.

https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump

Cain, R. M. (2011). Embedded Advertising on Television: Disclosure, Deception, and Free Speech Rights. *Journal of Public Policy & Marketing*, *30*(2), 226–238. https://doi.org/10.1509/jppm.30.2.226

Carr, C. T., and Hayes, R. A. (2014). The Effect of Disclosure of Third-Party Influence on an Opinion Leader's Credibility and Electronic Word of Mouth in Two-Step Flow. *Journal of Interactive Advertising*, *14*(1), 38–50. https://doi.org/10.1080/15252019.2014.909296

Deng, X., Li, M., and Suh, A. (2020). Recommendation or advertisement? The influence of advertising-disclosure language with pictorial types on influencer credibility and consumers' brand attitudes. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12427 LNCS*, 234–248. https://doi.org/10.1007/978-3-030-60152-2_19

Di Plinio, S., Aquino, A., Haddock, G., Alparone, F. R., and Ebisch, S. J. H. (2022). Brain and behavioral contributions to individual choices in response to affective–cognitive persuasion. *Cerebral Cortex*, bhac213. https://doi.org/10.1093/cercor/bhac213

Digital Services Act, (2022). https://eur-lex.europa.eu/eli/reg/2022/2065/oj

Dillard, J. P., and Shen, L. (2005). On the Nature of Reactance and its Role in Persuasive Health Communication. *Communication Monographs*, *72*(2), 144–168. https://doi.org/10.1080/03637750500111815

Dobber, T., Kruikemeier, S., Helberger, N., and Goodman, E. (2023). Shielding citizens? Understanding the impact of political advertisement transparency information. *New Media & Society*. https://doi.org/10.1177/14614448231157640

Dobber, T., Metoui, N., Trilling, D., Helberger, N., and de Vreese, C. (2021). Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes? *The International Journal of Press/Politics*, *26*(1), 69–91. https://doi.org/10.1177/1940161220944364

Edell, J. A., and Burke, M. C. (1987). The Power of Feelings in Understanding Advertising Effects. *Journal of Consumer Research*, *14*(3), 421. https://doi.org/10.1086/209124

European Commission. (2022). *Regulation of the European Parliament and of The Council on a Single Market For Digital Services (Digital Services Act) and amending Directive*. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN

Evans, N. J., Phua, J., Lim, J., and Jun, H. (2017). Disclosing Instagram Influencer Advertising: The Effects of Disclosure Language on Advertising Recognition, Attitudes, and Behavioral Intent. *Journal of Interactive Advertising*, *17*(2), 138–149. https://doi.org/10.1080/15252019.2017.1366885

Fransen, M. L., Smit, E. G., and Verlegh, P. W. J. (2015). Strategies and motives for resistance to persuasion: An integrative framework. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01201

Fransen, M. L., Verlegh, P. W. J., Kirmani, A., and Smit, E. G. (2015). A typology of consumer strategies for resisting advertising, and a review of mechanisms for countering them. *International Journal of Advertising*, *34*(1), 6–16. https://doi.org/10.1080/02650487.2014.995284

Friestad, M., and Wright, P. (1994). The Persuasion Knowledge Model: How People Cope with Persuasion Attempts. *Journal of Consumer Research*, *21*(1), 1–31. https://doi.org/10.1086/209380

Funk, C. L. (1999). Bringing the Candidate into Models of Candidate Evaluation. *The Journal of Politics*, *61*(3), 700–720. https://doi.org/10.2307/2647824

Gandy, O. H. (2000). Dividing Practices: Segmentation and Targeting in the Emerging Public Sphere. In W. L. Bennett and R. M. Entman (Eds.), *Mediated Politics* (1st ed., pp. 141–159). Cambridge University Press. https://doi.org/10.1017/CBO9780511613852.008

Giasson, T., Le Bars, G., and Dubois, P. (2019). Is Social Media Transforming Canadian Electioneering? Hybridity and Online Partisan Strategies in the 2012 Quebec Election. *Canadian Journal of Political Science*, *52*(2), 323–341. https://doi.org/10.1017/S0008423918000902

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2014). *Multivariate data analysis* (7. ed., Pearson new internat. ed). Pearson.

Ham, C. D., Nelson, M. R., and Das, S. (2015). How to measure persuasion knowledge. *International Journal of Advertising*, *34*(1), 17–53. https://doi.org/10.1080/02650487.2014.994730

Hetherington, M. J. (1999). The Effect of Political Trust on the Presidential Vote, 1968–96. *American Political Science Review*, *93*(2), 311–326. https://doi.org/10.2307/2585398

Housholder, E. E., and LaMarre, H. L. (2014). Facebook PolTowards Toward a Process Model for Achieving Political Source Credibility Through Social Media. *Journal of Information Technology & Politics*, *11*(4), 368–382. https://doi.org/10.1080/19331681.2014.951753

Jansen, M.-P., and Krämer, N. C. (2023). Empty Transparency? The Effects on Credibility and Trustworthiness of Targeting Disclosure Labels for Micro-Targeted Political Advertisements. *Weizenbaum Journal of the Digital Society*, *3*(1). https://doi.org/10.34669/WI.WJDS/3.1.5

Jost, P., Kruschinski, S., Sülflow, M., Haßler, J., and Maurer, M. (2022). Invisible transparency: How different types of ad disclaimers on Facebook affect whether and

how digital political advertising is perceived. *Policy & Internet*, poi3.333. https://doi.org/10.1002/poi3.333

Jung, A. R., and Heo, J. (2019). Ad Disclosure vs. Ad Recognition: How Persuasion Knowledge Influences Native Advertising Evaluation. *Journal of Interactive Advertising*, *19*(1), 1–14. https://doi.org/10.1080/15252019.2018.1520661

Knowles, E. S., and Linn, J. A. (Eds.). (2004). *Resistance and persuasion*. Lawrence Erlbaum Associates.

Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Kruikemeier, S., Sezgin, M., and Boerman, S. C. (2016). Political Microtargeting: Relationship between Personalized Advertising on Facebook and Voters' Responses. *Cyberpsychology, Behavior, and Social Networking*, *19*(6), 367–372. https://doi.org/10.1089/cyber.2015.0652

Kumkale, G. T., Albarracín, D., and Seignourel, P. J. (2010). The Effects of Source Credibility in the Presence or Absence of Prior Attitudes: Implications for the Design of Persuasive Communication Campaigns: SOURCE CREDIBILITY AND PERSUASION. *Journal of Applied Social Psychology*, *40*(6), 1325–1356. https://doi.org/10.1111/j.1559-1816.2010.00620.x

Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, *8*(1), 33267. https://doi.org/10.1525/collabra.33267

Lakens, D., Scheel, A. M., and Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Madsen, J. K. (2019). *The psychology of micro-targeted election campaigns*. Palgrave Macmillan.

Markus, G. B. (1982). Political Attitudes during an Election Year: A Report on the 1980 NES Panel Study. *American Political Science Review*, *76*(3), 538–560. https://doi.org/10.1017/S0003055400188392

Matz, S. C., Kosinski, M., Nave, G., and Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48), 12714–12719. https://doi.org/10.1073/pnas.1710966114

Miller, A. H., Wattenberg, M. P., and Malanchuk, O. (1986). Schematic Assessments of Presidential Candidates. *American Political Science Review*, *80*(2), 521–540. https://doi.org/10.2307/1958272

Montoya, R. M., Horton, R. S., Vevea, J. L., Citkowicz, M., and Lauber, E. A. (2017). A re-examination of the mere exposure effect: The influence of repeated exposure on recognition, familiarity, and liking. *Psychological Bulletin*, *143*(5), 459–498. https://doi.org/10.1037/bul0000085

Murray, G. R., and Scime, A. (2010). Microtargeting and electorate segmentation: Data mining the American National Election Studies. *Journal of Political Marketing*, *9*(3), 143–166. https://doi.org/10.1080/15377857.2010.497732

Oh, J., and Sundar, S. S. (2015). How Does Interactivity Persuade? An Experimental Test of Interactivity on Cognitive Absorption, Elaboration, and Attitudes: Persuasive Effects of Interactivity. *Journal of Communication*, *65*(2), 213–236. https://doi.org/10.1111/jcom.12147

Ohanian, R. (1990). Construction and Validation of a Scale to Measure Celebrity Endorsers' Perceived Expertise, Trustworthiness, and Attractiveness. *Journal of Advertising*, *19*(3), 39–52.

Ostrom, T. M. (1969). The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology*, *5*(1), 12–30. https://doi.org/10.1016/0022-1031(69)90003-1

Petty, R. E., and Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, *46*(1), 69–81. https://doi.org/10.1037/0022-3514.46.1.69

Petty, R. E., and Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205. https://doi.org/10.1016/S0065-2601(08)60214-2

Pornprasertmanit, S., Miller, P., Schoemann, A., Jorgensen, T. D., and Quick, C. (2021). *simsem: SIMulated Structural Equation Modeling* (0.5-16) [Computer software]. https://CRAN.R-project.org/package=simsem

Quinn, J. M., and Wood, W. (2004). Forewarnings of Influence Appeals: Inducing Resistance and Acceptance. In *Resistance and persuasion.* (pp. 193–213). Lawrence Erlbaum Associates Publishers.

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rosenberg, M. J., Hovland, C. I., McGuire, W. J., Abelson, R. P., and Brehm, J. W. (1960). *Attitude organization and change: An analysis of consistency among attitude components. (Yales studies in attitude and communication.), Vol. III* (pp. x, 239). Yale Univer. Press.

Rosselli, F., Skelly, J. J., and Mackie, D. M. (1995). Processing Rational and Emotional Messages: The Cognitive and Affective Mediation of Persuasion. *Journal of Experimental Social Psychology*, *31*(2), 163–190. https://doi.org/10.1006/jesp.1995.1008

Sagarin, B. J., and Cialdini, R. B. (2004). Creating Critical Consumers: Motivating Receptivity by Teaching Resistance. In *Resistance and persuasion.* (pp. 259–282). Lawrence Erlbaum Associates Publishers.

Sagarin, B. J., Cialdini, R. B., Rice, W. E., and Serna, S. B. (2002). Dispelling the illusion of invulnerability: The motivations and mechanisms of resistance to persuasion. *Journal of Personality and Social Psychology*, *83*(3), 526–541. https://doi.org/10.1037/0022-3514.83.3.526

Schwarz, N., Bless, H., and Bohner, G. (1991). Mood and Persuasion: Affective States Influence the Processing of Persuasive Communications. In *Advances in Experimental Social Psychology* (Vol. 24, pp. 161–199). Elsevier. https://doi.org/10.1016/S0065-2601(08)60329-9

See, Y. H. M. (2018). Cognitive-Affective Processing System. In V. Zeigler-Hill and T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1–7). Springer International Publishing. https://doi.org/10.1007/978-3-319-28099-8_1788-2

Segijn, C. M., Opree, S. J., and Van Ooijen, I. (2022). The validation of the Perceived Surveillance Scale. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *16*(3), Article 3. https://doi.org/10.5817/CP2022-3-9

Segijn, C. M., Van Ooijen, I. (2022). Differences in consumer knowledge and perceptions of personalized advertising: Comparing online behavioural advertising and synced advertising. *Journal of Marketing Communications*, *28*(2), 207–226. https://doi.org/10.1080/13527266.2020.1857297

Stubb, C., and Colliander, J. (2019). “This is not sponsored content” – The effects of impartiality disclosure and e-commerce landing pages on consumer responses to social media influencer posts. *Computers in Human Behavior*, *98*, 210–222. https://doi.org/10.1016/j.chb.2019.04.024

The jamovi project. (2022). *Jamovi* (2.3) [Computer software]. http://www.jamovi.org

Tormala, Z. L., Briñol, P., and Petty, R. E. (2006). When credibility attacks: The reverse impact of source credibility on persuasion. *Journal of Experimental Social Psychology*, *42*(5), 684–691. https://doi.org/10.1016/j.jesp.2005.10.005

Tormala, Z. L., and Petty, R. E. (2002). What doesn’t kill me makes me stronger: The effects of resisting persuasion on attitude certainty. *Journal of Personality and Social Psychology*, *83*(6), 1298–1313. https://doi.org/10.1037/0022-3514.83.6.1298

Tormala, Z. L., and Petty, R. E. (2004). Resistance to Persuasion and Attitude Certainty: The Moderating Role of Elaboration. *Personality and Social Psychology Bulletin*, *30*(11), 1446–1457. https://doi.org/10.1177/0146167204264251

Van Reijmersdal, E. A., Brussee, E., Evans, N., and Wojdynski, B. W. (2023). Disclosure-Driven Recognition of Native Advertising: A Test of Two Competing Mechanisms. *Journal of Interactive Advertising*, 1–13. https://doi.org/10.1080/15252019.2022.2146991

Van Reijmersdal, E. A., Fransen, M. L., Van Noort, G., Opree, S. J., Vandeberg, L., Reusch, S., Van Lieshout, F., and Boerman, S. C. (2016). Effects of Disclosing Sponsored Content in Blogs: How the Use of Resistance Strategies Mediates Effects on Persuasion. *American Behavioral Scientist*, *60*(12), 1458–1474. https://doi.org/10.1177/0002764216660141

Van Reijmersdal, E. A., Opree, S. J., and Cartwright, R. F. (2021). Brand in focus: Activating adolescents' persuasion knowledge using disclosures for embedded advertising in music videos. *Communications*. https://doi.org/10.1515/commun-2019-0168

Vergara, A., Siles, I., Castro, A. C., and Chaves, A. (2021). The Mechanisms of "Incidental News Consumption": An Eye Tracking Study of News Interaction on Facebook. *Digital Journalism*, *9*(2), 215–234. https://doi.org/10.1080/21670811.2020.1813047

Wojdynski, B. W., and Evans, N. J. (2016). Going Native: Effects of Disclosure Position and Language on the Recognition and Evaluation of Onlin. *Journal of Advertising*, *45*(2), 157–168. https://doi.org/10.1080/00913367.2015.1115380

Wojdynski, B. W., and Evans, N. J. (2020). The Covert Advertising Recognition and Effects (CARE) model: Processes of persuasion in native advertising and other masked formats. *International Journal of Advertising*, *39*(1), 4–31. https://doi.org/10.1080/02650487.2019.1658438

Yoon, K., Pinkleton, B. E., and Ko, W. (2005). Effects of Negative Political Advertising on Voting Intention: An Exploration of the Roles of Involvement and Source Credibility in the Development of Voter Cynicism. *Journal of Marketing Communications*, *11*(2), 95–112. https://doi.org/10.1080/1352726042000315423

Youn, S., and Kim, S. (2019). Newsfeed native advertising on Facebook: Young millennials' knowledge, pet peeves, reactance and ad avoidance. *International Journal of Advertising*, *38*(5), 651–683. https://doi.org/10.1080/02650487.2019.1575109

Zarouali, B., Dobber, T., De Pauw, G., and de Vreese, C. (2020). Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media. *Communication Research*. https://doi.org/10.1177/0093650220961965

Zuiderveen Borgesius, F. J. (2016). Singling out people without knowing their names—Behavioural targeting, pseudonymous data, and the new Data Protection Regulation. *Computer Law and Security Review*, *32*(2), 256–271. https://doi.org/10.1016/j.clsr.2015.12.013

Zuwerink Jacks, J., and Cameron, K. A. (2003). Strategies for Resisting Persuasion. *Basic and Applied Social Psychology*, *25*(2), 145–161. https://doi.org/10.1207/S15324834BASP2502_5

Article III

Jansen, M.-P., & Krämer, N.C. (2023). Balancing perceptions of targeting: An investigation of political microtargeting transparency through a calculus approach. *PLOS ONE, 18*(12), e0295329. https://doi.org/10.1371/journal.pone.0295329

PLOS ONE

RESEARCH ARTICLE

Balancing perceptions of targeting: An investigation of political microtargeting transparency through a calculus approach

Martin-Pieter Jansen[1]*, **Nicole C. Krämer**[1,2]

1 Social Psychology: Media and Communication, University of Duisburg-Essen, Duisburg, Germany, **2** Research Center Trustworthy Data Science and Security, Dortmund, Germany

* martin-pieter.jansen@uni-due.de

OPEN ACCESS

Citation: Jansen M-P, Krämer NC (2023) Balancing perceptions of targeting: An investigation of political microtargeting transparency through a calculus approach. PLoS ONE 18(12): e0295329. https://doi.org/10.1371/journal.pone.0295329

Editor: Pierluigi Vellucci, Roma Tre University: Universita degli Studi Roma Tre, ITALY

Received: March 31, 2023

Accepted: November 16, 2023

Published: December 7, 2023

Data Availability Statement: All data files, code and stimuli are available from the OSF database (url: https://osf.io/2rbqu/).

Funding: We acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Over the last few years, political advertisers have moved with their audiences: to social media platforms. Advertisers on these platforms aim to persuade voters by sending messages tailored to them based on their own data: political microtargeting (PMT). A considerable problem with PMT is that users are often unaware that they are being targeted, while current transparency advances do not seem to suffice in informing users. However, increasing transparency may have consequences on users' privacy perceptions. Thus, the current work investigates whether disclosures, as a measure to increase transparency, increase users' recognition of a microtargeted ad, and subsequently what this means for their perceived benefits, privacy concerns, and their likelihood of engaging in privacy protection behavior, based on the privacy calculus. In a preregistered online one-factorial between-subjects experiment ($N = 450$) we exposed participants to either an Instagram post containing a currently used disclosure or a more salient disclosure. Our results show that exposure to this disclosure increases recognition of the ad being microtargeted, and that this relates to perceived benefits but not privacy concerns. However, the results show that users' privacy concerns are related to their increased privacy protection behavior. Finally, we found that over four-fifths of our participants who were exposed to the more salient disclosure recalled it correctly.

Introduction

A by-product of people's use of social networking sites (SNS) is the data they leave behind. These digital breadcrumbs, left behind through liking posts, commenting on videos, and simply interacting with or viewing content, are valuable to advertisers and political consultancy agencies. Through these breadcrumbs or data points, they build profiles of users and allocate them into small groups while narrowly targeting them with specific messages that are developed to resonate most effectively within these groups, a practice commonly titled *microtargeting* or in a political context *political microtargeting* (PMT) [1–4].

PMT can not only be used to persuade voters, but also to discourage political participation, potentially negatively affecting voter turnout, which damages the democratic process [5, 6].

Moreover, sending messages to parts of the public while nobody else, but that part of the public knows of the existence of that message avoids scrutiny in the democratic process as well, harming it instead [7]. Besides, campaigns could base their actions on personal data, while users are not aware of the information used to target them, if they are even aware that they are being targeted at all, potentially opening the door for voter manipulation [8]. However, PMT also provides opportunities for voter mobilization [4], increasing political interest [8], and supporting voters to access relevant information [9]. Recent work [10] finds both positive and negative effects and speaks of a paradox of PMT, where the technique might benefit individuals through usefulness but might be harmful for society at large.

In the most commonly known example of political targeting British political consultancy firm, *Cambridge Analytica* presumably gathered and used data from more than 50 million Facebook users to form psychological profiles and target users with messages that would persuade them as strongly as possible [11]. This allegedly contributed to Trump's presidential victory and the Leave campaign's success in the Brexit referendum. However, critics claim that users' privacy is at risk here and that this is mostly without their informed consent [12–14].

A major problem with PMT is that users do not recognize the targeted, persuasive attempts that campaigning agencies and parties use to try to affect them [2, 15]. A potential solution to this problem is the use of disclosure labels. Platforms such as Facebook, Instagram, and Twitter have been using these labels to be transparent and to provide users with a tool to distinguish sponsored posts or advertisements from regular content on the platform. Marketing and advertising research has investigated these labels for a long time [16–18]. In the context of PMT, however, research on the labels is upcoming but still scarce [1, 2, 19, 20]. Disclosures have been found to increase users' knowledge and perceptions of persuasive attempts, which can be investigated through the Persuasion Knowledge Model [21], or in a PMT context through targeting knowledge, which is based on the Persuasion Knowledge Model and adapted to the context of targeting [1].

One of the theoretical frameworks that helps us understand users' privacy perception is the privacy calculus by Culnan and Armstrong [22]. The privacy calculus assumes that users' behavior regarding their privacy is influenced by their perceived benefits and perceived costs or risks. If users are aware of targeting that occurs, this might lead to different perceptions concerning their online privacy, as has been found in research on advertising personalization [23]. To the best of our knowledge, an empirical investigation of the privacy calculus in the sphere of PMT has not yet been conducted.

Prior research on PMT has mainly focused on Meta's Facebook as a platform [1, 2, 20]. While the company's other major platform, Instagram, has grown faster than Facebook [24], an empirical investigation of the use of targeting disclosures on Instagram has not been conducted. Since the platforms have different focuses on content and users might have different needs to fulfill while using them, there might be different effects of these disclosures.

Therefore, the current study investigates two different disclosures: a regular disclosure that is currently used on Instagram showing that content is sponsored and a more elaborate and salient disclosure that is a partial replica of the disclosures that were used on Instagram to inform users about potential fake news regarding COVID-19 information. We investigate the effects of these two disclosures on users targeting knowledge. Moreover, we will focus on users' privacy perceptions by investigating a micro-level perspective on the privacy calculus, where perceived benefits and privacy concerns potentially lead to privacy protection behavior.

Theoretical background

One of the major problems regarding PMT is the presumed black box, where neither laypersons, journalists, nor researchers know exactly what political consultancy firms do to microtarget users and how and which data are used. Although this black box makes it difficult to research PMT, it also makes it practically impossible to estimate the effects of campaigns run by firms and parties. Because the lack of transparency and information about the models used makes it difficult to focus on the sender side of PMT, one solution could lie within the user side of this way of campaigning. Regulatory bodies are starting to implement regulations to provide users with appropriate transparency measures. For example, the EU Digital Services Act states that targeted information and advertisements should include information on when and on whose behalf, content is displayed. Simultaneously, these measures should go beyond that and include background information about individualized data used to target users and parameters for those data points [25, 26].

Disclosures

The way the EU Digital Services Act aims to have platforms implement disclosures is the first time that targeting disclosures have been initiated in Germany. Currently, the 'fake news' bill, that also includes targeting transparency on platforms, in Brazil is being reviewed in congress [27]. Within the EU, earlier regulations were proposed in Ireland [28] and adopted in France [29, 30]. Earlier versions of disclosures only aimed to inform users that content is sponsored (e.g., YouTube, Instagram, and Facebook), and in some cases who the party paying for the sponsored content is (e.g., Twitter and political advertisements on Facebook). Research on the application of disclosures in a microtargeting context has emerged for a couple of years and shows mixed results regarding the recall of disclosures and their effectiveness in helping users recognize political advertising [1, 2, 20]. Nonetheless, one of the main implications of disclosures is that the prominence, position, and degree to which they stand out in contrast to a user's timeline are important [16, 31]. Furthermore, the combination of text and a symbol leads to the highest visual fixation [17]. While being more prominent and therefore standing out compared to regular content, disclosures lead to higher levels of attention and, in turn, to better perception of the disclosed information, and thus contribute to transparency [1, 2]. The status quo for platforms seems to be a short gray sentence stating that content is sponsored or sponsored, and by whom it is paid for. However, this status quo seems to be insufficient in informing users, thus contributing to transparency, as most of the disclosures that are currently used on platforms lead to low levels of recallment [1, 19, 20].

Since 2021, differently designed disclosures have been used on various platforms. During the COVID-19 pandemic, both Instagram and Facebook took measures to counter the spread of misinformation and to inform users. One of these measures was the use of disclosures on posts and stories (disappearing posts) on the platform that guided users with both information to directly inform them that information could be false, or was not yet proven, and a link to Instagram's 'COVID-19 Information Center', where users could find credible information about COVID-19 in general [32, 33].

To the best of our knowledge, these new disclosures have not been investigated in a scientific setting, nor are there results or information publicly available from Meta. It seems promising that the platform experiments with different types of disclosures. From a business perspective, it would make sense to not label advertisements on platforms as large as the platforms did with the COVID-19 information because selling room for advertisements is a major part of their business models. However, since an increasing number of legislations seem to be upcoming (e.g., the EU Digital Services Act), using these already existing disclosures might be

a great solution for platforms to oblige to these legislations. Therefore, the current work will use Instagram's existing sponsored disclosure as a control condition while also going beyond that and consequently investigate disclosures that inform users about targeting.

Targeting knowledge

Persuasion is something that users are exposed to every single day, both in the 'real world' and on social networking sites. To cope with persuasive messages, users develop beliefs and knowledge about persuasion based on their previous experiences with and exposure to advertisements: *persuasion knowledge* [21]. Existing work on disclosures shows that if a message contains a disclosure, it is more likely that users' persuasion knowledge will be activated compared to when a message does not contain a disclosure, in both research on regular advertising [34, 35] and microtargeting [2, 36]. However, as persuasive tactics evolve and move to places where the potential targets of these messages are, online, it is deemed important to also evolve the measurement instruments to reflect the specific context. In recent work on PMT, Binder and colleagues [1] developed a scale for *targeting knowledge* based on earlier work on personalization and the persuasion knowledge model [21, 37, 38]. For this construct, the authors use the following definition: "Individuals' beliefs of agents' use of their online data to tailor messages to recipients" [1]. The authors also underline the importance of the construct covering not only the recognition of targeted advertisements, which is usually lower than that of a regular advertisement [39], but also users' perceptions of targeted messages, as the way the message is designed, constructed, and delivered, in line with the persuasion knowledge model [21]. Binder and colleagues found good reliability for their newly developed measure [1].

Therefore, this study builds on this new concept and attempts to empirically investigate it with a more prominent, salient, and potentially more attention-grabbing disclosure, including both a textual aspect and a symbol informing users that the disclosure contains additional information, which has been found to improve transparency regarding the persuasive or targeted nature of the message [2, 20, 40]. Taking this into account, we expect that a more salient and prominent disclosure increases users' recognition of targeting practices:

H_1: Exposure to a more prominent targeting disclosure, compared to Instagram's regular sponsored disclosure increases targeting knowledge.

Privacy

One of the problems with using data for advertising purposes, as done in microtargeting, is that it crosses boundaries regarding users' privacy on SNS. An example of this is work that shows that online behavioral advertising (OBA), which could be viewed as the precursor of PMT, is often perceived as a privacy risk and intrusive [23]. The same work, however, also sheds light on the perceived benefits of OBA and shows that personal relevance, added advertising value, and economic benefits are the top three perceived benefits in their US national sample. One framework that helps us understand online privacy and users' perceptions of it is Culnan and Armstrong's privacy calculus theory [22]. The privacy calculus theory states that people weigh privacy costs and benefits before they disclose personal information and are more likely to disclose this information if the benefits are at least balanced (if not greater than) the privacy costs [41]. Regarding privacy on social media, this could be seen as the subjective experience of privacy, as described in the model by Trepte [42]. According to this framework, the user experiences access that results from their usage goals on the one hand and the social media boundary conditions and privacy mechanisms on the other. Subsequently, these privacy

perceptions lead to different forms of privacy regulation behaviors, such as altering self-disclosure and control by restricting access to information [42]. Moreover, in empirical work, Dienlin and Metzger [43] showed that in the context of SNS, their extended privacy calculus, including users' self-withdrawal behaviors and privacy self-efficacy, holds true. However, regarding self-disclosure, the benefits outweigh privacy concerns, while in the case of self-withdrawal, privacy concerns outweigh both self-efficacy and benefits. Other work shows that, in the case of app usage concerning a COVID-19 warning app, both privacy concerns and perceived benefits predict app usage [44]. Although there have been other studies that provided a theoretical framework [45], investigating aspects of privacy regarding microtargeting over time [13, 46], and with an adolescent sample [47], an empirical investigation of the privacy calculus in a microtargeting context has not yet been conducted.

In the context of microtargeting, we propose that the calculated rational approach grounded in the privacy calculus can explain users' behavior regarding their online privacy. Users might experience privacy concerns through the use of their personal and behavioral data while also having concerns regarding data protection [1, 13, 23]. Simultaneously, users may experience higher usefulness by receiving political advertisements that match their personal preferences or interests [1, 48]. These experiences and perceptions can coexist and lead to changes in attitudes towards PMT techniques and users' privacy protection behavior. Engaging in more protection behavior could protect them against data usage by platforms and advertisers on the one hand, while on the other hand, engaging less in this behavior might give platforms and advertisers more information and data to work with, potentially leading to better targeted ads and thus higher usefulness. Thus, we aim to investigate political microtargeting through the propositions of the privacy calculus theory while analyzing the perceived benefits and privacy concerns users experience concerning PMT. Subsequently, we scrutinize whether these perceptions and concerns lead to differences in privacy protection behavior.

Perceived benefits of microtargeting

The benefits side of the calculus rationale can be described as the advantages that users receive in exchange for their online data, both self-disclosed and gathered through behavior. In existing work, the benefits of PMT are explained primarily on a societal (or macro) level, and the overall consensus seems to be rather negative. However, PMT may also be beneficial. The use of the technique can, for instance, activate potential voters who are usually deemed to have a lower propensity to vote by reaching out to them in personally relevant ways, with messages that are personally relevant as well [4], which has the potential to strengthen general political interest [8, 10]. Nevertheless, these benefits occur at the macro level. At a personal level, personalization of content leads to higher levels of attention, more accurate recall, and more positive evaluations of content [49]. Other work does not find that higher levels of personalization lead to higher levels of perceived relevance, but this could be a result of the fact that SNS users might not be aware that so much information is gathered concerning them and is used to tailor advertisements to them [50]. Furthermore, in a study on the AdChoices Icon, which could be seen as a version of a disclosure, and personalization Brinson and Eastin [51] found that when consumers, due to a disclosure, recognize an advertisement that contains personalization, they have more favorable attitudes towards that ad.

Moreover, the personal benefits of PMT might lead to societal benefits when a reduction in time and cognitive effort in obtaining information and higher content relevance mobilizes individuals to vote [2]. Furthermore, previous work on online behavioral advertising found that the technique can narrow down alternative solutions to the most relevant and helpful information [9]. Similarly, Zarouali and colleagues [4] conclude that PMT can serve as an

effective way to provide relevant information to citizens on the issues they really care about, which could lead them to be more informed and knowledgeable about these issues. Moreover, personal relevance could lead to higher motivation to process information, which could subsequently lead to a more central attitude change, as described in the Elaboration Likelihood Model (ELM) [52]. Taking the macro-level benefits into account, this study empirically investigates the micro- or individual benefits of PMT, and we expect that users who recognize targeted advertisements experience higher levels of benefit perceptions:

H_2: Targeting knowledge is positively related to perceived benefits

Privacy concerns

As stated above, the overall consensus on PMT and its effect on society appears more negative. Multiple authors underline their concerns on PMT potentially: discouraging political participation [5], decreasing scrutiny in the democratic process by leaving out certain target groups [7], enlarging the gap in representation in governments [6], and manipulating voters by targeting users without their knowledge or consent [8]. These concerns, however, are all on a societal level, meaning that they potentially harm individual users but are not necessarily concerns users have if they are aware of a message being targeted at them.

On an individual level, users perceived privacy concerns regarding personalization and microtargeting; for example, an advertisement is perceived as a privacy risk and intrusive [23]. Other work shows that higher levels of personalization lead to higher perceived *creepiness* of advertisements [50]. Recent work validating a scale on perceived surveillance concerning personalization effects found that users experience creepiness, concerns about surveillance, perceptions of privacy risks overall, and privacy concerns [53]. In addition, other work found that in the case of data-driven OBA, persuasion knowledge, on which the concept of targeting knowledge is built, positively affects privacy risks, which could be considered as the cost side of the privacy calculus theory [54]. Dobber and colleagues [13] found that privacy concerns regarding PMT lead to more negative attitudes towards the technique and reversibly a higher attitude towards the technique leads to a decrease in privacy concerns. Taking this into account, we propose that on an individual or micro level, users who recognize targeting advertisements feel that the technique violates their privacy standards:

H_3: Targeting knowledge is positively related to privacy concerns

Attitude towards the platform

While making users aware of the targeting practices taking place on SNS might lead them to change their perceptions regarding those practices, user attitudes might also influence this relation. In an earlier study on Facebook, Debatin and colleagues [55] found that while users did recognize the potential privacy issues of the platform, they simultaneously uploaded large amounts of personal information, and that this behavior may be explained through high levels of gratifications of using the platform. Furthermore, the mood congruency hypothesis assumes that a recipient's mood state may influence the associations generated during exposure to a message, leading to more positive elaboration of the content or more positive reactions to peripheral cues [56]. Other work shows that in the domain of intrapersonal communication, the medium used to send a message affects its persuasiveness [57]. Moreover, in work on personalized advertisements, De Keyzer and colleagues [50] found that the source type can

mediate the effect of perceived relevance on source attitude. Furthermore, in work on personalization and the privacy calculus, Hayes and colleagues [58] found that the consumer-brand relationship has a positive moderating effect on the benefits side of the privacy calculus, meaning that for users with a stronger consumer-brand relationship, the effect of perceived benefits on the value of information is larger than for users with a weaker consumer-brand relationship. In addition, in the case of adolescents, previous work finds that privacy perception and data protection are positively affected by social media activity [59], meaning that users who find social media use important have higher levels of privacy perception and find data protection more important.

In this study, we scrutinize if the attitude users have towards the platform (Instagram) beforehand, leads to a difference in the relations of targeting knowledge, perceived benefits, and privacy concerns regarding PMT. More specifically, we propose that for users who have a more positive attitude towards Instagram, the relationship between targeting knowledge and perceived benefits is stronger, meaning that if a user is happy with the platform and likes to use it, the microtargeted advertisement might be perceived as more beneficial. Similarly, we propose that for users who have a more negative attitude towards Instagram, the relationship between targeting knowledge and perceived benefits is weaker, meaning that if a user is not happy with the platform and, for instance, is already contemplating leaving it, the microtargeted advertisement might be perceived as less beneficial, just because the user is less happy with Instagram. Leading us to propose the following two moderation hypotheses:

H_{4a}: The relation between targeting knowledge and perceived benefits is moderated by users' attitude towards the platform

H_{4b}: The relation between targeting knowledge and privacy concerns is moderated by users' attitude towards the platform

Privacy protection behavior

As mentioned, the rationale described in the privacy calculus leads users to engage in more or less self-disclosure on social media. However, recent work provides us with a different outcome of the calculation between privacy costs and benefit perceptions: privacy protection behavior [60]. Examples of privacy protection behavior range from altering the privacy settings on a platform or using software to disguise oneself to deregistering from the platform altogether. We deem privacy protection behavior a more fitting outcome measure in the case of PMT because users are not always aware of the information that is used to target them, which is not always self-disclosed. Earlier work found that perceived risks lead to a lower intention to self-disclose and a higher desire for protection, which in turn leads to an intention to use a tool to protect oneself online [60]. Moreover, perceived manipulative intent has been found to increase users' privacy behavior [1]. Additionally, higher levels of privacy concerns lead to higher intentions to withdraw information from Facebook and lower intentions to disclose information on the platform as well [43]. Besides, Büchi and colleagues [61] found that when people feel that their online privacy has been violated, they implement greater privacy protection.

Furthermore, other work on the personalization of advertisements suggests that the perceived costs (or privacy concerns) outweigh the perceived benefits, which we propose leads to engaging in more privacy protection behavior [50]. In a microtargeting context, recent work shows that privacy concerns lead to more privacy protection behavior, but interestingly, it does not lead to users applying ad blockers to block the advertisement [46]. This is in line with

other work, showing that in the case of regular online advertising, opt-out rates via AdChoices (a platform that uses cookies to tailor ads to website users) are 0.26% in the European Union and 0.24% in Germany [62]. This could raise the question of whether there are boundaries for privacy protection behavior, meaning that users will engage in it if the measure is not too technical or too much of a procedure to take.

Conversely to what is explained above, we expect that users that have higher levels of perceived benefits regarding PMT, will engage less in privacy protection behavior, which is in line with a lot of work that shows that perceived benefits lead to more self-disclosure. We recognize that even though the outcome measure might be less fitting for this work, the mechanism behind it might still hold true. For instance, someone who is happy with a tailored ad will be less motivated to disable the ad by altering their advertising preferences on social media. In this study, we investigate privacy protection behavior as the outcome from a privacy calculus perspective on PMT. We propose that perceived benefits are negatively related to privacy protection behavior, meaning that users who perceive high levels of benefits of PMT intend to take fewer precautions to protect their privacy online. Conversely, we propose that privacy concerns are positively related to privacy protection behavior, meaning that users who perceive high levels of privacy concerns concerning PMT intend to engage in more behavior to protect their online privacy:

H_5: Perceived benefits are negatively related to intended privacy protection behavior

H_6: Privacy concerns are positively related to intended privacy protection behavior

Algorithmic user agency

In addition to investigating privacy protection behavior, we propose an exploratory outcome for users who experience the benefits of personalization or targeting. Users' social media feed is filled with content, content from people and companies they follow, but also with content that they are more likely to engage with disregarding the sender. This content is usually targeted at them through an algorithm that knows what they previously engaged with, look at for a longer time, or scrolled back to.

Nevertheless, there is work that shows that users are not as surrendered to algorithms as one might think. In an exploratory study, Kapsch [63] found that some users influence what content they see by interacting with profiles, liking posts, commenting, or even texting via direct messages (DM): *algorithmic user agency*. Users who exercise these techniques try to gain autonomy by actively showing the algorithms of social platforms what they like, which potentially is a proxy for their willingness to actively use algorithms to their advantage. To the best of our knowledge, there has not been work discussing this concept concerning PMT. Using this construct, we aim to better understand users' behavior that is intended to inform the algorithm about their preferences by consciously interacting with content in order to receive recommended content that is tailored better in the future. In the case of PMT, this behavior could lead to even more fitting targeting because the user is actively feeding the algorithm information about their personal preferences. This behavior can be exemplified by liking a cat video on Instagram to see more cat videos in the future. As an exploratory research question, we aim to investigate whether users' perceived benefits of PMT relates to these users trying to influence the algorithm by employing algorithmic agency:

RQ_1: What is the potential relation between users' perceived benefits and behavior that would lead to users actively interacting with certain content: algorithmic user agency?

Method

This study was approved by the ethics committee of the University of Duisburg-Essen (approval number 2211SPJM9646). We preregistered this study before collecting data: [link deleted for peer-review]. All participants agreed with our online informed consent form by checking a box on the page showing the consent form before participation; otherwise, participation was not possible, and participants were redirected to the website of our panel provider (Prolific). Supplementary materials and our measures are publicly accessible on OSF: https://osf.io/2rbqu/?view_only=7923aff35ecd44caa90f7c7603912e03.

Design

To test our hypotheses, we conducted an online experiment using a factorial between-subjects design containing two groups. Participants were exposed to a political Instagram advertisement with a statement aligned with their beliefs. The advertisement contained either a sponsored disclosure, in line with the disclosures Instagram uses ('Sponsored') in the control condition, or a sponsored disclosure and a more salient targeting disclosure based on the false information disclosures the platform used during the COVID-19 pandemic. However, the disclosure was adapted to contain information about the post being targeted ('This sponsored message is targeted at you based on your age, gender, and online behavior') and was highlighted by a red square. Both stimuli can be found on OSF (https://osf.io/2rbqu/?view_only=7923aff35ecd44caa90f7c7603912e03). In total, we created four different Instagram ads, meaning that for every condition, we had pro- or anti-climate change regulations to fit the view of participants to simulate targeting (see Procedure). For both conditions, the number of likes, comments, and timestamp of posting (14 hours ago) were the same.

Procedure

After providing informed consent, participants were briefed about the study and that the next page would include 10 political statements we needed them to either agree or disagree with. After answering the statements, which we used to investigate their point of view on climate change regulations, we briefed them again, this time about the next page containing an Instagram post which we asked them to closely look at, and that a 'next' button would appear after 15 seconds. After this briefing we exposed participants to our stimulus material containing a statement that was either pro- or anti-climate change regulations, fitting their earlier answers. Subsequently, we asked them questions regarding our variables and a manipulation check. We then asked the participants about their demographic information before we finally debriefed them and thanked them. The average completion time was five minutes.

Sample

We recruited 464 adult German Instagram users through panel provider Prolific from December 13 to 16, 2022. We were unable to identify the participants individually in this study. Seven participants quit our questionnaire before completion, four participants failed our attention check, two participants timed out before completion, and one participant did not agree with our informed consent. Leaving us with a final sample of 450 participants that we included in our analyses. Within our sample, the age ranged from 18 to 70 years (M = 29.4, SD = 9.5). Of these participants, 220 identified as female, 223 as male, and 7 as diverse. Regarding education, most of our participants had a university entrance qualification (n = 150), 114 had a bachelor's degree, and 89 had a master's degree. Furthermore, 49 participants had an intermediate high school diploma, and 28 had an advanced technical college entrance diploma. Finally, 10

participants had a doctorate, five had a qualifying middle school diploma, and five participants responded that they had a different educational background. Randomization checks showed no differences between our two experimental groups regarding age ($F(1, 448) = 0.45$, $p = .505$), gender ($\chi 2(2, N = 450) = 0.17$, $p = .921$), or level of education ($\chi 2(7, N = 450) = 8.37$, $p = .301$).

Power

Given the budget for this study, we were able to gather responses from 450 participants [64]. To determine our statistical power, we used R (version 4.1.2) [65] and the simsem package (Version 0.5–13) [66]. To establish our smallest effect size of interest (SESOI) [67], which could still be interpreted meaningfully given our current sample size and method, we employed various simulation analyses using the structure of our structural equation model (SEM). We set our power level at 90%, which is desirable at least [68], and kept our α level (α = .05) and sample size ($N = 450$) constant while testing various effect sizes. We ran 1,000 replications for our model. Finally, we set our SESOI at β = |.15| and did not interpret results with effect sizes smaller than this.

Measures

All constructs used to measure our variables were tested for factor validity in confirmatory factor analyses (CFAs). The results of the CFAs and measures of reliability and internal consistency are shown in Table 1. We measured *targeting knowledge* through five items used validated in earlier research on PMT by [1] on a Likert scale ranging from 1 (= strongly disagree) to 7 (= strongly agree) (e.g., "The post is tailored to me"), one item "The post showed personalized advertising" decreased our indicators for a good fit in a CFA and thus was not included in our analyses. We measured *perceived benefits* through five items combining the most fitting items from Lavado-Nalvaiz and colleagues [69] that investigated the privacy calculus for smart home devices with items from Yang [70] investigating the privacy calculus for online behavioral advertising using a Likert scale ranging from 1 (= strongly disagree) to 7 (= strongly agree) (e.g., "The Instagram post is helpful"). Two items "Seeing an Instagram post that is targeted at me makes me happy" and "I know a targeted Instagram post fits and suits me" decreased our indicators for a good fit in a CFA and thus were not included in our analyses. We measured *privacy concerns* using a scale that has been used in recent microtargeting research by Dobber et al. [13], consisting of five items measured on a Likert scale ranging from 1 (= strongly disagree) to 5 (= strongly agree) (e.g., "I am worried that my personal data (*such as my online surf and search behavior, name, and location*) will be abused by others"). We measured *platform attitude* using an adapted version of the *Facebook Attitude Scale* (FAS) (tailoring it to Instagram) that Chua and Chua [71] adapted from the Facebook Questionnaire

Table 1. Results of the confirmatory factor analyses.

Measured Constructs	χ^2	*df*	*p*	CFI	TLI	RMSEA	SRMR	α	ω	AVE
Targeting Knowledge	213.08	2	.008	.99	.98	.09	.01	.90	.90	.69
Perceived Benefits	85.62	5	< .001	.95	.91	.19	.04	.91	.92	.68
Privacy Concerns	27.00	5	< .001	.98	.96	.10	.02	.87	.88	.60
Attitude Towards Instagram	54.09	5	< .001	.94	.88	.15	.04	.83	.83	.50
Intended Privacy Protection Behavior	43.25	5	< .001	.93	.87	.13	.05	.77	.77	.45

Reliability measures are Cronbach's α, McDonald's ω, and average variance extracted. CFI = Comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual; AVE = average variance extracted.

https://doi.org/10.1371/journal.pone.0295329.t001

developed by Ross et al. [72], consisting of seven items measured on a 5-point Likert scale ranging from 1 (= strongly disagree) to 5 (= strongly agree). Two items, "I feel out of touch when I haven't logged on to Instagram in a while" and "How satisfied are you with Instagram" decreased our indicators for a good fit in a CFA and thus were not included in our analyses. We measured *intended privacy protection behavior* using the scale used in recent microtargeting research by Binder et al. [1], consisting of five items measured on a Likert scale ranging from 1 (= strongly disagree) to 7 (= strongly agree) (e.g., "I will use software that disguises my identity online"). One item "I will deregister from an app or account to protect my data" decreased our indicators for a good fit in a CFA and thus was not included in our analyses. While further evaluating our items for this measurement, we found that the other items concerned changing settings on a platform, using software, or informing oneself. The excluded item might be a very rigorous measure to take, meaning that it could not be as uniformative compared to the rest of the items, which could explain why this item decreased our indicators. In addition, we recognize that the average variance explained for this scale is below the threshold of .50, but we decided not to further alter this scale because other indicators were above their respective thresholds. However, this scale combines various behaviors and does not represent a uniform scale.

Exploratory variable. As an exploratory variable, we measured *algorithmic user agency*. With this construct, we aim to reveal users' behavior, which is intended to inform the algorithm about their preferences to receive better-tailored content. This means that users consciously interact with certain content in their timelines to see more content like it (e.g., liking a cat video on Instagram to see more cat videos in the future). We included three items: "I like pictures and videos on social media to see more of that content", "I try to inform the algorithm about myself, to receive better-tailored content", and "I am selective in what I like or interact with, to let the algorithm know that". We measured these three items on a Likert scale ranging from 1 (= strongly disagree) to 7 (= strongly agree) (Cronbach's α = .89, McDonald's ω = .89, M = 3.1, SD = 1.2, AVE = .73).

Manipulation check

As a manipulation check, we asked participants if they recalled if there was a disclosure on the Instagram post and what the disclosure stated by asking them to check one of the following statements:

- The Instagram post was a regular post
- The Instagram post was labeled as targeted at me and sponsored
- The Instagram post was labeled as sponsored

Results

Statistical analyses were conducted using R (version 4.1.2) [65] and jamovi (version 2.0.0.0) [73]. To investigate the differences between the two conditions, we conducted a path model with our mean scores through structural equation modeling in lavaan [74]. The code of our conducted analyses is available on the OSF (https://osf.io/qa2en?view_only=7923aff35ecd44caa90f7c7603912e03). The bivariate correlations, means, and standard deviations for the measured variables are shown in Table 2. For all our hypotheses, we used our SESOI of $\beta = |.15|$ and an alpha level of .05 as thresholds for acceptance or rejection.

Table 2. Means, standard deviations, and bivariate correlations of the measured constructs.

Measured construct	*M* (*SD*)	1	2	3	4	5	6
1 Experimental condition	-	-					
2 Targeting Knowledge	4.1 (1.6)	.27***	-				
3 Perceived Benefits	3.8 (1.5)	.05	.54***	-			
4 Privacy Concerns	4.6 (1.3)	-.05	.03	.00	-		
5 Attitude Towards Instagram	2.8 (0.9)	.02	.03*	.13**	-.04	-	
6 Intended Privacy Protection Behavior	3.7 (1.3)	-.01	.02	.03	.42***	.15**	-
7 Algorithmic User Agency	3.1 (1.2)	.10*	.14**	.21***	-.08	.15**	.00

* $p < .05$,
** $p < .01$,
*** $p < .001$.

https://doi.org/10.1371/journal.pone.0295329.t002

Regarding our manipulation check, in our sponsored condition (n = 225) 132 participants did not recall a disclosure, 72 participants rightfully recalled a sponsored label, and 21 participants stated the post contained a targeting disclosure. In our targeting disclosure condition (n = 225) 189 participants rightfully recalled a targeting disclosure, 19 recalled no disclosure and 17 recalled a sponsored disclosure.

As preregistered, we tested our hypotheses, without our moderation hypotheses, in a path model. Our model fit was evaluated in line with frequently used fit indices [75]. Our model showed an adequate fit: $\chi2(5) = 8.09$, p = .151, $\chi2/df = 1.62$, CFI = 0.99, TLI = 0.98, RMSEA = .04, 90% CI [.00, .08], SRMR = .03. This model is illustrated in Fig 1. Due to our available resources and our maximum sample size, our path model with an inclusion of the moderation would be too complicated and would decrease the power and interpretation of our results. However, we tried to test the model with moderation included, but this led to model fit indices that were far from reasonable to interpret.

Our first hypothesis predicted that exposure to a targeting disclosure would lead to higher levels of targeting knowledge than exposure to a sponsored disclosure. In line with this hypothesis, we found a significant positive effect (β = .27, $p < .001$), leading us to accept H_1. The second hypothesis predicted a positive relationship between targeting knowledge and perceived

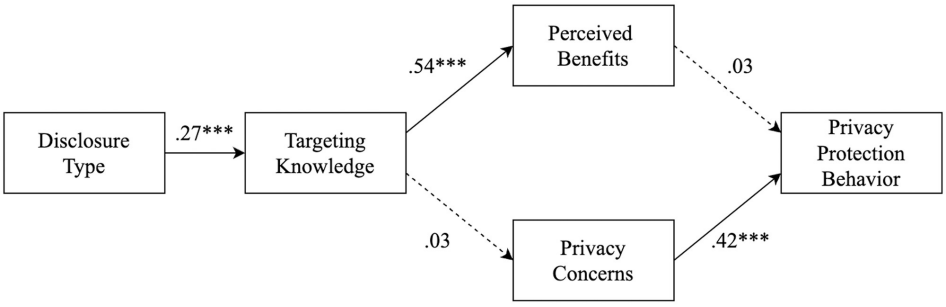


Fig 1. Path model. Numbers represent standardized regression coefficients. Dashed lines indicate a path is either statistically meaningless (not significant) or theoretically meaningless (below β = |.15|). * $p < .05$, ** $p < .01$, *** $p < .001$.

https://doi.org/10.1371/journal.pone.0295329.g001

benefits. We found a significant positive relationship (β = .54, $p < .001$), supporting this hypothesis. Moreover, through our third hypothesis, we predicted that targeting knowledge would be positively related to privacy concerns, which we found no evidence for in our data (β = .03, $p = .531$), leading us to reject this hypothesis.

Our fifth hypothesis predicted that perceived benefits would be negatively related to intended privacy protection behavior, which we found no evidence for in our data (β = .03, $p = .471$), leading us to reject H_5. Our sixth hypothesis predicted that privacy concerns would be positively related to intended privacy protection behavior, which we found evidence of in our data (β = .42, $p < .001$), leading us to accept H_6.

Moderation

For our fourth hypothesis (H_4), we proposed that the relationship between targeting knowledge and both perceived benefits (H_{4a}) and privacy concerns (H_{4b}) would be moderated by users' attitude towards Instagram. We investigated these hypotheses using the medmod [76] module in jamovi with 1,000 bootstrapped samples. For H_{4a}, we did not find a direct relationship between users' attitude towards Instagram and users' perceived benefits ($b = .11$, $p = .088$), nor did we find an interaction between users' attitude and targeting knowledge ($b = .03$, $p = .441$), leading us to reject H_{4a}. For H_{4b}, the direct effect of users' attitude towards Instagram was not found to be related to users' privacy concerns ($b = -.06$, $p = .450$), as was the case for the interaction between users' attitude and targeting knowledge ($b = .05$, $p = .279$), leading us to reject H_{4b}.

Exploration

Finally, through research question one, we investigate our exploratory variable, algorithmic user agency, and its potential relationship with perceived benefits. Our data showed this relationship ($r = .21$, $p < .001$), meaning that there was indeed a positive correlation between the two constructs.

Discussion

The current study aimed to investigate whether targeting disclosures on political microtargeted advertisements on Instagram would increase users' awareness and perceptions of their online privacy, and how this related to their privacy behavior. To do so, we investigated microtargeting disclosures through an integrative path model with a rationale based on the privacy calculus that expected users privacy benefits and costs to be related to increases or decreases in their privacy protection behavior.

We expected that exposure to a targeting disclosure, compared to a sponsored disclosure, would lead to higher levels of targeting knowledge, and thus help users recognize targeted political ads [1, 2, 20]. Although prior research shows mixed results concerning users' recall and interpretations of disclosures, we found proof of this assumption in our data. One of the consistencies in the existing work on disclosures as a measure to increase transparency is that users do not always recall disclosures correctly [1, 2, 77]. In the current study, we found results that are partially in line with these findings. In our control condition, in which the post was labeled with Instagram's current sponsored disclosure, 32% of our participants correctly recalled the disclosure. However, in our experimental condition, where we exposed participants to a targeting disclosure based on the COVID-19 misinformation disclosures that Instagram and Facebook used during the pandemic, we found different results. In this group, the vast majority (84%) recalled the targeting disclosure. Although it was not a core question of this study, we, in line with other works, were able to show that the current disclosures

regarding sponsored content on Instagram do not work as intended and suffer from a lack of recall and are thus not able to contribute to transparency regarding targeting [1, 19]. On the other hand, we were also able to show that larger, more salient disclosures led to greater recall within our sample, which eventually could lead to greater efficiency of disclosures.

Besides testing the potential effects of our disclosures on targeting knowledge, we expected that users' awareness of microtargeting and its processes would be positively related to both perceived benefits and perceived privacy risks. We found this relationship only for the perceived benefits. This shows that when users are aware of the advertisement being targeted at them, they have higher levels of perceived benefits, which is in line with work that found more positive evaluations of content when it is personalized [49, 51]. A potential mechanism behind this could be that the fluency of the content being personalized at the user disregards the message being perceived as a warning. While previously disclosures were mainly used to inform or even warn users about content being tailored and therefore more personalized than a 'regular' message or advertisement, it could be that users' are more used to tailoring of timelines on platforms, and even advertisements, that they mainly perceive the benefits of tailoring as a technique and therefore perceive it as a useful or beneficial technique.

Regarding privacy concerns, we did not find a relationship with targeting knowledge. Prior work shows that in the case of behavioral advertising, persuasion knowledge affects privacy risk perceptions. While existing research shows that if people know that they are exposed to personalized messages, this affects privacy risks [54] or that users even feel that these messages are intrusive or creepy, our results did not align with this [23, 53]. Moreover, we would like to point out that our result aligns with the novel work by Dobber and colleagues, who also did not find a relationship between exposure to transparency information and users' privacy concerns [36]. However, looking at the mean scores for privacy benefits and concerns, we see that the mean for privacy concerns is substantially higher than the mean for perceived benefits, which could imply that even though we did not find a relationship with targeting knowledge, users still have relatively high privacy concerns regarding PMT. A certain privacy threat awareness could have already existed, which might explain why explicitly informing users about the use of their data might not make a difference.

Regarding the relation between targeting knowledge and privacy concerns we would like to point out that based on the fact that we did not find a relation could also, on a speculative basis, be explained by a negativity bias [78]. By default, users might be more focused on the concerns regarding their online privacy, which could explain why there was no relation. In contrast, this might be why we did find a relation between targeting knowledge and perceived benefits, because the benefits of microtargeting might not be the default thing people think about if they are informed about the technique. In addition, a ceiling effect could also limit the potential relationship between targeting knowledge and privacy concerns since the mean score was already relatively high. This might be a threshold that does not increase or decrease by explicitly explaining that users are targeted.

Furthermore, we assumed that the relationship between users' awareness and perceived benefits and privacy risks would be moderated by users' attitude towards Instagram, meaning that for users with a more positive attitude towards Instagram, the relation between targeting knowledge and perceived benefits would be stronger and that for those users, the relation between targeting knowledge and privacy concerns would be weaker. Conversely, we expected that, for users with less favorable attitudes towards Instagram, the relationship between targeting knowledge and perceived benefits would be weaker, and the relationship between targeting knowledge and privacy concerns would be stronger. Contrary to existing work, we did not find proof of our moderation hypotheses in our data [58]. One reason for this could be that PMT as a technique is not related to the platform in the perception of users. Even though the

implementation of the technique is the same for both Instagram and Facebook, it could be that users' awareness of tailoring in any form might have risen beyond just the platforms. Tailoring of both content and advertisements is something that happens not only on SNS but also on regular websites and search engines. It is possible that users do not distinguish between tailoring and PMT specifically, and see this separately from the platform they are using.

Moreover, we expected that perceived benefits would be negatively related to intended privacy protection behavior, meaning that users who perceive the benefits of being targeted and receiving personalized advertisements engage less in measures to protect their online privacy. However, we did not find proof for this in our data, which is confirmatory for work on personalization [79] and SNSs [80] (the latter used self-withdrawal, which can be seen as a form of privacy protection behavior). While we made the assumption based on the privacy calculus, we do recognize other work that found no relation between perceived benefits and users' desire to protect their privacy online. It is possible that benefits and privacy protection behavior are rather independent constructs and that privacy protection behavior might not be just the outcome of a calculation that users make between benefits and risks [43, 60].

Furthermore, we assumed that privacy concerns would be positively related to intended privacy protection behavior, which we indeed found proof for in our data, consistent with other work regarding social media [43, 60] and microtargeting [1, 46]. However, we would like to emphasize that this does not show a calculation or weighing between privacy concerns and perceived benefits. This calculation could be visible in direct relations or interactions between the two constructs, or in the perceived benefits being larger than privacy concerns, which is not the case in the current study. The fact that we did find a relationship between privacy concerns and protection behavior but did not find a relationship between benefits and protection behavior would not necessarily imply that users weigh the factors beforehand. This mainly shows that when users have more privacy concerns, independent of whether they perceive the targeted ad to be beneficial or not, they aim to protect themselves online.

Additionally, as an exploratory research question, we investigated the relationship between users' perceived benefits and algorithmic user agency [63]. This means that users are aware of the algorithmic processes taking place (i.e., receiving content because the algorithm 'knows' that you interacted with content like that before), and try to actively control these processes by consciously interacting with content, which could be perceived as a potential proxy for their willingness to actively use algorithms to their advantage. We found a correlation between perceived benefits and algorithmic user agency. In addition, even though they were smaller, we also found correlations between the construct and our manipulation, targeting knowledge and attitudes towards the platform. This could imply that users who are explicitly informed about targeting practices might make use of their agency, but also that users with higher levels of knowledge about targeting and users with more favorable attitudes towards the platforms might operate in the same way. However, we emphasize that these findings are correlations and that we do not attempt to make causal claims regarding this subject.

Moreover, we argue that algorithmic user agency might fit as a secondary outcome of the calculus rationale for PMT. Users who experience higher benefits might interact with the advertised content and the algorithm more than users who are more focused on the risks, which in turn might lead them to protect their online privacy. Finally, although algorithmic user agency is a newly developed construct and, to the best of our knowledge, the first time this has been empirically investigated in an experiment, we encourage other researchers to build on these findings and explore users' behaviors and potential interactions with algorithms regarding PMT.

Limitations and future work

To avoid influencing participants beyond the scope of our experiment, we created a political party that supposedly posted the Instagram content in our stimulus material. However, we recognize that political behavior and, as an extension, voting are behaviors that are complex and do not change within a moment of looking at an Instagram post. While avoiding ethical risks, we also recognize that a novelty effect might have taken place for participants. It is not possible that participants ever saw content from this political party, which might have led them to focus more on the Instagram post and less on the disclosure, potentially underestimating the power of the effect we found. On the other hand, we forced participants to view the post, including the disclosure, for a minimum of 15 seconds, which might have led to an overestimation of not only the effect of the disclosure on targeting knowledge, but also on the percentage of participants who correctly recalled the disclosures in our sample, which is remarkably higher than in other studies. We believe that the field of personalization and microtargeting research will benefit from large-scale field experiments with actual usage data gathered through web scraping or even data donations.

Concerning the statements used in our stimuli, even though we used them to improve ecological validity since they were used in the German Wahl-o-mat, we acknowledge that the statements are not mutually exclusive and that participants who agree with a pro-climate change regulations statement could also agree with an anti-climate change statement. After investigating these subgroups, however, we found that only 42 participants were exposed to the anti-climate change regulations statement (21 per condition). Therefore, we did not perform our analyses with separate subgroups. Future research could benefit from statements that are mutually exclusive, as this could make for better comparisons.

Regarding the effect of our manipulation on targeting knowledge, we would like to point out that participants saw our manipulations in a controlled experimental setting (i.e., a timer that tried to force them to look at the Instagram post for 15 seconds). This might have led them to use not only more cognitive resources to process the message, but also the disclosure itself, compared to a less superficial setting. In a setting with higher ecological validity, SNS users might be exposed to more messages in a shorter period, which leads to the question of whether they would process the message, and accordingly the disclosure, less central and more peripheral, as explained in the Elaboration Likelihood Model [52]. In a real-world setting, users might focus more on the actual content before focusing on a disclosure. Earlier work, for example, showed that users focus on a video or image in their timelines before they focus on anything else [81].

In the current study, our aim was to simulate microtargeting by asking participants if they agreed with statements concerning climate change regulations. Even though we showed our participants' captions on the Instagram post that we assumed fit their views more than a general statement would, we recognize that there is a chance that this was not the case, and the post might not have been perceived as microtargeted. In addition, we recognize that there is a chance that participants were aware that the post we showed them was not actually based on their personal data. Besides, there are many different regulations regarding climate change reduction, and people might generally agree with the regulations but could also disagree with the ones we selected (e.g., the speed limit on the German highway). However, we would like to point out that in a real-world setting, ads that users receive might not always be perceived as perfectly fitting, even though this is what senders try to achieve.

We believe that the field of disclosure research, as well as legislators and governmental institutions, will benefit from research that includes a more practical approach and not only investigates the effects and relations regarding disclosures, but also the design of the disclosures

themselves and which designs are easier recalled by users. Recallment is a large part of disclosures effectiveness, as disclosures need to be recalled and perceived to inform people. We also want to emphasize that research would benefit from investigating whether these disclosures are perceived as annoying or disturbing for users. Besides, from a policy and transparency perspective, future research would benefit from a clearer investigation of why users recall (or do not) certain disclosures, which, for instance, might be due to a lack of exposure but could also have to do with a lack of interest in the information about an ad, if an ad is directly recognized as such, which might lead to the user directly scrolling to other content. We believe that eye-tracking studies combined with more ecologically valid experiments where users scroll through the timelines of their own SNS accounts might provide clearer insights into user behavior on SNS and the effectiveness of disclosures if they are embedded on the platforms.

Regarding our moderation hypothesis, we would like to emphasize that to detect potential moderation, this study might have had a sample size that was not sufficient to detect moderation with sufficient power [82]. Finally, apart from the effect of our manipulation on targeting knowledge, this study has a cross-sectional design and, therefore, does not allow us to investigate causal relations, meaning that other paths in our model are bidirectional.

Conclusion

The targeting of political advertising on social media is something that users do not see occurring in the foreground, or recognize at all. However, governments, regulators, and researchers have reached consensus on the need to improve transparency. The current study investigated the effectiveness of targeting disclosures as a means to improve transparency, and subsequently, scrutinizes users' perceptions of their online privacy based on the privacy calculus. In an integrative model, we found that our disclosure affected users' targeting knowledge, which was positively related to their perceived benefits of PMT. Nevertheless, we did not find a relationship between targeting knowledge and users' privacy concerns regarding PMT. Additionally, neither of these relationships was moderated by users' attitudes towards the platform they were using. In addition, we did not find a relationship between users' perceived benefits and intended privacy protection behavior; however, we did find a relationship between privacy concerns and intended protection behavior. Together, our findings show that if users are alerted about targeting practices taking place on platforms, they see the benefits of personalization, and that this does not relate to behavior that protects their privacy. In addition, we see that only if users view personalization as a privacy violation they might engage in behavior to protect their privacy. Although the exact role of targeting disclosures and their desired designs may still be a topic of debate for legislators, this study provides a first interpretation of what these disclosures mean to users' privacy perception if they are made aware that they are microtargeted.

Acknowledgments

The authors thank Nur Efsan Cetinkaya for helping with the questionnaire and translations.

Author Contributions

Conceptualization: Martin-Pieter Jansen, Nicole C. Krämer.

Data curation: Martin-Pieter Jansen.

Formal analysis: Martin-Pieter Jansen.

Investigation: Martin-Pieter Jansen.

Methodology: Martin-Pieter Jansen, Nicole C. Krämer.

Project administration: Martin-Pieter Jansen.

Resources: Martin-Pieter Jansen, Nicole C. Krämer.

Supervision: Nicole C. Krämer.

Writing – original draft: Martin-Pieter Jansen.

Writing – review & editing: Martin-Pieter Jansen, Nicole C. Krämer.

References

1. Binder A, Stubenvoll M, Hirsch M, Matthes J. Why Am I Getting This Ad? How the Degree of Targeting Disclosures and Political Fit Affect Persuasion Knowledge, Party Evaluation, and Online Privacy Behaviors. J Advert. 2022; 1–17. https://doi.org/10.1080/00913367.2021.2015727
2. Kruikemeier S, Sezgin M, Boerman SC. Political Microtargeting: Relationship between Personalized Advertising on Facebook and Voters' Responses. Cyberpsychology Behav Soc Netw. 2016; 19: 367–372. https://doi.org/10.1089/cyber.2015.0652 PMID: 27327063
3. Matz SC, Kosinski M, Nave G, Stillwell DJ. Psychological targeting as an effective approach to digital mass persuasion. Proc Natl Acad Sci U S A. 2017; 114: 12714–12719. https://doi.org/10.1073/pnas.1710966114 PMID: 29133409
4. Zarouali B, Dobber T, De Pauw G, de Vreese C. Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media. Commun Res. 2020. https://doi.org/10.1177/0093650220961965
5. Bodó B, Helberger N, De Vreese CH. Political micro-targeting: A manchurian candidate or just a dark horse? Internet Policy Rev. 2017; 6. https://doi.org/10.14763/2017.4.776
6. Endres K, Kelly KJ. Does microtargeting matter? Campaign contact strategies and young voters. J Elections Public Opin Parties. 2018; 28: 1–18. https://doi.org/10.1080/17457289.2017.1378222
7. Jamieson KH. Messages, micro-targeting, and new media technologies. Forum Ger. 2013; 11: 429–435. https://doi.org/10.1515/for-2013-0052
8. Zuiderveen Borgesius FJ, Möller J, Kruikemeier S, Fathaigh R, Irion K, Dobber T, et al. Online political microtargeting: Promises and threats for democracy. Utrecht Law Rev. 2018; 14: 82–96. https://doi.org/10.18352/ulr.420
9. Kerem K, Ulla M. Perceived Intrusiveness of Personalized Marketing. Digital Transformation–Meeting the challenges. Univerzitetna založba Univerze v Mariboru / University of Maribor Press; 2018. pp. 147–158.
10. Matthes J, Hirsch M, Stubenvoll M, Binder A, Kruikemeier S, Lecheler S, et al. Understanding the democratic role of perceived online political micro-targeting: longitudinal effects on trust in democracy and political interest. J Inf Technol Polit. 2022; 1–14. https://doi.org/10.1080/19331681.2021.2016542
11. Cadwalladr C. 'I made Steve Bannon's psychological warfare tool': meet the data war whistleblower. The Guardian. 18 Mar 2018. https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump. Accessed 12 Apr 2022.
12. Bennett Collin J. Trends in Voter Surveillance in Western Societies: Privacy Intrusions and Democratic Implications. Surveill Soc. 2015; 13: 370–384. https://doi.org/10.24908/ss.v13i3/4.5373
13. Dobber T, Trilling D, Helberger N, de Vreese C. Spiraling downward: The reciprocal relation between towardsde toward political behavioral targeting and privacy concerns. New Media Soc. 2019; 21: 1212–1231. https://doi.org/10.1177/1461444818813372
14. Tene O. Privacy: The new generations. Int Data Priv Law. 2011; 1: 15–27. https://doi.org/10.1093/idpl/ipq003
15. Boerman SC, Kruikemeier S, Zuiderveen Borgesius FJ. Online Behavioral Advertising: A Literature Review and Research Agenda. J Advert. 2017; 46: 363–376. https://doi.org/10.1080/00913367.2017.1339368
16. Amazeen MA, Wojdynski BW. The effects of disclosure format on native advertising recognition and audience perceptions of legacy and online news publishers. Journalism. 2020; 21: 1965–1984. https://doi.org/10.1177/1464884918754829
17. Boerman SC, Van Reijmersdal EA, Neijens PC. Using Eye Tracking to Understand the Effects of Brand Placement Disclosure Types in Television Programs. J Advert. 2015; 44: 196–207. https://doi.org/10.1080/00913367.2014.967423

18. Cain RM. Embedded Advertising on Television: Disclosure, Deception, and Free Speech Rights. J Public Policy Mark. 2011; 30: 226–238. https://doi.org/10.1509/jppm.30.2.226

19. Binford MT, Wojdynski BW, Lee Y-I, Sun S, Briscoe A. Invisible transparency: Visual attention to disclosures and source recognition in Facebook political advertising. J Inf Technol Polit. 2021; 18: 70–83. https://doi.org/10.1080/19331681.2020.1805388

20. Jost P, Kruschinski S, Sülflow M, Haßler J, Maurer M. Invisible transparency: How different types of ad disclaimers on Facebook affect whether and how digital political advertising is perceived. Policy Internet. 2022; poi3.333. https://doi.org/10.1002/poi3.333

21. Friestad M, Wright P. The Persuasion Knowledge Model: How People Cope with Persuasion Attempts. J Consum Res. 1994; 21: 1–31. https://doi.org/10.1086/209380

22. Culnan MJ, Armstrong PK. Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation. Organ Sci. 1999; 10: 104–115. https://doi.org/10.1287/orsc.10.1.104

23. Segijn CM, Van Ooijen I. Differences in consumer knowledge and perceptions of personalized advertising: Comparing online behavioural advertising and synced advertising. J Mark Commun. 2022; 28: 207–226. https://doi.org/10.1080/13527266.2020.1857297

24. Pew Research Center. Use of online platforms, apps varies–sometimes widely–by demographic group. In: Pew Research Center: Internet, Science & Tech [Internet]. 5 Apr 2021 [cited 14 Dec 2022]. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/pi_2021-04-07_social-media_0-03/

25. European Commission. Regulation of the European Parliament and of The Council on a Single Market For Digital Services (Digital Services Act) and amending Directive. 2022. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN

26. King JM. Microtargeted Political Ads: An Intractable Problem. Boston Univ Law Rev. 2022; 102: 39.

27. Lubianco J. Fake News bill gets stuck in Brazilian Congress and it's unlikely to be voted on before the elections; remuneration proposal for journalistic organizations is a sensitive topic. In: LatAm Journalism Review by the Knight Center [Internet]. 12 Apr 2022 [cited 23 Jun 2023]. https://latamjournalismreview.org/articles/fake-news-brazil-payment-journalism/

28. Lawless J. Online Advertising and Social Media (Transparency) Bill. 150–2017 Mar 21, 2021. https://data.oireachtas.ie/ie/oireachtas/bill/2017/150/eng/initiated/b15017d.pdf

29. Ferrand R, Braun-Pivet Y, Studer B, Moutchou N, Attal G, Anato P. Lutte contre la manipulation de l'information. 0297, 2018, 1202 Mar 21, 2018. https://www.senat.fr/dossier-legislatif/ppl17-623.html

30. Dobber T. Effectiveness of Online Political Ad Disclosure Labels: Empirical Findings. 2021; 6.

31. Wojdynski BW, Evans NJ. Going Native: Effects of Disclosure Position and Language on the Recognition and Evaluation of Onlin. J Advert. 2016; 45: 157–168. https://doi.org/10.1080/00913367.2015.1115380

32. Clark M. Instagram is rolling out new notifications about COVID-19 information. In: The Verge [Internet]. 17 Dec 2020 [cited 18 Jan 2023]. https://www.theverge.com/2020/12/17/22187298/instagram-coronavirus-covid19-misinformation-notifications

33. Instagram. Helping to Inform People about COVID-19 Vaccines | Instagram Blog. 16 Mar 2020 [cited 18 Jan 2023]. https://about.instagram.com/blog/announcements/continuing-to-keep-people-safe-and-informed-about-covid-19

34. Boerman SC, Van Reijmersdal EA, Neijens PC. Sponsorship Disclosure: Effects of Duration on Persuasion Knowledge and Brand Responses. J Commun. 2012; 62: 1047–1064. https://doi.org/10.1111/j.1460-2466.2012.01677.x

35. Boerman SC, Kruikemeier S. Consumer responses to promoted tweets sent by brands and political parties. Comput Hum Behav. 2016; 65: 285–294. https://doi.org/10.1016/j.chb.2016.08.033

36. Dobber T, Kruikemeier S, Helberger N, Goodman E. Shielding citizens? Understanding the impact of political advertisement transparency information. New Media Soc. 2023. https://doi.org/10.1177/14614448231157640

37. Aguirre E, Mahr D, Grewal D, de Ruyter K, Wetzels M. Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. J Retail. 2015; 91: 34–49. https://doi.org/10.1016/j.jretai.2014.09.005

38. Dijkstra A. Working mechanisms of computer-tailored health education: evidence from smoking cessation. Health Educ Res. 2005; 20: 527–539. https://doi.org/10.1093/her/cyh014 PMID: 15701665

39. Dobber T, Trilling D, Helberger N, de Vreese CH. Two crates of beer and 40 pizzas: the adoption of innovative political behavioural targeting techniques. Internet Policy Rev. 2017; 6. https://doi.org/10.14763/2017.4.777

40. Carr CT, Hayes RA. The Effect of Disclosure of Third-Party Influence on an Opinion Leader's Credibility and Electronic Word of Mouth in Two-Step Flow. J Interact Advert. 2014; 14: 38–50. https://doi.org/10.1080/15252019.2014.909296

41. Dinev T, Hart P. An Extended Privacy Calculus Model for E-Commerce Transactions. Inf Syst Res. 2006; 17: 61–80. https://doi.org/10.1287/isre.1060.0080

42. Trepte S. The Social Media Privacy Model: Privacy and Communication in the Light of Social Media Affordances. Commun Theory. 2021; 31: 549–570. https://doi.org/10.1093/ct/qtz035

43. Dienlin T, Metzger MJ. An Extended Privacy Calculus Model for SNSs: Analyzing Self-Disclosure and Self-Withdrawal in a Representative U.S. Sample: THE EXTENDED PRIVACY CALCULUS MODEL FOR SNSs. J Comput-Mediat Commun. 2016; 21: 368–383. https://doi.org/10.1111/jcc4.12163

44. Meier Y, Meinert J, Krämer NC. Investigating factors that affect the adoption of COVID-19 contact-tracing apps: A privacy calculus perspective. Technol Mind Behav. 2021; 2: 1–10. https://doi.org/10.1037/tmb0000040

45. Schäwel J, Frener R, Trepte S. Political Microtargeting and Online Privacy: A Theoretical Approach to Understanding Users' Privacy Behaviors. Media Commun. 2021; 9: 158–169. https://doi.org/10.17645/mac.v9i4.4085

46. Stubenvoll M, Binder A, Noetzel S, Hirsch M, Matthes J. Living is Easy With Eyes Closed: Avoidance of Targeted Political Advertising in Response to Privacy Concerns, Perceived Personalization, and Overload. Commun Res. 2022; 009365022211308. https://doi.org/10.1177/00936502221130840

47. Zarouali B, Verdoodt V, Walrave M, Poels K, Ponnet K, Lievens E. Adolescents' advertising literacy and privacy protection strategies in the context of targeted advertising on social networking sites: implications for regulation. Young Consum. 2020; 21: 351–367. https://doi.org/10.1108/YC-04-2020-1122

48. Barocas S. The price of precision: voter microtargeting and its potential harms to the democratic process. Proceedings of the first edition workshop on Politics, elections and data. New York, NY, USA: Association for Computing Machinery; 2012. pp. 31–36.

49. Tam KY, Ho SY. Understanding the Impact of Web Personalization on User Information Processing and Decision Outcomes. MIS Q. 2006; 30: 865–890. https://doi.org/10.2307/25148757

50. De Keyzer F, Van Noort G, Kruikemeier S. Going Too Far? How Consumers Respond to Personalized Advertising From Different Sources. J Electron Commer Res. 2022; 23: 22.

51. Brinson NH, Eastin MS. Juxtaposing the persuasion knowledge model and privacy paradox: An experimental look at advertising personalization, public policy and public understanding. J Psychosoc Res. 2016; 10.

52. Petty RE, Cacioppo JT. The Elaboration Likelihood Model of Persuasion. Adv Exp Soc Psychol. 1986; 19: 123–205. https://doi.org/10.1016/S0065-2601(08)60214-2

53. Segijn CM, Opree SJ, Van Ooijen I. The validation of the Perceived Surveillance Scale. Cyberpsychology J Psychosoc Res Cyberspace. 2022; 16. https://doi.org/10.5817/CP2022-3-9

54. Jain S, Purohit HC. Privacy concerns and avoidance behaviour towards data-driven online behavioural advertising. Bus Anal J. 2022; 43: 1–12. https://doi.org/10.1108/BAJ-08-2022-0024

55. Debatin B, Lovejoy JP, Horn A-K, Hughes BN. Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. J Comput-Mediat Commun. 2009; 15: 83–108. https://doi.org/10.1111/j.1083-6101.2009.01494.x

56. Schwarz N, Bless H, Bohner G. Mood and Persuasion: Affective States Influence the Processing of Persuasive Communications. Advances in Experimental Social Psychology. Elsevier; 1991. pp. 161–199.

57. Heim J, Asting T, Schliemann T. Medium effects on persuasion. Proceedings of the second Nordic conference on Human-computer interaction—NordiCHI '02. Aarhus, Denmark: ACM Press; 2002. p. 259.

58. Hayes JL, Brinson NH, Bott GJ, Moeller CM. The Influence of Consumer–Brand Relationship on the Personalized Advertising Privacy Calculus in Social Media. J Interact Mark. 2021; 55: 16–30. https://doi.org/10.1016/j.intmar.2021.01.001

59. Soffer T, Cohen A. Privacy Perception of Adolescents in a Digital World. Bull Sci Technol Soc. 2014; 34: 145–158. https://doi.org/10.1177/0270467615578408

60. Meier Y, Schäwel J, Krämer NC. Between protection and disclosure: applying the privacy calculus to investigate the intended use of privacy-protecting tools and self-disclosure on different websites. Stud Commun Media. 2021; 10: 283–306. https://doi.org/10.5771/2192-4007-2021-3-283

61. Büchi M, Just N, Latzer M. Caring is not enough: the importance of Internet skills for online privacy protection. Inf Commun Soc. 2017; 20: 1261–1278. https://doi.org/10.1080/1369118X.2016.1229001

62. Johnson GA, Shriver SK, Du S. Consumer Privacy Choice in Online Advertising: Who Opts Out and at What Cost to Industry? Mark Sci. 2020; 39: 33–51. https://doi.org/10.1287/mksc.2019.1198

63. Kapsch PH. Exploring user agency and small acts of algorithm engagement in everyday media use. Media Int Aust. 2022; 183: 16–29. https://doi.org/10.1177/1329878X211067803

64. Lakens D. Sample Size Justification. Collabra Psychol. 2022; 8: 33267. https://doi.org/10.1525/collabra.33267

65. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. https://www.R-project.org/

66. Pornprasertmanit S, Miller P, Schoemann A, Jorgensen TD, Quick C. simsem: SIMulated Structural Equation Modeling. 2021. https://CRAN.R-project.org/package=simsem

67. Lakens D, Scheel AM, Isager PM. Equivalence Testing for Psychological Research: A Tutorial. Adv Methods Pract Psychol Sci. 2018; 1: 259–269. https://doi.org/10.1177/2515245918770963

68. Curran-Everett D. CORP: Minimizing the chances of false positives and false negatives. J Appl Physiol. 2017; 122: 91–95. https://doi.org/10.1152/japplphysiol.00937.2016 PMID: 27909232

69. Lavado-Nalvaiz N, Lucia-Palacios L, Pérez-López R. The role of the humanisation of smart home speakers in the personalisation–privacy paradox. Electron Commer Res Appl. 2022; 53: 101146. https://doi.org/10.1016/j.elerap.2022.101146

70. Yang K. Online Behavioral Advertising: Why and how online customers respond to it?: An experimental study into the effects of personalized levels, rewards on click-through intentions towards ads between Chinese and Dutch. Essay, University of Twente. 2020. http://essay.utwente.nl/80840/

71. Chua YP, Chua YP. Do computer-mediated communication skill, knowledge and motivation mediate the relationships between personality traits and attitude toward Facebook? Comput Hum Behav. 2017; 70: 51–59. https://doi.org/10.1016/j.chb.2016.12.034

72. Ross C, Orr ES, Sisic M, Arseneault JM, Simmering MG, Orr RR. Personality and motivations associated with Facebook use. Comput Hum Behav. 2009; 25: 578–586. https://doi.org/10.1016/j.chb.2008.12.024

73. The jamovi project. jamovi. 2022. http://www.jamovi.org

74. Rosseel Y. *lavaan*: An R Package for Structural Equation Modeling. J Stat Softw. 2012; 48. https://doi.org/10.18637/jss.v048.i02

75. Hooper D, Couglan J, Mullen MR. Structural equation modelling: guidelines for determining model fit. J Bus Res Methods. 2008; 6: 53–60.

76. Selker R. medmod: Simple Mediation and Moderation Analysis. 2019. https://github.com/raviselker/medmod

77. Evans NJ, Phua J, Lim J, Jun H. Disclosing Instagram Influencer Advertising: The Effects of Disclosure Language on Advertising Recognition, Attitudes, and Behavioral Intent. J Interact Advert. 2017; 17: 138–149. https://doi.org/10.1080/15252019.2017.1366885

78. Baumeister RF, Bratslavsky E, Finkenauer C, Vohs KD. Bad is Stronger than Good. Rev Gen Psychol. 2001; 5: 323–370. https://doi.org/10.1037/1089-2680.5.4.323

79. Van Ooijen I, Segijn CM, Opree SJ. Privacy Cynicism and its Role in Privacy Decision-Making. Commun Res. 2022; 009365022110609. https://doi.org/10.1177/00936502211060984

80. Meier Y, Schäwel J, Kyewski E, C. Krämer N. Applying Protection Motivation Theory to Predict Facebook Users' Withdrawal and Disclosure Intentions. International Conference on Social Media and Society. Toronto ON Canada: ACM; 2020. pp. 21–29.

81. Vergara A, Siles I, Castro AC, Chaves A. The Mechanisms of "Incidental News Consumption": an Eye Tracking Study of News Interaction on Facebook. Digit Journal. 2021; 9: 215–234. https://doi.org/10.1080/21670811.2020.1813047

82. Memon MA, Cheah J-H, Ramayah T, Ting H, Chuah F, Cham TH. Moderation analysis: issues and guidelines. J Appl Struct Equ Model. 2019; 3: i–xi. https://doi.org/10.47263/JASEM.3(1)01

Article IV

Jansen, M.-P., & Van Ooijen, I. (In revision). For your eyes only? An eye-tracking experiment investigating microtargeting transparency, visual attention, and critical processing. *Journal of Media Psychology.*

For your eyes only? An eye-tracking experiment investigating political microtargeting transparency, visual attention, and critical processing

Martin-Pieter Jansen[1], and Iris van Ooijen[2]

[1]Social Psychology: Media and Communication, University of Duisburg-Essen, 47057, Duisburg, Germany

[2]Behavioural Science Institute, Radboud University, 6525 GD, Nijmegen, The Netherlands

Author Note

Martin-Pieter Jansen [https://orcid.org/0000-0003-2876-8073]

Iris van Ooijen [https://orcid.org/0000-0002-1907-6158]

We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Martin-Pieter Jansen, Social Psychology: Media and Communication, University of Duisburg-Essen, Forsthausweg 2, 47057 Duisburg, Germany. E-Mail: martin-pieter.jansen@uni-due.de

Abstract

By means of their personal data, users on social media platforms get targeted to be persuaded by political advertisers with messages that are specifically tailored to them: political microtargeting. A substantial problem with political microtargeting is that receivers of targeted advertisements are often unaware that they are being targeted. Increasing transparency to improve awareness potentially leads to a more critical way to process information in an advertisement. Moreover, the use of these measures could lead to differences in users' attention to the advertisement. Therefore, this study investigates whether a disclosure that informs users about targeting practices occurring on Instagram leads to more visual attention to the disclosure, whether this leads to more critical processing of the advertisement, and whether this relationship is mediated by users' visual attention to the advertisement. In a preregistered one-factorial between-subjects laboratory eye-tracking experiment (N = 134) we exposed participants to an Instagram timeline containing either a sponsorship disclosure (control condition) or a more salient and informative targeting disclosure (experimental condition). Results show that the targeting disclosure did not lead to higher levels of visual attention to the disclosure, but that attention to the disclosure was positively related to more critical processing as well as more visual attention to the ad. Nevertheless, we did not find a mediating effect of attention to the advertisement on the relationship between attention to the disclosure and critical processing. Finally, we believe that these disclosures represent a step forward in improving transparency and shielding users.

Keywords: microtargeting, transparency, eye-tracking, disclosures, political

Introduction

Recently, strategies used in political advertising have shifted from canvassing–like door-to-door interactions with potential voters–to online interactions. The majority of communication between voters and politicians occurs online, more specifically on social networking sites (SNS) (Giasson et al., 2019). This shift to an online environment creates opportunities for political advertisers. Online platforms collect the interactions and behavior of their users (e.g., liking posts, commenting on videos, and exposure time to content). This information can be used to build profiles of certain user groups and allocate them into small groups to target them with specific messages (Murray & Scime, 2010). These messages are developed to resonate more effectively within these groups, a practice called *microtargeting*, or in a political context: *political microtargeting* (PMT) (Binder et al., 2022; Kruikemeier et al., 2022; Zuiderveen Borgesius et al., 2018).

The most commonly known example of PMT is when *Cambridge Analytica*, a British consultancy firm, gathered and used data from a large number of Facebook users to form psychological profiles and, based on these profiles, sent target messages to users that would persuade them as much as possible (Cadwalladr, 2018). Hereby, the firm allegedly contributed to Trump's presidential race and the Leave campaign's success in the Brexit referendum. However, PMT also has the potential to benefit political systems within societies. The technique has the potential to, for instance, activate voters who are usually deemed to have a lower propensity to vote by reaching out to them in personally relevant ways with messages that potential voters really care about (Zarouali et al., 2020), and also strengthen general political interest (Matthes et al., 2022) and general political knowledge (Holt et al., 2013).

While at first glance PMT might seem to be an effective strategy in political campaigning, the phenomenon also introduces risks that might not outweigh the benefits. For

instance, PMT discourages political participation by showing people more issues that they do not find interesting (Bodó et al., 2017) and decreases scrutiny through the exclusion of certain target groups if they are not interesting for advertisers (Jamieson, 2013). Critics claim that users' privacy is at risk when profiles are developed based on personal information and (online) behavioral data, which occurs mostly without their informed consent (Bennett, 2015; Dobber et al., 2019; Tene, 2011). A major problem with PMT is that it may manipulate voters by targeting them while they are unaware of it (Zuiderveen Borgesius et al., 2018). Users generally do not recognize that they have received persuasive messages that are specifically targeted at them (Binder et al., 2022; Kruikemeier et al., 2016). A potential solution is the use of disclosures.

Disclosures

Disclosures are labels that are usually implemented to inform users about the sponsored or commercial nature of a message (e.g., advertorials and YouTube reviews). Within research on sponsored content, disclosures have been shown to facilitate the recognition of advertisements (Amazeen & Wojdynski, 2020; Boerman et al., 2015). On platforms such as Facebook, Instagram, and Twitter, disclosures are a measure to improve transparency, while simultaneously providing users with tools to distinguish sponsored content from regular content (Boerman & Van Reijmersdal, 2016; Janssen et al., 2021).

Regulatory bodies are starting to implement regulations to provide users with disclosures as a transparency measure. A prime example is the European Commission's Digital Service Act (DSA) (European Commission, 2022). The DSA states that targeted messages and ads should include information on when and on whose behalf, content is displayed. Simultaneously, these measures aim to go beyond that and include background information about the individual data used to target users and parameters for those data points (European Commission, 2022; King, 2022). In the Dutch context, the DSA would be the first

time that targeting disclosures have been initiated. However, other regulators have attempted to implement targeting disclosures in their markets. In Brazil, the 'fake news bill', which also includes targeting transparency on platforms, is being reviewed in Congress (Lubianco, 2022). Two examples within the EU are a proposal in Ireland (Online Advertising and Social Media (Transparency) Bill, 2021) and an adapted version in France (Dobber, 2021; Lutte contre la manipulation de l'information, 2018).

One of the main implications of current disclosure research is that the prominence, positioning, and degree to which disclosures stand out, in contrast to regular content on a timeline, are important (Amazeen & Wojdynski, 2020; Wojdynski & Evans, 2016). Moreover, a combination of text and a logo or symbol leads to the highest visual fixation (Boerman et al., 2015). If disclosures are more prominent (i.e., bigger, containing more text, or more colorful), they stand out compared to regular content on a timeline. In turn, they lead to higher levels of attention and clearer perceptions of disclosed information, thus contributing to greater transparency (Binder et al., 2022; Kruikemeier et al., 2016). The currently used disclosures are mostly designed with a grey and thin font (i.e., lacking prominence) that state content is sponsored and by whom it is paid for regarding political ads. However, research shows that these disclosures seem insufficient in informing users and increasing transparency (Binder et al., 2022; Binford et al., 2021; Jansen & Krämer, 2023; Jost et al., 2022).

Nonetheless, since 2020, differently designed disclosures have been used by platforms. During the COVID-19 pandemic, Instagram has implemented measures to counter the spread of misinformation. One of these measures was to place disclosures on posts and stories (i.e., disappearing posts) that informed users that the information could be false or not yet proven. The disclosure was located below the content and was designed as a bar in the system font and color. Moreover, the disclosure contained a large "I" symbol to show that it

contained additional information and a link to the 'COVID-19 Information Center' (Clark, 2020; Instagram, 2020). The design choices made regarding these disclosures might be a solution to the regulations in the DSA. While these disclosures provide platforms with more space to disclose information, they could also attract more attention as they are more prominent. To the best of our knowledge, these specific disclosures have not yet been investigated by other researchers as transparency measures for PMT. Since more legislation and regulations seem to appear concerning ads on SNSs (e.g., the DSA), the aforementioned COVID-19 disclosures have the potential to be a solution for platforms to comply with these regulations.

Visual attention to disclosures

For disclosures to work and contribute to transparency users need to see them first, which is why eye-tracking is deemed a fitting method to investigate the effectiveness of disclosures. Recent work combining an experiment and eye-tracking in the context of disclosures for influencer marketing on Instagram shows that while users seem to be aware of influencer marketing they also make mistakes in identifying it as such within a feed (Boerman & Müller, 2021). Nonetheless, users pay the most attention to paid partnership' labels (compared to no sponsor cue, #paidad or a brand tag). In turn, this influences the recognition of ads through conceptual persuasion knowledge (i.e., ad recognition, understanding of persuasive intent, and understanding of the commercial source) (Boerman & Müller, 2021). Moreover, research on product placement disclosures in television programs showed that a combination of a logo and text was more effective concerning users' recognition of ads than either of those aspects separately (Boerman et al., 2015).

In the field of PMT, research shows that even though they are promising, targeting disclosures do not always get recognized or recalled by users (Binder et al., 2022; Dobber et al., 2023; Jost et al., 2022; Kruikemeier et al., 2016). However, in one of the first eye-

tracking studies regarding disclosures on microtargeted ads, Jost and colleagues (2022) found that while a prominent disclosure did not lead to more visual attention to the disclosure, participants were still better able to recall sponsored ads in a timeline. Oppositely, in another eye-tracking experiment, Binford and colleagues (2021) found that although a political ad disclosure on Facebook did not effectively enhance users' comprehension of who paid for the political ad, users did pay visual attention to the disclosure. Therefore, we expect that a targeting disclosure (i.e., a disclosure consisting of a sponsorship element and a targeting element) that provides more information and is more prominent because it consists of more text, leads to more visual attention to the disclosure compared to a sponsorship disclosure (i.e., a disclosure consisting of merely sponsorship elements). This leads to the following hypothesis:

> H1: A more prominent and informative targeting disclosure (compared to a sponsored disclosure) leads to more visual attention towards this disclosure.

Critical processing

One of the constructs that have been identified as important with regard to disclosures and their reception, but is not yet studied in the domain of disclosures and PMT, is critical processing (Boerman & Van Reijmersdal, 2016). When users are made aware of a persuasive attempt, they have to figure out how they manage their response to this attempt effectively (Friestad & Wright, 1994). As a result of this, the awareness of that persuasive attempt encourages more systematic and biased processing of the advertisement (Janssen et al., 2010). Critical processing is defined as the adoption of an evaluative style of processing in which the content is criticized (Boerman et al., 2014, p. 217). This style of processing has been found to get elicited through the use of sponsoring disclosures (compared to no disclosure) in the case of sponsored content on television (Boerman et al., 2014). Moreover,

Zarouali and colleagues (2017) showed that adolescents engage in critical reflection of targeted advertising if they are exposed to a cue that triggers their critical elaboration. In other work, Zarouali and colleagues (2018) found that if adolescents get exposed to a disclosure in a timeline regarding the targeted nature of a Facebook ad, they engage in more critical processing compared to a dropdown menu containing that information, or an ad without any transparency measure. When a disclosure enhances the recognition of a post being an ad instead of organic content, users could realize that the post is indeed an ad, not neutral, and has a persuasive purpose. Given the embedded design of the ad, users might even feel deceived. Because of this awareness and feeling of deception, users might cope with the persuasive attempt by criticizing it and being more skeptical while interpreting the message (Boerman et al., 2014). Moreover, awareness of the persuasive intent has been shown to lead to less favorable perceptions of the communicator (Allyn & Festinger, 1961; Boerman et al., 2014; Campbell & Kirmani, 2000). Thus, we expect that visual attention to the disclosure leads to more critical processing, which leads to the following hypothesis:

> H2: Visual attention towards the disclosure is positively related to critical processing of the advertisement.

Visual attention to the advertisement

However, we propose that another mechanism might get activated through visual attention to the disclosure. Disclosures can not only protect users from the persuasive effects of the ad but also have the potential of unintentionally increasing, rather than decreasing the persuasive effects (Janssen et al., 2016). The main reasoning is that when users allocate more visual attention to a disclosure and read its message, they might become curious about a political party's perception of them. This curiosity, in turn, might lead to an interest in the party's persuasive message. This could mean that attention to the disclosure would lead to

more attention to the advertisement which in turn, assuming that the advertisement stays as persuasive as it is intended to be, might be a negative unintended effect.

Prior research shows that, indeed, viewers' recognition of advertising increases their brand memory (Boerman et al., 2015), which could be seen as an outcome of a persuasive message. Moreover, earlier work by Evans and colleagues (2019) shows that sponsorship transparency can have a positive effect on receivers' attitude towards the ad and on receivers' purchase intention. Furthermore, Pan and Zinkhan (2006) found that a privacy disclosure message can lead to more trust in an online retailer. Besides, Zarouali and colleagues (2018) found that when adolescents get exposed to a disclosure in a timeline explaining the targeted nature of a Facebook ad, they perceive the ad as more persuasive. Other work finds that if platforms and/or advertisers are transparent about the personalization practices taking, this decreases perceptions of vulnerability and as a result, consumers are more inclined to engage with advertisements (Aguirre et al., 2015). In this study, we view visual attention to the advertisement as a proxy of users' interest in the advertisement itself, which could eventually lead to the brand –or party– related effects mentioned above. Thus, we expect that users' visual attention to the disclosure relates to more visual attention towards the advertisement itself, leading to the following hypothesis:

> H3: Visual attention towards the disclosure is positively related to visual attention towards the advertisement.

While most research that is mentioned above uses the effectiveness of an advertisement as an outcome measure, we aim to investigate our initial dependent variable, critical processing, as such. Therefore, in this work, we also investigate the potential mediating effect of users' visual attention to the advertisement on the effect of users' visual

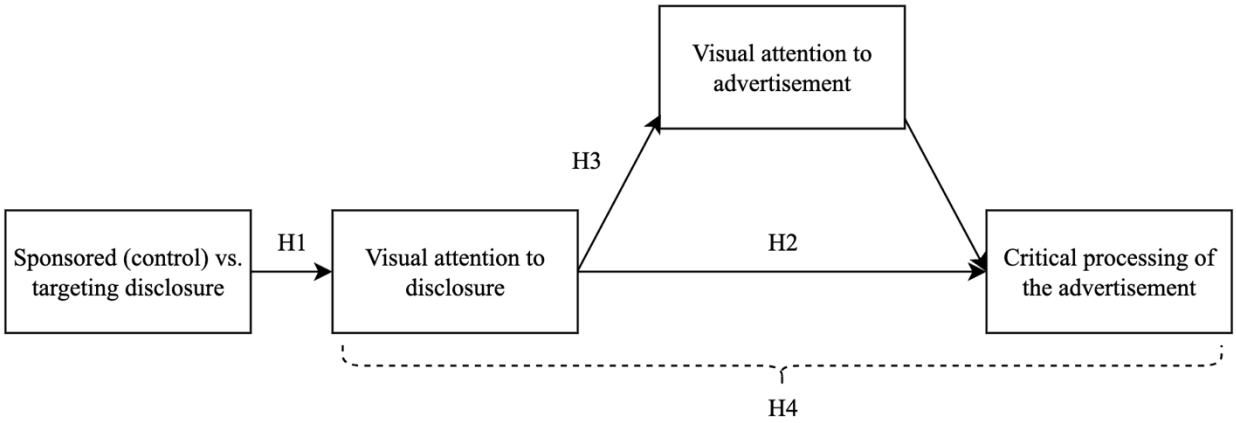
attention to the disclosure and their critical processing of the ad. This leads to the following hypothesis:

> H4: Visual attention towards the disclosure is positively related to visual attention towards the advertisement which in turn positively relates to critical processing of the advertisement.

The proposed path model visualizing our hypotheses can be found in Figure 1.

Figure 1

Proposed path model



Method

The study was approved by the ethics committee of Radboud University under approval number ECSW-LT-BC-2023-3-30-47641We preregistered this study before collecting data: https://osf.io/mpurd. Every participant gave written informed consent before participation. Supplementary materials, stimuli and areas of interest and measures are publicly accessible on OSF (original stimulus material, areas of interest and timelines are not accessible due to local references, they will be after review): https://osf.io/zkjcb/?view_only=6d884a3377534df4a83151314c76cc28.

Design

To test our hypotheses we conducted an eye-tracking laboratory experiment with a factorial between-subjects design containing two groups. Participants were exposed to an Instagram timeline containing six Instagram posts in total: three posts from Dutch news outlets, a post from a local hotel/restaurant, the university (which were all filler stimuli), and subsequently from a Dutch political party which contained either a sponsored disclosure (control condition) or a targeting disclosure (experimental condition) which can be found in Figure 2. The political advertisement was the third post in the timeline. The image of the full timelines can be found on OSF.

In both conditions, participants were exposed to the disclosure that Meta uses on Instagram to inform users that content is sponsored ('Sponsored', between the ad and account name). Moreover, underneath the advertisement, users were exposed to a more prominent disclosure, designed based on the disclosures Meta used in the COVID-19 pandemic. This disclosure underneath the post contained an "I" to show users that there was extra information to be found there. For participants in the control condition, the disclosure explained that the content was sponsored (in line with the term "sponsored" that was already visible above the ad). However, in the experimental condition, the disclosure contained additional information about the targeting practices taking place on the platform and that the ad was targeted at the users based on their personal information and online behavior. Thus, in both conditions a sponsorship disclosure was visible above the ad, while in only the experimental condition additional targeting information was presented. To ensure that possible effects of the manipulation would be due to the targeting disclosure and not just any text or disclosure being visible underneath the advertisement, in the control condition the sponsorship disclosure was repeated underneath the ad (See Figure 2).

To ensure that most of our participants at least could have been targeted by this advertisement in a real-world setting, we used the same political advertisement from two Dutch parties, one more center-left (Partij van de Arbeid; here "Labor") and one more center-right (Democrats 66; here "D66"), participants got shown one of the advertisements through random selection. Besides, since we predicted that the majority of our sample would be students and we aimed to make the ad about a relevant topic, the topic of the ad was the housing market, which is one of the most important political issues in the Netherlands (Statista, 2023) and stated: 'Housing is a fundamental right, not something that should be speculated with'. The ad was designed in the corporate design of both parties and contained a logo and the same picture of houses, we replicated the design from two local bodies of both political parties that we found through the Facebook Ad Library (Labor: https://www.facebook.com/ads/library/?id=122486037322954, D66: https://www.facebook.com/ads/library/?id=1240248669857229). The advertisement for the Labor Party also included a part of their larger logo (a rose), which was not the case for D66 since their logo is just the abbreviation of their name. On both Instagram posts, the number of likes (188) and comments (12) was the same, the ads differed from regular content regarding the timestamp of posting, which Instagram does not provide for ads. The account names were from the local bodies of both parties (@pvdanijmegen and @nijmegen d66 and we used their profile pictures (in April 2023) as profile pictures.

Figure 2

Stimuli: Instagram posts for both the control condition (left) and targeting disclosure condition (right) exemplified for one of the political parties.

Measures

In this study, we measured both visual attention to the disclosure and visual attention to the ad in seconds, with the eye-tracker. To measure attention to the disclosure two areas of interest (AOIs) were combined into a measurement for visual attention to the disclosure in seconds To measure attention to the ad we did the same thing but with different AOIs We combined participants' visual fixation on the account name, the image, and the caption below the image into a measure for visual attention to the ad in seconds. Finally, as a measure for the critical processing of the ad, we adapted two items that have been used in targeting research by Zarouali and colleagues (2018) but originally were used in disclosure research by Boerman and colleagues (2014) on a Likert scale ranging from 1 (= strongly disagree) to 7 (= strongly agree): *"While scrolling the Instagram timeline I criticized the advertisement of the political party"*, and *"While scrolling the Instagram timeline I was skeptical toward the*

advertisement of the political party". We computed the mean for these two items as our measure for critical processing (Cronbach's α = .76, McDonald's ω = .76).

Manipulation check

As a manipulation check, we asked participants if they recalled if there was a disclosure on the Instagram ad and what the disclosure stated by asking them to check one of the following statements: *"The Instagram post was a regular post"*, *"The Instagram post was labeled as targeted at me and sponsored"*, or *"The Instagram post was labeled as sponsored"*.

Besides, as a stimulus check, we asked participants if they thought that this advertisement could have been shown to them while browsing a SNS in real life on a Likert scale ranging from 1 (= definitely not) to 5 (= definitely yes; M = 3.42, SD = 1.15).

Sample

We recruited 150 participants that had at least one active social media account, through the online participant system of the university or by approaching them on campus between the 18th of April and the 15th of May 2023. We excluded participants with a weighted gaze percentage of less than 80%, leaving us with 134 participants (see Procedure). Participants' age ranged between 18 to 35 years (M = 22.73, SD = 2.96). Of these participants, 85 identified as female, 45 as male, 3 as non-confirming, and 1 as a transgender male. Regarding education, most of our participants had a high school diploma (n = 70), 55 participants had a bachelor's degree, 8 had a master's degree and 1 participant had a vocational education diploma. Randomization checks showed no differences between our two experimental groups regarding age (F(1, 132) = 0.05, p = .817), gender (χ2(3, N = 134) = 1.46, p = .691), or level of education (χ2(3, N = 134) = 7.50, p = .058). Moreover, there were no differences in which version of the political advertisement (D66 or Labor) participants saw between our experimental groups (χ2(1, N = 134) = 0.48, p = .489).

Power

Given the budget for this study, we were able to gather responses from 150 participants. As described, we only analyzed data from participants with a weighted gaze percentage above 80, leaving us with a final sample of 134 participants. To determine our statistical power, we used R (version 4.1.2; R Core Team, 2021) and the simsem package (version 0.5-13; Pornprasertmanit et al., 2021). The power analysis aimed to set the smallest effect size that can be found with a recommended minimum power of 90% (Curran-Everett, 2017). The smallest effect size of interest (SESOI; Lakens et al., 2018) is complementary to and comparable with the alpha-level, i.e., a threshold for rejecting results. We calculated our SESOI while using a power level of 90%, an α level of .05, and a sample size of 134 participants in total. Consequently, we set our SESOI at $\beta = |.28|$ and do not interpret results with effect sizes below that.

Procedure

Participants were invited to the laboratory and were informed that their eye movements and visual fixation would be recorded while they were looking at social media content. After calibrating the eye-tracker, we informed participants that they would see an Instagram timeline that they should browse ordinarily and that they could exit the timeline by pressing a button at the bottom of the screen. After the instruction, the Instagram timeline appeared, and visual fixation was recorded while the participants scrolled through the timeline. Next, participants were asked to fill in a questionnaire containing questions to measure the dependent variable and their demographic information. Finally, we debriefed and thanked our participants. The average time of completion for viewing the stimuli and answering the questions was 2 minutes and 19 seconds in total.

Apparatus

Visual fixation was recorded with a Tobii TX 300 eye-tracker (sampling rate of 300Hz). Participants were seated 60–70cm away from the monitor. A 9-point calibration was performed before the start of the experiment. We excluded participants that had a weighted gaze of 80 percent during stimulus reception. This is a method for determining the quality of a recording which consists of a ratio of the number of samples that the eye-tracker correctly identified divided by the attempts. Finally, 134 participants met the criteria and could be included in data analyses (average proportion of recorded weighted gaze data: $M = 95.01\%$, $SD = 3.85$).

Results

We conducted the statistical analyses for this work using R (version 4.1.2; R Core Team, 2021) and jamovi (version 2.3; The jamovi project, 2022). To investigate differences between our conditions we conducted a path model through structural equation modeling in lavaan (version 0.6-12.; Rosseel, 2012). The code that we used in both software packages is available on the OSF: https://osf.io/zkjcb/?view_only=6d884a3377534df4a83151314c76cc28. Means, standard deviations, and bivariate correlations for our measured variables can be found in Table 1. Descriptives per advertisement per group can be found on OSF. Other than preregistered, we changed the wording for our hypotheses to non-causal language (i.e., changed from lead to, to be related with).

Table 1

Means, Standard Deviations, and Bivariate Correlations of the Measured Constructs.

Measured construct	*M* (*SD*)	1	2	3
1 Experimental condition	-	-		
2 Visual attention to disclosure in *s*	2.20 (2.47)	.24**	-	

3 Critical processing	3.53 (1.46)	0.09	.37***	
4 Visual attention to ad in *s*	6.69 (4.33)	-.00	.59***	.19*

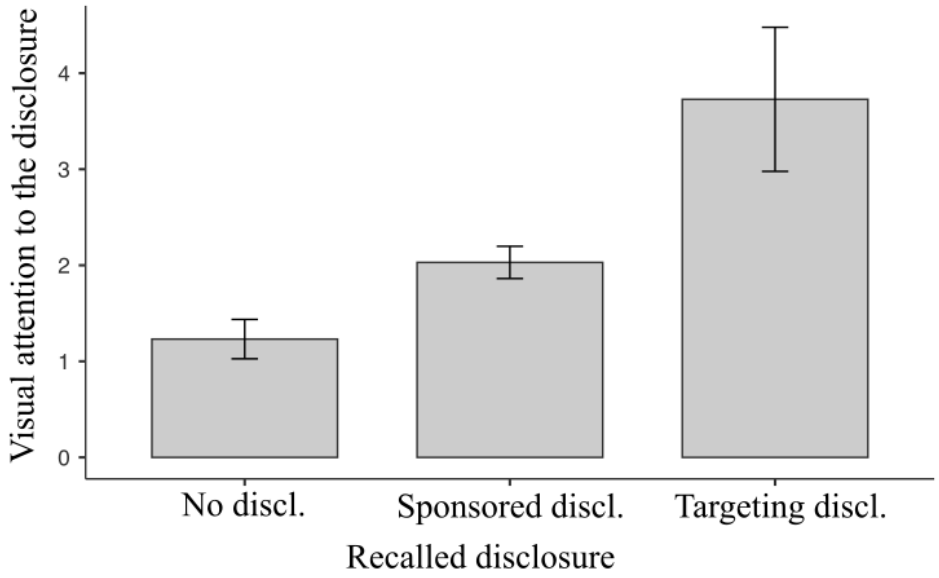
Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

Manipulation check

Concerning our manipulation check, in our sponsored disclosure condition ($n = 67$), 40 participants correctly recalled this disclosure, while 21 participants did not recall a disclosure at all and 6 participants wrongfully recalled a targeting disclosure. In our targeting disclosure condition ($n = 67$), 25 participants recalled the correct disclosure, while 25 other participants wrongfully recalled a sponsored disclosure and 17 participants did not recall a disclosure at all. However, because of the eye-tracking we were able to investigate the recalled disclosure and participants' visual attention to the disclosure. Interestingly, we found that the participants recalling no disclosure, had the lowest visual attention to the disclosure, followed by participants that recalled a sponsored disclosure, and finally that participants who recalled a targeting disclosure, spent the most seconds looking at the disclosure. A visual representation of these results can be found in Figure 3.

Figure 3

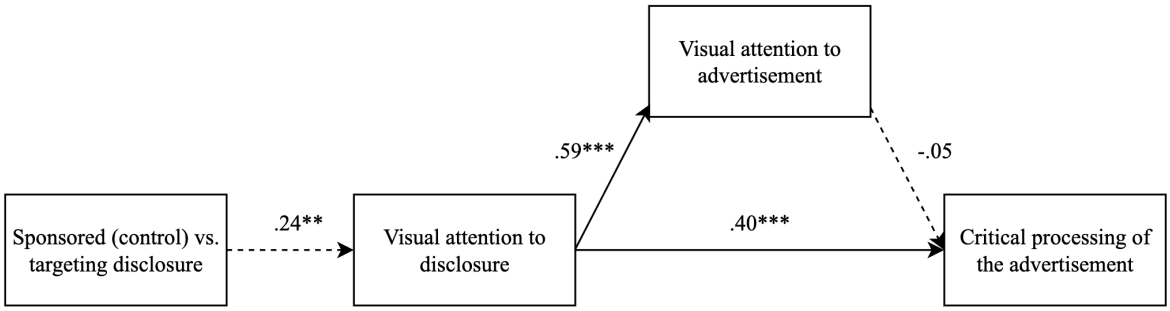
Visual attention to the disclosure in seconds and recalled disclosure.



As preregistered, we tested all our hypotheses in an integrated path model. We evaluated our model fit in line with frequently used fit indices (Hooper et al., 2008). Our model showed a near decent fit: $\chi2(2) = 4.57$, $p = .102$, $\chi2/df = 2.29$, CFI = 0.97, TLI = 0.91, RMSEA = .10, 90% CI [.00, .22], SRMR = .04. Our path model is shown in Figure 4.

Figure 4

Path Model



Note. Numbers represent standardized regression coefficients. Dashed lines indicate a path is either statistically meaningless (not significant) or theoretically meaningless (below β = |.15|). * $p < .05$, ** $p < .01$, *** $p < .001$.

For our first hypothesis, we expected that if an Instagram ad contained a targeting disclosure, this would affect participants' visual attention to the disclosure. Our results show a significant positive effect (β = .24, $p = .004$). However, this effect size is below our SESOI (β = |.28|), and thus, we are not able to interpret this effect, leading us to reject our first hypothesis. For our second hypothesis, we predicted that visual attention towards the disclosure would be positively related to users' critical processing of the message. We found proof for this in our data (β = .40, $p < .001$), leading us to accept this hypothesis. For our third hypothesis, we expected that visual attention towards the disclosure, would be positively related to visual attention towards the ad itself. We found proof for this in our data, (β = .59, $p < .001$), leading us to accept this hypothesis.

Finally, for our fourth hypothesis, we predicted that visual attention to the disclosure would be positively related to users' visual attention to the ad and that this, in turn, would positively relate to users' critical processing of the ad. We investigated this through mediation analyses in our path model in lavaan (version 0.6-12; Rosseel, 2012) while bootstrapping with 5,000 samples. We found no proof for mediation within our data as our indirect effect was not significant (indirect effect = -0.02, *SE* = 0.04, 95% *BCBCI* [-0.09, 0.05]), leading us to reject hypothesis four.

Discussion

In this work, we aimed to investigate whether a more informative targeting disclosure on a political microtargeted ad on Instagram would lead to more attention to the disclosure, if

this led to more critical processing, and if this effect gets mediated by attention to the ad itself. To do so, we investigated targeting disclosures through a laboratory eye-tracking experiment in which we showed participants an Instagram timeline containing a post from a political party containing either a sponsored or a targeting disclosure.

Firstly, we found that exposure to a targeting disclosure, compared to a sponsored disclosure in our control condition, led to more attention to the disclosure (Binford et al., 2021; Boerman et al., 2015). Even though prior research shows mixed results concerning users' attention to a disclosure, we found a significant effect of our targeting disclosure on attention to this disclosure. Nevertheless, the effect size that our data showed, was below our SESOI and thus not meaningfully interpretable. However, the results also indicates that the more prominent targeting disclosure was positively correlated with users' visual attention, which might imply prominence being an important factor regarding disclosures (Binford et al., 2021; Jost et al., 2022)

Besides the effect of our disclosure type on attention to this disclosure, we found that that users that had more visual fixation on the disclosure, engaged in more critical processing of the advertisement (Boerman et al., 2014; Zarouali et al., 2017). Even though we were not able to conclude that the targeting disclosure, compared to the mere sponsorship disclosure, increased visual attention, we still think that the disclosure had an effect on subsequent processing of the advertisement. The rationale behind this is that even without a correct recalling of the disclosure that states the ad is targeted; the majority of participants still recalled a disclosure. Therefore, the disclosure could have been effective in helping users distinguish the ad from regular content which could have helped them be aware of the persuasive attempt (Friestad & Wright, 1994), and subsequently manage their response by evaluating the ad systematically and critically (Boerman et al., 2014; Janssen et al., 2010).

However, looking at the low and non-significant correlation between our conditions and critical processing, this is something that we can able to speculate about.

Furthermore, we found that attention to the disclosure would lead to more attention to the advertisement. This effect can be an unintentional effect of disclosures where the label, besides protecting users from the persuasive effects of the ad, increases rather than decreases the persuasive effects (Janssen et al., 2016). This is in line with existing work that shows that viewers' recognition of an ad increases their brand memory (Boerman et al., 2015). Besides, other research shows that sponsorship transparency can have a positive effect on receivers' attitude towards the advertisement (Evans et al., 2019). Moreover, in privacy and e-commerce research, a disclosure about privacy has been shown to lead to more trust in a retailer (Pan & Zinkhan, 2006). Finally, earlier work on microtargeting shows that users' awareness of the persuasive meaning and the personalization taking place on the platform, positively influences party evaluations (Binder et al., 2022). A reason for this effect might be that users are curious about the microtargeting, in a skeptical way that is different from critical processing of the ad. This might imply that users, if they are made aware of the persuasive nature of the ad as well as the targeting that took place to deliver that specific ad to them via their timeline, have a surveillance motive and are curious about the advertisement (Choi, 2016; Flavián & Gurrea, 2009). This could even lead to them allocating more cognitive resources toward the advertisement because they feel like they are researching what the platform thinks it knows about them as an individual.

Regarding our final hypothesis, we expected a mediating effect of attention to the ad on the effect of attention to the disclosure on critical processing. However, we did not find this mediating effect of attention to the ad in our data. This could imply that even though we think that people could be curious about the advertisement that is shown to them because they are informed that it is a persuasive message that is targeted to them, this might not lead them

to be more critical about it after they spend some seconds on the advertisement. However, it is also a possibility that users just do not care that much, or do not have cognitive responses after receiving the advertisement itself, as has been found in work regarding cognitive resistance to advertisements (van Reijmersdal et al., 2016).

Within existing disclosure research, results regarding correct recall of the disclosure are very diverse, and disclosures do not always get noticed (Binder et al., 2022; Evans et al., 2017; Jost et al., 2022; Kruikemeier et al., 2016). The current study's results are in line with this regarding recall. In our control condition with the sponsored disclosure 60% of our participants recalled the correct disclosure, while in the experimental condition 37% of our participants recalled the correct disclosure, and the same number of participants incorrectly recalled a sponsored disclosure. However, if we combine this information with the visual attention to the disclosure, we find that participants who recalled the targeting disclosure had 66% more fixation time on the disclosure compared to participants that recalled the sponsored disclosure. Even though it was not a core question in this study, we still think that the finding that 74% of our participants at least recalled a disclosure, is a step forward compared to other work (Binford et al., 2021; Jost et al., 2022; Kruikemeier et al., 2016). Moreover, the finding that the same number of people recalled a sponsored and targeting disclosure in the targeting condition and that 60% of people recalled the correct disclosure in the control condition, might imply that users need some time to adapt to this new type of disclosure. Over time, they might internalize it and subsequently enhance their recall of it, as they did with the sponsored disclosure.

Limitations and future work

In this work, we aimed to randomize the political ads participants saw by using a center-left and center-right party out of the Dutch party system. Besides, we choose the statement about the housing market to at least try to make the ad suitable for a sample that

would be mainly students. Even though our randomization check did not show any differences between our two conditions, this does not specifically mean that the content was targeted at the participants. The field of microtargeting research would benefit from more studies that use large-scale field experiments with actual targeting data and existing advertisements, to improve ecological validity. This data and these ads could be gathered through web scraping or even data donations. Furthermore, a combination of those designs combined with surveys or experiments that measure attitudes and evaluations towards PMT as a technique and the advertisements themselves would also benefit the field.

Also, in the current study, we had an initial sample size of 150 participants that mainly were students. While this is more than the average of 73 participants and two-thirds being students, in studies published in the top 25 communication journals between 2005 and 2019 (King et al., 2019), future research might find different results with more diverse and larger samples. Moreover, we think that future work would benefit the field, as well as legislators and regulators, by taking a more practical approach and scrutinizing more precisely what design factors influence the correct recalling and remembrance of disclosure.

We would like to emphasize that the difference in color regarding the advertisement could have played a role in participants visual attention to the ad, as red is a color that attracts attention quicker than green, which our data confirm (Visual attention to the ad: $M_{D66} = 6.08$, $SD = 4.07$; $M_{PvdA} = 7.28$, $SD = 4.52$) (Baik et al., 2013). Furthermore, this study was carried out in a laboratory and that this could have influenced our results as participants in a laboratory could have been very aware that they took part in a study. However, while eye-tracking studies enable us to gain valuable insights into participants' attention distribution, it is also a possibility that users perceive parts of the stimuli through peripheral vision without actually fixating on an area of interest. Besides, the timeline that participants saw, was not their own, which might lead to an even stronger consciousness that they were in a laboratory

viewing a made-up timeline. Finally, this study did not have a control condition without a disclosure, since without a disclosure there would be no visual attention to the disclosure that would be measurable. However, we advise future research to include a condition like that in order to scrutinize if users process a message more critically due to a targeting or sponsored disclosure, compared to no disclosure.

Conclusion

Political microtargeting is something that users on social media do not actively recognize in most cases. Nevertheless, regulatory bodies, legislative authorities, and researchers alike agree that there is a need for improving transparency. The current study investigated the effects of targeting disclosures on attention to these disclosures and the effects of attention to the disclosure on critical processing of, and attention to, the ad. Finally, we investigated the potential mediation of attention to the ad on the effect of attention to the disclosure and critical processing. The results show that a more prominent disclosure was not related to more attention to the disclosure. But that visual attention, in turn, was positively related to both critical processing and attention to the displayed ad. Finally, attention to the ad did not relate to critical processing and attention to the ad did not mediate the relation between attention to the disclosure and critical processing of the ad. Together, these findings show that there is a potential for newly designed and more prominent disclosures, and that these would be able to inform users about the targeting taking place while obliging with the Digital Services Act. However, users that see the disclosure, might process an ad more critically, which could be fruitless for political advertisers.

Research Transparency Statement

The authors are willing to share their data, analytics methods, and study materials with other researchers. The material will be available at OSF.

Preregistration Statement

The authors have pre-registered this research with an analysis plan which is retrievable at: https://osf.io/mpurd/?view_only=0cec88db924245fd88d1255922417946

References

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing*, *91*(1), 34–49. https://doi.org/10.1016/j.jretai.2014.09.005

Allyn, J., & Festinger, L. (1961). The effectiveness of unanticipated persuasive communications. *The Journal of Abnormal and Social Psychology*, *62*(1), 35–40. https://doi.org/10.1037/h0048493

Amazeen, M. A., & Wojdynski, B. W. (2020). The effects of disclosure format on native advertising recognition and audience perceptions of legacy and online news publishers. *Journalism*, *21*(12), 1965–1984. https://doi.org/10.1177/1464884918754829

Baik, M., Suk, H.-J., Lee, J., & Choi, K. (2013). *Investigation of eye-catching colors using eye tracking* (B. E. Rogowitz, T. N. Pappas, & H. De Ridder, Eds.; p. 86510W). https://doi.org/10.1117/12.2001141

Bennett, Collin. J. (2015). Trends in Voter Surveillance in Western Societies: Privacy Intrusions and Democratic Implications. *Surveillance & Society*, *13*(3/4), 370–384. https://doi.org/10.24908/ss.v13i3/4.5373

Binder, A., Stubenvoll, M., Hirsch, M., & Matthes, J. (2022). Why Am I Getting This Ad? How the Degree of Targeting Disclosures and Political Fit Affect Persuasion Knowledge, Party Evaluation, and Online Privacy Behaviors. *Journal of Advertising*, 1–17. https://doi.org/10.1080/00913367.2021.2015727

Binford, M. T., Wojdynski, B. W., Lee, Y.-I., Sun, S., & Briscoe, A. (2021). Invisible transparency: Visual attention to disclosures and source recognition in Facebook

political advertising. *Journal of Information Technology & Politics*, *18*(1), 70–83. https://doi.org/10.1080/19331681.2020.1805388

Bodó, B., Helberger, N., & De Vreese, C. H. (2017). Political micro-targeting: A manchurian candidate or just a dark horse? *Internet Policy Review*, *6*(4). https://doi.org/10.14763/2017.4.776

Boerman, S. C., & Müller, C. M. (2021). Understanding which cues people use to identify influencer marketing on Instagram: An eye tracking study and experiment. *International Journal of Advertising*, 1–24. https://doi.org/10.1080/02650487.2021.1986256

Boerman, S. C., & Van Reijmersdal, E. A. (2016). Informing Consumers about "Hidden" Advertising: A Literature Review of the Effects of Disclosing Sponsored Content. In P. De Pelsmacker (Ed.), *Advertising in New Formats and Media* (pp. 115–146). Emerald Group Publishing Limited. https://doi.org/10.1108/978-1-78560-313-620151005

Boerman, S. C., Van Reijmersdal, E. A., & Neijens, P. C. (2014). Effects of Sponsorship Disclosure Timing on the Processing of Sponsored Content: A Study on the Effectiveness of European Disclosure Regulations: EFFECTS OF SPONSORSHIP DISCLOSURE TIMING. *Psychology & Marketing*, *31*(3), 214–224. https://doi.org/10.1002/mar.20688

Boerman, S. C., Van Reijmersdal, E. A., & Neijens, P. C. (2015). Using Eye Tracking to Understand the Effects of Brand Placement Disclosure Types in Television Programs. *Journal of Advertising*, *44*(3), 196–207. https://doi.org/10.1080/00913367.2014.967423

Cadwalladr, C. (2018, March 18). 'I made Steve Bannon's psychological warfare tool': Meet the data war whistleblower. *The Guardian*.

https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump

Campbell, M. C., & Kirmani, A. (2000). Consumers' Use of Persuasion Knowledge: The Effects of Accessibility and Cognitive Capacity on Perceptions of an Influence Agent. *Journal of Consumer Research*, *27*(1), 69–83. https://doi.org/10.1086/314309

Choi, J. (2016). Why do people use news differently on SNSs? An investigation of the role of motivations, media repertoires, and technology cluster on citizens' news-related activities. *Computers in Human Behavior*, *54*, 249–256. https://doi.org/10.1016/j.chb.2015.08.006

Clark, M. (2020, December 17). *Instagram is rolling out new notifications about COVID-19 information*. The Verge. https://www.theverge.com/2020/12/17/22187298/instagram-coronavirus-covid19-misinformation-notifications

Curran-Everett, D. (2017). CORP: Minimizing the chances of false positives and false negatives. *Journal of Applied Physiology*, *122*(1), 91–95. https://doi.org/10.1152/japplphysiol.00937.2016

Dobber, T. (2021). *Effectiveness of Online Political Ad Disclosure Labels: Empirical Findings*. 6.

Dobber, T., Kruikemeier, S., Helberger, N., & Goodman, E. (2023). Shielding citizens? Understanding the impact of political advertisement transparency information. *New Media & Society*.

Dobber, T., Ó Fathaigh, R., & Zuiderveen Borgesius, F. J. (2019). The regulation of online political micro-targeting in Europe. *Internet Policy Review*, *8*(4). https://doi.org/10.14763/2019.4.1440

European Commission. (2022). *Regulation of the European Parliament and of The Council on a Single Market For Digital Services (Digital Services Act) and amending Directive*. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN

Evans, N. J., Phua, J., Lim, J., & Jun, H. (2017). Disclosing Instagram Influencer Advertising: The Effects of Disclosure Language on Advertising Recognition, Attitudes, and Behavioral Intent. *Journal of Interactive Advertising*, *17*(2), 138–149. https://doi.org/10.1080/15252019.2017.1366885

Evans, N. J., Wojdynski, B. W., & Grubbs Hoy, M. (2019). How sponsorship transparency mitigates negative effects of advertising recognition. *International Journal of Advertising*, *38*(3), 364–382. https://doi.org/10.1080/02650487.2018.1474998

Lutte contre la manipulation de l'information, 2018, 1202, 0297 (2018). https://www.senat.fr/dossier-legislatif/ppl17-623.html

Flavián, C., & Gurrea, R. (2009). Users' motivations and attitude towards the online press. *Journal of Consumer Marketing*, *26*(3), 164–174. https://doi.org/10.1108/07363760910954109

Friestad, M., & Wright, P. (1994). The Persuasion Knowledge Model: How People Cope with Persuasion Attempts. *Journal of Consumer Research*, *21*(1), 1–31. https://doi.org/10.1086/209380

Giasson, T., Le Bars, G., & Dubois, P. (2019). Is Social Media Transforming Canadian Electioneering? Hybridity and Online Partisan Strategies in the 2012 Quebec Election. *Canadian Journal of Political Science*, *52*(2), 323–341. https://doi.org/10.1017/S0008423918000902

Holt, K., Shehata, A., Strömbäck, J., & Ljungberg, E. (2013). Age and the effects of news media attention and social media use on political interest and participation: Do social

media function as leveller? *European Journal of Communication*, *28*(1), 19–34. https://doi.org/10.1177/0267323112465369

Hooper, D., Couglan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Journal of Business Research Methods*, *6*(1), 53–60.

Instagram. (2020, March 16). *Helping to Inform People about COVID-19 Vaccines | Instagram Blog*. https://about.instagram.com/blog/announcements/continuing-to-keep-people-safe-and-informed-about-covid-19

Jamieson, K. H. (2013). Messages, micro-targeting, and new media technologies. *Forum (Germany)*, *11*(3), 429–435. https://doi.org/10.1515/for-2013-0052

Jansen, M.-P., & Krämer, N. C. (2023). Empty Transparency? The Effects on Credibility and Trustworthiness of Targeting Disclosure Labels for Micro-Targeted Political Advertisements. *Weizenbaum Journal of the Digital Society*, *3*(1). https://doi.org/10.34669/WI.WJDS/3.1.5

Janssen, L., Fennis, B. M., & Pruyn, A. Th. H. (2010). Forewarned is forearmed: Conserving self-control strength to resist social influence. *Journal of Experimental Social Psychology*, *46*(6), 911–921. https://doi.org/10.1016/j.jesp.2010.06.008

Janssen, L., Fransen, M. L., Wulff, R., & Van Reijmersdal, E. A. (2016). Brand placement disclosure effects on persuasion: The moderating role of consumer self-control: Self-control and brand placement disclosure. *Journal of Consumer Behaviour*, *15*(6), 503–515. https://doi.org/10.1002/cb.1587

Janssen, L., Schouten, A. P., & Croes, E. A. J. (2021). Influencer advertising on Instagram: Product-influencer fit and number of followers affect advertising outcomes and influencer evaluations via credibility and identification. *International Journal of Advertising*, 1–27. https://doi.org/10.1080/02650487.2021.1994205

Jost, P., Kruschinski, S., Sülflow, M., Haßler, J., & Maurer, M. (2022). Invisible transparency: How different types of ad disclaimers on Facebook affect whether and how digital political advertising is perceived. *Policy & Internet*, poi3.333. https://doi.org/10.1002/poi3.333

King, A. J., Bol, N., Cummins, R. G., & John, K. K. (2019). Improving Visual Behavior Research in Communication Science: An Overview, Review, and Reporting Recommendations for Using Eye-Tracking Methods. *Communication Methods and Measures*, *13*(3), 149–177. https://doi.org/10.1080/19312458.2018.1558194

King, J. M. (2022). Microtargeted Political Ads: An Intractable Problem. *Boston University Law Review*, *102*, 39.

Kruikemeier, S., Sezgin, M., & Boerman, S. C. (2016). Political Microtargeting: Relationship between Personalized Advertising on Facebook and Voters' Responses. *Cyberpsychology, Behavior, and Social Networking*, *19*(6), 367–372. https://doi.org/10.1089/cyber.2015.0652

Kruikemeier, S., Vermeer, S., Metoui, N., Dobber, T., & Zarouali, B. (2022). (Tar)getting you: The use of online political targeted messages on Facebook. *Big Data & Society*, *9*(2), 20539517221089626. https://doi.org/10.1177/20539517221089626

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Online Advertising and Social Media (Transparency) Bill, 150–2017 (2021). https://data.oireachtas.ie/ie/oireachtas/bill/2017/150/eng/initiated/b15017d.pdf

Lubianco, J. (2022, April 12). *Fake News bill gets stuck in Brazilian Congress and it's unlikely to be voted on before the elections; remuneration proposal for journalistic*

organizations is a sensitive topic. LatAm Journalism Review by the Knight Center. https://latamjournalismreview.org/articles/fake-news-brazil-payment-journalism/

Matthes, J., Hirsch, M., Stubenvoll, M., Binder, A., Kruikemeier, S., Lecheler, S., & Otto, L. (2022). Understanding the democratic role of perceived online political micro-targeting: Longitudinal effects on trust in democracy and political interest. *Journal of Information Technology & Politics*, 1–14. https://doi.org/10.1080/19331681.2021.2016542

Murray, G. R., & Scime, A. (2010). Microtargeting and electorate segmentation: Data mining the American National Election Studies. *Journal of Political Marketing*, *9*(3), 143–166. https://doi.org/10.1080/15377857.2010.497732

Pan, Y., & Zinkhan, G. M. (2006). Determinants of retail patronage: A meta-analytical perspective. *Journal of Retailing*, *82*(3), 229–243. https://doi.org/10.1016/j.jretai.2005.11.008

Pornprasertmanit, S., Miller, P., Schoemann, A., Jorgensen, T. D., & Quick, C. (2021). *simsem: SIMulated Structural Equation Modeling* (0.5-16) [Computer software]. https://CRAN.R-project.org/package=simsem

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rosseel, Y. (2012). **lavaan**: An *R* Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Statista. (2023). *Most important issues facing Netherlands 2022* [Chart]. Statista. https://www.statista.com/statistics/547685/public-opinion-of-important-issues-facing-the-netherlands/

Tene, O. (2011). Privacy: The new generations. *International Data Privacy Law*, *1*(1), 15–27. https://doi.org/10.1093/idpl/ipq003

The jamovi project. (2022). *Jamovi* (2.3) [Computer software]. http://www.jamovi.org

van Reijmersdal, E. A., Fransen, M. L., Van Noort, G., Opree, S. J., Vandeberg, L., Reusch, S., Van Lieshout, F., & Boerman, S. C. (2016). Effects of Disclosing Sponsored Content in Blogs: How the Use of Resistance Strategies Mediates Effects on Persuasion. *American Behavioral Scientist*, *60*(12), 1458–1474. https://doi.org/10.1177/0002764216660141

Wojdynski, B. W., & Evans, N. J. (2016). Going Native: Effects of Disclosure Position and Language on the Recognition and Evaluation of Onlin. *Journal of Advertising*, *45*(2), 157–168. https://doi.org/10.1080/00913367.2015.1115380

Zarouali, B., Dobber, T., De Pauw, G., & de Vreese, C. (2020). Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media. *Communication Research*. https://doi.org/10.1177/0093650220961965

Zarouali, B., Poels, K., Ponnet, K., & Walrave, M. (2018). "Everything under control?": Privacy control salience influences both critical processing and perceived persuasiveness of targeted advertising among adolescents. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *12*(1). https://doi.org/10.5817/CP2018-1-5

Zarouali, B., Ponnet, K., Walrave, M., & Poels, K. (2017). "Do you like cookies?" Adolescents' skeptical processing of retargeted Facebook-ads and the moderating role of privacy concern and a textual debriefing. *Computers in Human Behavior*, *69*, 157–165. https://doi.org/10.1016/j.chb.2016.11.050

Zuiderveen Borgesius, F. J., Möller, J., Kruikemeier, S., Fathaigh, R., Irion, K., Dobber, T., Bodo, B., & de Vreese, C. (2018). Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, *14*(1), 82–96. https://doi.org/10.18352/ulr.420

DOI: 10.17185/duepublico/81516
URN: urn:nbn:de:hbz:465-20240216-142518-5