Online Trolls: Unaffectionate Psychopaths or Just Lonely Outcasts and Angry Partisans?

Verbalyte, Monika; Keitel, Christoph; Howard, Christa

Article

Online Trolls: Unaffectionate Psychopaths or Just Lonely Outcasts and Angry Partisans?

Monika Verbalyte [1],*, Christoph Keitel [1], and Krista Howard [2]

[1] Interdisciplinary Center for European Studies, Germany
[2] Department of Psychology, Texas State University, USA

* Corresponding author (monika.verbalyte@uni-flensburg.de)

Abstract
The main objective of the article is to attempt to provide a more sociological explanation of why some people attack and insult others online, i.e., considering not only their personality structure but also social and situational factors. The main theoretical dichotomy we built on is between powerful high-status and low-on-empathy "bullies" trolling others for their own entertainment, and people who are socially isolated, disempowered, or politically involved, therefore feel attacked by others' beliefs and opinions expressed online, and troll defensively or reactively instead of primarily maliciously. With an MTurk sample of over 1,000 adult respondents from the US, we tested these assumptions. We could confirm that there are two categories and motivations for trolling: for fun and more defensive/reactive. Further, we checked how strongly precarious working conditions, low social status, social isolation, and political as well as religious affiliation of the person increase or decrease the probability of trolling as well as enjoyment levels from this activity. We controlled for personality traits, social media use and patterns, as well as sociodemographic factors. We could confirm that political identities and religiosity increase the likelihood of, but not the enjoyment of trolling; however, socio-economic factors do not have the same differentiating effect.

Issue
This article is part of the issue "Negative Politics: Leader Personality, Negative Campaigning, and the Oppositional Dynamics of Contemporary Politics" edited by Alessandro Nai (University of Amsterdam), Diego Garzia (University of Lausanne), Loes Aaldering (Free University Amsterdam), Frederico Ferreira da Silva (University of Lausanne), and Katjana Gattermann (University of Amsterdam).

1. Introduction

Social media is often seen as a polarizing, if not toxic, environment. In part, this heated online climate might be attributed to constant conflicts with so-called internet "trolls." It is still not fully clear why some people find it appropriate and even fun to attack and insult others online. The most widespread explanation, offered mainly by psychological studies, is that those who engage in such behavior are aggressive individuals with low impulse control, psychotic personality traits (Buckels et al., 2014; Lopes & Yu, 2017), and a lack of empathy

(Sest & March, 2017). However, this research rarely considers how the social environment and living conditions of trolls—such as stress level, precarious working conditions, social isolation, economic deprivation, and other hardships—may influence their engagement in different kinds of social and online deviance. Also, scholars have not yet examined how trolling behavior is influenced by political identities, i.e., those known to strongly increase affective polarization by driving people to react aggressively to their opponents, especially within the anonymous internet space. On this subject, one qualitative interview study recently found that the most aggressive

people online are not only extreme in their political attitudes but also disadvantaged in their social lives, such that they seek needed attention, empowerment, and confidence from hostile online activities (Bail, 2021).

The goal of this study is to theoretically contrast and comprehensively analyze these different modes of trolling, focusing on the personality traits of individuals who engage in such behavior and the socio-political factors that may impact their decision to do so. We see this phenomenon in the broader context of current developments: rising affective political and social polarization and negative partisanship in both the US and Europe (Reiljan, 2020), the increasing negativity of electoral campaigns (Nyhuis et al., 2020), and the general incivility of political discourse (Gidron et al., 2019). Very often, this augmented hostility is related to populist and extremist (especially far-right) parties, whose candidates are more prone to use attack politics and harsh campaigns (Nai, 2021), and more often loath and are loathed by the partisans of other parties (Harteveld et al., 2021).

Against this background, a puzzle arises. While dark personality traits allegedly underlie trolling behavior, the phenomenon also seems to represent a reactive response to felt powerlessness. Former US President Donald Trump can serve as an illustrative example here. Trump can be described as a real "bully": narcissistic, aggressive, power-obsessed (Nai & Maier, 2018), and seemingly liked by individuals with similar personality profiles (Nai et al., 2021). It is logical to expect that people like Trump and his supporters are also active and spread hostility online. Yet so-called "rednecks" and other populist supporters are also often seen as social "losers": individuals whose economic situation objectively worsened in the last decades (Gidron & Hall, 2017), who feel misunderstood, unjustly treated, disempowered, and excluded due to their beliefs (Abts & Baute, 2021; Hochschild, 2016), and who fuel *ressentiment* and reactionarism (Salmela & Capelos, 2021). This negative emotionality is channeled not only through populist support, but also vented on social media, for example by trolling out-party supporters (Bail, 2021)—thereby further fueling affective polarization (Bulut & Yörük, 2017).

Considering these factors, this article seeks to take a more comprehensive look at trolling behavior. First, we clarify the definition of trolling and differentiate between trolling for fun and more defensive/reactive trolling. Second, we review the possible explanations of online trolling behavior, including the more sociological and political ones. We continue by performing explanatory and confirmatory factor analyses (EFA and CFA, respectively) to test the assumed differences between concepts of and motivating factors behind trolling, as well as structural equation path modeling to show their relationship. We then conduct regression analyses in order to introduce a wider range of explanatory and control variables into the equation. Finally, we conclude by presenting our results and reflecting on the limitations of our study.

2. Definition of Trolling

The term "trolling" has become a catchphrase to describe a range of online deviant behaviors. It is crucial, therefore, that we differentiate between trolling and other forms of anti-social, or dark, online activities, the most similar of which are hate speech, online harassment, and cyberbullying (Hardaker, 2010; Shachaf & Hara, 2010). All four must be clearly distinguished from impoliteness, online incivility, offensive, derogatory, and abusive language, or so-called "flaming." The latter depict forms of discursive hostility (Andersen, 2021), or hostile ways of presenting content—not the behavior itself. Accordingly, uncivil language could be used in connection with all forms of online hostility, albeit for different purposes and with different intentions. What differentiates trolling from other forms of hostile online behavior is, first and foremost, the target. There is a pre-existing relationship between the cyberbully and the victim(s), so damage is caused to the specific target (Fichman & Sanfilippo, 2016). Harassment is related to real-world stalking and is often sexual in nature (May et al., 2012), and hate speech is a form of verbal violence directed at particular groups, especially on the basis of race, religion, or sexual orientation (Paasch-Colberg et al., 2021). By contrast, in the case of trolling the target does not need to be predefined, and the network of trolling is usually wide, not limited as it is with cyberbullying (Hardaker, 2010).

All in all, while trolling has become a widely known internet phenomenon, its definition remains blurry due to the variety of practices it encompasses. Trolling activities range from mocking others for self-amusement to disrupting online communities (Pao, 2015). In general, trolling can be defined as the act of posting/sharing content or comments that may irritate or cause conflict among those who receive it, and/or starting a circular discussion that deliberately provokes an aggressive response by the targets of the trolling (Hardaker, 2010). Fichman and Sanfilippo (2015) regard online trolling as a deviant and antisocial behavior in which the user acts provocatively against normative expectations. However, when we look at this phenomenon with the aim to understand the desired outcomes and intentions of trolling, the picture becomes rather vague. From this perspective, we encounter enjoyment and thrill-seeking as well as revenge and self-gratification (Cook et al., 2018; Fichman & Sanfilippo, 2015; Quandt et al., 2022; Shachaf & Hara, 2010). This article seeks to manage the complexity of the trolling dynamic by categorizing trolling into two main types: (a) trolling for fun and (b) defensive/reactive trolling.

Trolling as a form of entertainment has been widely documented (Buckels et al., 2014; Shachaf & Hara, 2010). Hardaker (2010) explains that trolls aim to cause disruption and trigger unpleasant feelings in their victims for the sake of their own amusement. The trolling behavior is therefore associated with positive emotions, such

as the pleasure that comes from attracting attention. Yet the internal trigger of boredom also appears to play an important role in this context (Shachaf & Hara, 2010).

In contrast, defensive/reactive trolling is related to negative emotions and represents a reaction to another person's actions. In this vein, Cheng et al. (2017) found that exposure to previous trolling behavior by others can increase the users' likelihood of engaging in trolling behavior themselves. Furthermore, witnessing how another group member is subjected to trolling may motivate a third party to step in and troll back (Hopkinson, 2013). Another possibility is the so-called "white and gray hat trolls"—or people who not only troll back, but also "troll the trolls" for the sake of an alleged higher cause, i.e., to direct public attention to certain issues (Matthews & Goerzen, 2019). In this sense, trolling might be motivated by genuine beliefs and in-group identities, and not only by feelings of boredom or the expectation of fun. People who engage in this kind of trolling might feel threatened, attacked (Liu et al., 2018), and vengeful (Shachaf & Hara, 2010). They may respond spontaneously and impulsively, without the intent to harm (Hardaker, 2010), and later regret their actions. To sum up, this trolling is defensive/reactive in many respects: It is a neurotic or moody reaction to the provocation of another, as well as a more considerate response strategy based on strong beliefs and calculations.

3. Explanations of Trolling Behavior

The different types of trolling already provide some hints as to the underlying reasons behind such behavior. This section reviews the known explanatory factors for trolling, focusing on those related to defensive/reactive trolling. To gain access to a broader pool of potential explanations, we consulted a wide range of studies on online aggressiveness, deviance, and criminality in general. We start by reviewing what Cook et al. (2018) classified as the internal triggers (personality traits) and circumstantial factors (mainly, the social media environment) of trolling behavior. While these are the most commonly considered factors with regard to trolling, they are rather narrow in their explanatory power. After that, we present what we believe to be the most important social and political determinants of defensive/reactive trolling.

3.1. Personality Structure

Among scholars, personality structure appears to be the most popular and attractive explanation of trolling behavior. The dark personality types—Machiavellianism, psychopathy, sadism, and narcissism—are distinguished by remorseless, impulsive, thrill-seeking, and socially offensive behavior (Buckels et al., 2019; Jones & Paulhus, 2011), including trolling. Psychopathy is related to low self-control and low empathy, as well as a lack of respect for social norms and conventions (Foster & Trimm, 2008).

The absence of inhibitory mechanisms permits people with this trait to engage in a wide range of antisocial activities (Craker & March, 2016; Hare, 2006; Jones et al., 2011). Sadistic psychopaths also believe that it is totally acceptable to manipulate and use others, for example by blackmailing and bullying popular individuals (Buckels et al., 2014; Lopes & Yu, 2017). Like psychopaths, narcissists and Machavellianists have a distorted view of their own self-importance and do not hesitate to use manipulation and deception for self-enhancement and to protect their self-esteem, although their behavior is less motivated by pleasure-seeking or harming others and is more self-serving than that of psychopaths (Campbell & Miller, 2011; Jakobwitz & Egan, 2006). Our data set, however, only enables us to measure empathy, which has already been shown to negatively relate to both psychopathy and trolling (Sest & March, 2017).

Our analysis does include the Big5 inventory, of which low levels of conscientiousness and agreeableness are associated with psychopathy and a higher likelihood of enjoying trolling. High levels of extraversion, neuroticism, or openness also have a positive impact on trolling (Gylfason et al., 2021); however, extraversion and openness might well indicate an outgoing and self-confident personality and an interest in communicating in various different ways, rather than malignant intentions. Moreover, neuroticism, low levels of agreeableness, and low levels of conscientiousness are correlated with a specific kind of social media use—that is, the kind that seeks to gain from social media the attention and social support people might lack in their offline lives (Seidman, 2013), so also possibly defensive trolling. Another known personality-related explanation of online deviance has to do with the difficulties some people experience in regulating their own emotions, and the ramifications of their anger (Toro et al., 2020). This negative affect is known to spiral, completely consume the person, and reduce the person's inhibitions, prompting an urge to engage in corrective action towards the culprit (the person perceived to have caused the anger; Agnew, 1995). In this sense, it could also be related to the troll-back reaction.

Thus, we expect empathy to have a strong impact in that it can decrease a person's enjoyment of trolling for fun (but not defensive trolling). However, the contradictory reasoning and differing results of the existing studies on the Big5 hinder us from formulating any clear expectations on this relation. Because our survey provides no measure for emotional regulation as a personality trait, we count the motivations to troll related to anger and other negative emotions as situational, and therefore as reactive trolling.

3.2. Social Media Environment

With the recent development of new technologies, trolling opportunities have become more available. Many online platforms, especially those supported by social media, provide relatively open spaces for a large

number of individuals to engage in genuine, sincere inter-personal communication and debate. Trolls actively prey on such individuals by posting controversial and inappropriate messages to derail discussions into pointless tangents (Herring et al., 2002). Little to no moderation and the option to remain anonymous means that violations of conversational norms often go unpunished, or at least give that impression—which, in turn, invites deception, controlled self-presentation, and decreased self-control among some users (Hardaker, 2010). The ability to remain incognito or to disguise their identities gives trolling individuals an advantage over their victims, who often openly share personal information about themselves, thereby inadvertently inviting ad hominem attacks. Another known effect of anonymity is its deindividuating effect on users: Decreasing the salience of individual identity and increasing that of social/group identity deflects personal responsibility and enables deviant behavior (see social identity model of deindividuation effects by Spears, 2017, and our notes on group-driven trolling in the last passage of Section 2).

We also follow the logic of situational action theory, which addresses both the personal and environmental factors that tend to predominate in the offending individual (Li et al., 2022). Since the impact of the online environment is constant for everyone, what creates the difference is how individuals approach and use that environment. In this regard, time appears to be one key factor: Scholarship tends to show a strong association between time spent online and deviant behavior online (Lee, 2018). What about different kinds of trolling—Do they seem to align with different usage patterns regarding the online environment? On the one hand, more people show dark personality traits online than offline (Nevin, 2015) and hostile political discussions are mainly caused by status-driven individuals (Bor & Petersen, 2021). On the other hand, online attacks are frequently driven by people who feel their lack of status and want to boost it (Bail, 2021). In this respect, it seems that online environments do not feed into any specific type or types of trolling; they appear to promote trolling of all kinds. Also, we know that interactions between ideologically opposed users are significantly more negative than like-minded ones (Marchal, 2021), which means that the composition of online networks matters. Once again, this seems to present more opportunities for trolling in general—of all kinds.

How people shape social comparison on social media, however, is differently related to different sorts of trolling. Research suggests that mostly psychopathic people engage in upward social comparison, or comparison with people who are better off (Lopes & Yu, 2017); others who want to enhance their social status make a downward comparison, since upward comparison for them would only increase their anxiety, low self-esteem, insecurity, and loneliness (Bonnette et al., 2019; Howard et al., 2019). We would expect defensive trolling to be more related to downward online comparison,

and enjoyment trolling to be more related to upward online comparison.

3.3. Social Status and Identity

Independently of a user's propensity to anger ramifications, experiencing online hostility from others increases the likelihood that one will respond in a trolling manner (Liu et al., 2018). This could be triggered not only by vengefulness or a lost temper but also by the feeling that social norms allow for this kind of behavior (Sentse et al., 2007)—a feeling encouraged by the relative freedom and anonymity of online space, as we noted in Subsection 3.2. The inhibition of social controls is even stronger in group settings: Research suggests that exposure to trolling does not have to be experienced directly by individuals but may gear itself more towards group identities (see also social identity model of deindividuation effects theory in Subsection 3.2. Trolling out-group and in-group members who violated some inner rules also helps to establish community boundaries (Rafalow, 2015).

In that sense, partisanship represents a form of defensive/reactive trolling, even if it sometimes goes beyond mere revenge and represents some form of ideological impetus (Fichman & Sanfilippo, 2016). Driven by certain political beliefs or ideologies, trolls seek to draw attention to the social issues they care about (Sanfilippo et al., 2017), or they use social media as communication and mobilization channels through which to find allies and/or to legitimate and inspire a group action (Flores-Saviaga et al., 2018). Political trolling practices can also include baiting ideological opponents into arguments through coordinated behavior and conscious attack, in order to spam adversary online platforms and individuals (Frischlich et al., 2021; Sanfilippo et al., 2017). However, in this case, the action is done in response to the troll's perception of a group threat, rather than for personal fun. A somewhat different function could be ascribed to another identification: religion. Religious convictions are often politicized and important in defensive/reactive trolling; yet they are also sometimes presented as moral convictions that deter the holder of those convictions from deviance (Adamczyk, 2012). In that case, the person might feel shame or guilt when trolling, rather than enjoyment.

General socio-economic factors also belong to the important determinants of social deviance. Most of the studies relate to poverty, low levels of education, and destructive family relations to anti-social behavior (Cioban et al., 2021; Hagan & Foster, 2001). Uncivil online behavior has been even more strongly linked to a generally low or reduced social status (for example, due to a recent divorce or unemployment) and subjective feelings of social isolation (such as lack of social support) and economic deprivation. The loneliness of people with these characteristics leads them to attach greater importance to other social identities, such as political

partisanship (cf. Salmela & von Scheve, 2018) and to vent online more often than other individuals (Bail, 2021), which in turn ostensibly enhances their power and social status (Juvonen & Graham, 2014). The feeling of injustice might be more relevant than the objective socio-economic situation; thus, a person could have a fairly solid social standing and still lack the experience of social mobility (cf. Hochschild, 2016; see also Salmela & von Scheve, 2018) or feel deprived of what they legitimately deserve (Agnew, 1995; Cioban et al., 2021). This frustration causes anger and aggression as well as anxiety, depression, and general irritability (Aseltine et al., 2000), which leads us to believe that it also belongs to a group of factors causing defensive trolling.

To summarize, we expect to find more defensive/reactive trolling among people with strong political identities than among other individuals. However, we cannot formulate a clear hypothesis for religious identity, which might also inhibit trolling. Second, we suggest that people with low socio-economic status, precarious working conditions, and relatively low levels of social support are more prone to troll than others. On the other hand, an elevated social status appears more related to trolling for fun, since having a successful career and a position of power might also signal the presence of a dark personality (i.e., given the deep disregard for others and strong will to dominate that tend to characterize individuals with dark personality traits; Festl & Quandt, 2013; see also Section 3.1).

4. Research Data and Design

To test our theoretical model, we used survey data of 1,314 individuals gathered by Texas State University on MTurk (Amazon Mechanical Turk). The sample consists of 51.26% men and 48.29% women (the remaining respondents reported as transgender or gender non-conforming), 68.57% of whom self-identified as white, 40.79% as Democrats, and 26.64% as Republicans. The mean age of respondents was 35.74 years. In comparison to US census data, the sample overrepresents young people and Democrats; however, this is fairly common in online samples. While there are also some deviations for race, this might be based on the different race categories used and the fact that not all MTurk workers come from the US. Although scholars often express reservations about the quality of MTurk surveys, they still seem to be preferred over purely student samples (Pew Research Center, 2016).

Since the original purpose of the survey differs from the aim of our study, it does not cover all the variables that follow from our theoretical considerations, yet still provides a wide range of variables that could help to answer some of our research questions. In this sense, our study could be seen as a secondary data analysis. For this reason, we also present our study as rather explorative, even though it is generally led by theoretical considerations and expectations.

In analyzing these variables, we first present the available items related to trolling and the motivation to troll. We then draw on both EFA and CFA to group these into meaningful variables. Later, we present the expected links between those variables and test them with structural equation path modeling. To include more variables than would be feasible with the path model, we continue with regression analyses. Analyses were conducted with Stata15 and R package lavaan.

4.1. Trolling and Trolling Motivations

For our dependent variables, we decided to go with the items that would represent the greatest breadth of possible trolling and attacking behaviors and hypothesized that they could be clustered into two categories. In the first category, we grouped the items concerned with the extent to which one enjoys debating to upset or irritate others and to troll. These items were expected to represent the enjoyment of trolling. In the second category, we grouped together items related to the respondent's likelihood to make comments, create posts, send people "shock websites" to upset others, and whether the respondent enjoys posting with the intent to annoy others. We used the items in this second category to measure the respondent's likelihood to troll, since they do not specifically mention enjoyment and therefore probably also represent defensive/reactive trolling (for a more detailed description of variables, see the Supplementary Material).

This structure of variables could already be confirmed by means of a correlation matrix (Figure A1 in the Supplementary Material). Since all of the items represent anti-social and undesirable behaviors, they did not show a normal distribution and had to be dichotomized for further analyses (similar to Howard et al., 2019), i.e., all respondents who did say more than "extremely unlikely" or "strongly disagree" (for descriptives, see Figure A2 in the Supplementary Material). While this is a very strict exclusion criterion for the circle of trolls, it was done in order to capture measurements for accidental trolling behavior. EFA renders the model with two factors (for Table 1, Promax rotation was used because the respondent's likelihood to troll also partially includes trolling for fun, so these two sorts of trolling cannot be treated as independent from one another). CFA confirms this structure only with acceptable fit (Table A1 with loadings in the Supplementary Material), but we decided not to further modify our model due to the strong theoretical assumptions and other empirical evidence for this model. We created two continuous variables from the corresponding EFA factor scores, which ranged from 0 to 1. The pressing question, however, is how keen people are to admit that they are trolling, and how reasonable our partition of trolling in these two groups is. We present some validity checks to support our claims and data (see validity checks in the Supplementary Material), which confirm that both our numbers of trolls

Table 1. Factor loadings of trolling items.

	Likely to troll	Enjoy trolling
Comment to upset	**0.8007**	0.1280
Offensive posts	**0.9411**	0.0059
Send shocking websites	**0.8669**	0.0240
Post to annoy	**0.4228**	0.2661
Enjoy trolling	0.2742	**0.5294**
Enjoy debating to upset	−0.0064	**0.8909**

Notes: With Promax rotation, normalized.

as well as our inquiry into different trolling motivations are justified.

To merge deeper into these possible motivations, we look at the survey questions that address why respondents troll or how they felt while doing it. The questions address different emotional, mental, and social states of the people: "When you are commenting or posting in order to upset or irritate others, to what extent are you… Highly Stressed, Tired/Fatigued, Bored, Depressed, Anxious, Feeling Attacked, Lonely, Silly, Annoyed" and "When commenting on others' posts with the intent to upset or irritate others, to what extent do you feel… Powerful, Courageous, Intelligent, Levelheaded, Happy, Embarrassed, Devious, Superior, Cruel, Protected, Guilty, Confident, Fearful, Defensive." All of the answers are measured on a scale of 0 to 100. Since named motivations had a strong relation to trolling behavior in terms of distribution, we also dichotomized these items. Thus, everything higher than 10 was considered to include this motivation.

We used these items to build patterns of trolling which Maltby et al. (2016) called implicit trolling theories. First, we proceeded with EFA. The exclusion of the items Devious, Defensive, and Fearful, has rendered a good five-factor solution (Table 2). Already in the correlation matrix (Figure A3 in the Supplementary Material), excluded variables showed low correlations with other items or did not constantly load on any specific factor. The CFA also provided a good fit for this five-factor solution (Table A3 in the Supplementary Material), with the factors being Powerful and Happy, Silly and Bored, Attacked and Annoyed, Stressed and Anxious, and Embarrassed and Guilty. Here we also created continuous variables from corresponding EFA factor scores, which ranged from 0 to 1.

The literature we reviewed mentioned all of the motivation groups as possible motives for trolling. Here, we assumed that Powerful and Happy and Silly and Bored would be more related to trolling for fun, whereas Attacked and Annoyed (due to the defensive,

Table 2. Factor loadings of trolling motivations.

	Stressed and Anxious	Attacked and Annoyed	Powerful and Happy	Embarrassed and Guilty	Silly and Bored
Highly Stressed	**0.6849**	0.2568	−0.0167	0.0897	−0.1395
Tired/Fatigued	**0.6835**	0.0910	0.0831	−0.0008	0.0395
Depressed	**0.7731**	−0.0309	0.0385	0.0266	0.1032
Lonely	**0.6415**	−0.1190	0.1143	−0.0192	0.2980
Anxious	**0.8053**	0.1107	−0.0038	0.0257	−0.0383
Feeling Attacked	0.1606	**0.6424**	0.0424	0.1236	−0.0248
Annoyed	0.0906	**0.7981**	−0.0070	−0.0192	0.1166
Happy	−0.0298	−0.0458	**0.9403**	−0.0014	0.0078
Powerful	0.0475	−0.0155	**0.8744**	0.0383	−0.0000
Superior	0.0290	0.0447	**0.8112**	0.0396	0.0557
Confident	−0.0612	0.1128	**0.9076**	−0.0813	0.0164
Courageous	0.1219	−0.0596	**0.8657**	0.0425	−0.0628
Intelligent	−0.0322	0.0539	**0.9329**	−0.0452	−0.0015
Levelheaded	−0.0200	0.0331	**0.9054**	−0.0301	−0.0358
Protected	0.0823	−0.0995	**0.7474**	0.0794	0.0994
Embarrassed	0.1253	−0.0580	0.1400	**0.7165**	−0.0854
Guilty	0.0399	0.0048	−0.0914	**0.8997**	0.0180
Cruel	−0.0989	0.0830	−0.0105	**0.8575**	0.0739
Bored	0.0761	0.1061	0.0899	−0.0106	**0.7034**
Silly	0.2033	0.0039	0.1630	0.0416	**0.4881**

Notes: With Promax rotation, normalized.

possibly spontaneous reaction), Stressed and Anxious (in order to get attention and overcome one's insecurities), Embarrassed and Guilty (due to the inhibiting impact of social norms after the spontaneous outburst) would be indicators of defensive/reactive trolling. Our hypothesized relations do not mean that there could be no cross-loadings. For example, it is possible to feel some *Schadenfreude* (pleasure at another person's misfortune) even while trolling defensively; but since this is not the primary motivation for this behavior, the relation should be much less incisive. These cross-loadings we first and foremost expect between trolling for fun motivations and likelihood to troll, since defensive/reactive trolling could not be clearly extracted from variables on trolling behavior. On the other hand, enjoyment of trolling should clearly only be related to fun motivations (for the main hypothesized model, see Figure 1).

4.2. Other Independent Variables

For further analyses, we will conduct multiple linear regressions since further buildup of the path model would make it too confusing. Also, it would be hard to argue for some paths as mediations or causal links since we only have cross-sectional data.

We start with social status variables (more detailed description in the Supplementary material). Occupation includes jobs of different skill and status levels. Precarious Working Situation differentiates between people having one stable job, and those forced to juggle a few part-time positions. The Interpersonal Support Index

is a proxy of available support networks, the opposite of social isolation. Another group of main explanatory variables is identity. For Political and Religious Affiliation, we included both the strength and ideological direction of affiliation. To control for political involvement, we took people's self-assessed online political engagement or Percentage of Political Posts.

The rest of the variables represent controls. For aspects related to personality, we used the Big5 inventory and Empathy. For social media environment, we used the measure of Heavy Social Media Use and two additional variables of how people use them and the composition of their social networks. The former is measured through Upward and Downward Online Comparison, the latter with the Percentage of (Dis)Similar Groups in which they are involved. Finally, we also added a few important socio-demographic variables: Gender, Age, Race, and LGBTQAI+.

5. Results

First, we test the main measurement and path model (see Figure 2). It shows a good fit between our hypothesized model and our data (CFI = 0.951; TLI = 0.9445; RMSEA = 0.064). The measurement part confirms our previously presented results: Factor loadings are high for all of the relations discovered through EFA. The new part is the hierarchical model of motivations where five latent groups are further located under two groups of motivations, and high loadings confirm that the categories Embarrassed and Guilty, Attacked and Annoyed,
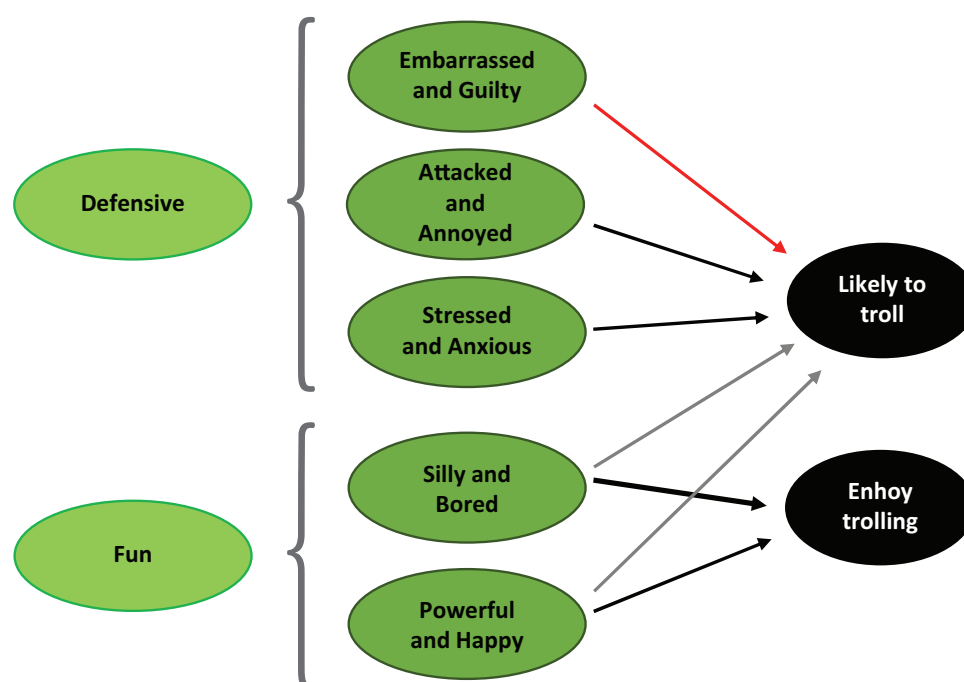

Figure 1. Tested measurement and path model. Notes: All the presented variables are latent variables; part of the measurement model with manifest variables is excluded in order to not overload the figure; different figure colors represent different variable groups; black arrows represent hypothesized positive relations, red ones hypothesized negative relations, and gray possible positive relations.
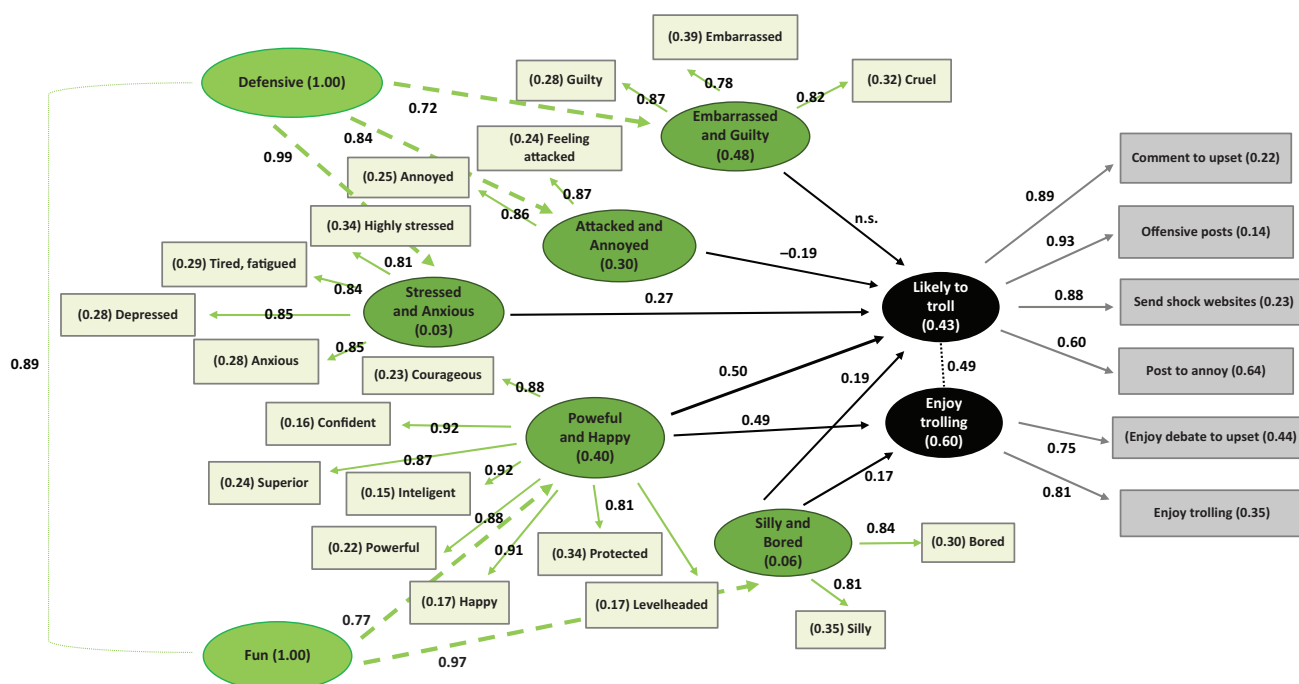
Figure 2. Main measurement and path model. Notes: Near variable names variances are depicted in parentheses; black and gray colors represent trolling; green colors represent motivations to troll; gray arrows are loadings of trolling behaviors on latent trolling variables; green arrows are loadings of specific motivations on latent motivation variables; green dashed lines are loadings of latent motivation variables on two main groups of motivations; black straight lines are regression coefficients; dotted lines (green for motivations, black for trolling behavior) are correlations.

and Stressed and Anxious belong to defensive motivations, whereas Silly and Bored and Powerful and Happy correlate to trolling for fun.

As for our path model, we see that Powerful and Happy as well as Silly and Bored have a significant positive relation with one's enjoyment of trolling and, as expected, one's likelihood to troll. Defensive trolling motivations demonstrate different relations to trolling. Thus, Stressed and Anxious has a positive influence on one's likelihood to troll, whereas—contrary to expectations—Embarrassed and Guilty has no significant relation to trolling behavior and Attacked and Annoyed has a negative coefficient for trolling. As for Embarrassed and Guilty, a nonsignificant relation was as plausible as a negative one. Indeed, in a sense, this feeling does not decrease or increase trolling: It is only felt during or after the act of trolling. As for Attacked and Annoyed, a negative coefficient sign emerges only after controlling for Powerful and Happy and Stressed and Anxious, whereas the binary correlation with trolling is positive (see Figure 5A in Supplementary Material). Thus, it seems that if the person is stressed, or feels self-confident and superior, attacking them does not serve as an additional driver for trolling; rather, it decreases the probability of a trolling incident. Possibly, this is due to a negative psychological effect—i.e., the attack deepens the person's anxiety or partially diminishes their courage to troll.

To control for a wider range of variables and possible cross-loadings between motivations and trolling, we now turn to regressions (see Table 3). Models 1a and 1b just confirm path model results. Models 2a and 2b demonstrate the influence of the main explanatory variables. From our social status variables, interpersonal support demonstrates the strongest impact. Yet we do find strong negative associations between this factor and enjoyment of trolling and likelihood to troll, although we have hypothesized that its impact is limited to general trolling. The explanation for this might be that trolling is not only more likely, but also more enjoyable if a person is bored and alone due to the excitement delivered by the act itself, and the satisfaction of not having to deal with one's lack of meaningful social connections. We also see a strong and positive correlation between the enjoyment of trolling and some occupations—not necessarily higher-status ones. This might simply indicate that, in comparison to unemployment, any job will increase one's feelings of power and therefore one's enjoyment of trolling. The likelihood to troll does not significantly correlate to any single occupation, but having a precarious position is a significant predictor of trolling behavior, even though it reduces the probability of trolling. When we also compare this result to the negative binary link between unemployment and trolling (Figure A4 in the Supplementary Material), we may conclude that having the lowest levels of social status (precariously employed or unemployed) is so disempowering that the person does not even try to gain a feeling of empowerment through trolling.

Table 3. Regression analyses on one's enjoyment of trolling and likelihood of troll.

	(1a) Enjoy trolling	(2a) Enjoy trolling	(3c) Enjoy trolling	(1b) Likely to troll	(2b) Likely to troll	(3b) Likely to troll
Feeling while trolling: Powerful and Happy	0.489*** (0.039)	0.392*** (0.042)	0.331*** (0.042)	0.515*** (0.037)	0.370*** (0.040)	0.302*** (0.037)
Feeling while trolling: Silly and Bored	0.110* (0.065)	0.060 (0.067)	0.017 (0.066)	0.161*** (0.057)	0.092* (0.059)	0.052 (0.057)
Feeling while trolling: Attacked and Annoyed	0.025 (0.055)	0.071 (0.056)	0.100* (0.055)	−0.145*** (0.043)	−0.089** (0.042)	−0.042 (0.041)
Feeling while trolling: Stressed and Anxious	−0.009 (0.068)	−0.091 (0.071)	−0.110 (0.068)	0.205*** (0.057)	0.116* (0.058)	0.062 (0.054)
Feeling while trolling: Embarrassed and Guilty	−0.011 (0.049)	−0.034 (0.048)	−0.015 (0.047)	0.009 (0.041)	−0.021 (0.041)	−0.021 (0.039)
Precarious working position (ref. no such position)		−0.020 (0.063)	0.007 (0.064)		−0.036* (0.039)	−0.012 (0.038)
Occupation (ref. unemployed)						
Manager		0.101* (0.046)	0.060 (0.045)		0.044 (0.036)	0.020 (0.037)
Professional		0.088 (0.040)	0.061 (0.038)		0.029 (0.032)	0.021 (0.032)
Clerk/asisstant professional		0.112** (0.044)	0.088* (0.042)		0.017 (0.035)	0.006 (0.034)
Skilled worker		0.050 (0.057)	0.026 (0.056)		0.055 (0.048)	0.033 (0.044)
Unskilled worker		0.067* (0.074)	0.061 (0.070)		0.011 (0.050)	−0.000 (0.053)
Other occupation		0.027 (0.080)	0.017 (0.072)		0.008 (0.061)	0.005 (0.054)
Interpersonal support		−0.176*** (0.002)	−0.115** (0.002)		−0.140*** (0.001)	−0.038 (0.002)
Political affiliation strength		−0.017 (0.014)	−0.019 (0.013)		0.001 (0.010)	0.001 (0.010)
Political affiliation (ref. no pol. affiliation)						
Democrat		0.008 (0.053)	0.041 (0.051)		0.083 (0.038)	0.136** (0.038)
Republican		0.045 (0.055)	0.059 (0.053)		0.104* (0.041)	0.123** (0.040)
Other political affiliation		0.015 (0.052)	0.042 (0.050)		0.054 (0.037)	0.088* (0.037)
Religiosity (strength)		−0.012 (0.010)	0.020 (0.010)		0.027 (0.008)	0.069* (0.008)
Religion (ref. nonreligious)						
Christian		0.048 (0.041)	0.053 (0.040)		−0.032 (0.032)	−0.032 (0.030)
Hindu		0.085* (0.047)	0.107** (0.054)		0.062 (0.044)	0.050 (0.044)
Spiritual		0.027 (0.060)	0.039 (0.062)		−0.018 (0.040)	−0.016 (0.042)

Table 3. (Cont.) Regression analyses on one's enjoyment of trolling and likelihood of troll.

	(1a) Enjoy trolling	(2a) Enjoy trolling	(3c) Enjoy trolling	(1b) Likely to troll	(2b) Likely to troll	(3b) Likely to troll
Religion (ref. nonreligious)						
Other religion		−0.023	−0.014		−0.052*	−0.051*
		(0.067)	(0.066)		(0.052)	(0.051)
Percentage political posts		0.108**	0.104*		0.229***	0.174***
		(0.001)	(0.001)		(0.001)	(0.001)
Empathy			−0.101*			−0.056
			(0.024)			(0.018)
Big5 Conscientiousness			−0.115**			−0.142***
			(0.002)			(0.002)
Big5 Agreeableness			−0.070			−0.106***
			(0.003)			(0.002)
Big5 Extroversion			0.098**			0.032
			(0.002)			(0.002)
Big5 Neuroticism			−0.061			−0.067*
			(0.002)			(0.002)
Big5 Openness			−0.021			−0.040
			(0.002)			(0.001)
Heavy social media use			0.042			0.010
			(0.034)			(0.024)
Upward online comparison			−0.006			−0.039
			(0.012)			(0.009)
Downward online comparison			0.027			0.156***
			(0.012)			(0.011)
Percentage of similar groups			−0.072*			−0.080***
			(0.000)			(0.000)
Percentage of dissimilar groups			−0.052			0.008
			(0.001)			(0.000)
Male (ref. female or diverse)			0.028			0.019
			(0.025)			(0.020)
Age			−0.066*			−0.030
			(0.001)			(0.001)
LGBTQIA+ (ref. heterosexual)			−0.030			0.003
			(0.031)			(0.023)
Race (ref. white)						
Black			0.009			0.002
			(0.047)			(0.032)
Other			−0.057			−0.015
			(0.034)			(0.027)
Constant	0.187***	0.487***	1.080***	−0.019	0.215***	0.816***
	(0.026)	(0.094)	(0.169)	(0.020)	(0.065)	(0.126)
Observations	909	909	909	909	909	909
R^2	0.32	0.37	0.43	0.51	0.59	0.65

Notes: Standardized beta coefficients, standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

As for identity and political variables, being more engaged in online politics is positively related to both enjoyment of trolling and the likelihood to troll, but being a Republican only increases the likelihood to troll, whereas being a Hindu only makes it more enjoyable. It seems that general political involvement does not differentiate between different types of trolling, but specific political affiliation does. It is not related to the enjoyment of trolling, since if you troll because you defend your position, fun might also not be the main motivation. The significance of Hinduism is related to Hindutva, a form of Hindu nationalism that strongly clashes with the Muslim community (Udupa, 2018) and therefore probably causes some extreme trolling. Religion, however, is not related to the general likelihood to troll.

As for models with further control variables, they do render most trolling motivations insignificant. The only variable maintaining its significance is feeling Powerful and Happy while trolling. Also, Attacked and Annoyed becomes significant with respect to the enjoyment of trolling in the third model, probably due to the stronger irritability and general preparedness of trolls who enjoy trolling. All in all, this means that the impact of personality factors empathy, conscientiousness, and extroversion on the enjoyment of trolling, and of conscientiousness, agreeableness, and neuroticism on the likelihood to troll do not completely confound the role of social or situational determinants. The influence of the former became even stronger (see political affiliations and religiosity in Model 3b). Being involved in more similar groups online stops one from both forms of trolling, whereas downward online comparison increases only the likelihood to troll, which might suggest that it is more related to defensive/reactionary trolling and to increase one's self-worth through trolling. Except for the association between being of younger age and having more fun while trolling, there are no other significant relations between sociodemographics and trolling. This might be due to the numerous other variables we control for in the regressions.

To summarize, the enjoyment of trolling and the likelihood to troll share some of the same determinants but also present some differences. First, defensive motivations are only associated with likelihood to troll. Second, having employment (i.e., a higher social status) increases trolling for fun, whereas precarious working conditions are only related to the likelihood to troll. Third, with the exception of Hinduism and its association with trolling for fun, political affiliation and religiosity do impact only on likelihood to troll. In this regard, we found that people with strong political identities and religious beliefs engage in trolling, but do not necessarily get any satisfaction from it.

6. Conclusion and Discussion

The objective of this study was not so much to demonstrate the untruth of present theories of trolling describ-

ing it as malicious behavior, as to give more place to other forms and causes of this behavior—especially the more defensive/reactive forms of trolling, which appear to be partially driven by socioeconomic and sociopolitical factors. The idea of trolls as cold psychopaths who gain pleasure from hurting others is strongly psychology-driven and accounts only for personality traits. Drawing on more general theories of social deviance and negative politics framework, we extended the scope of analysis and found that trolling behavior might also be motivated by social isolation, low social status, and strong political partisanship.

In the pages above, we were able to take some first steps towards analyzing how these socioeconomic and sociopolitical factors impact on two forms of trolling: trolling as entertainment and defensive/reactive trolling. First, we showed that enjoyment of trolling is distinguishable from general trolling behavior and that the latter is partially driven by other factors, such as religiosity or political affiliation. Second, we demonstrated that trolling can be perceived not only as an enjoyable activity motivated by silliness, boredom, and/or a desire to show one's power; it also sometimes occurs in response to an attack by a provoking third party, or due to anxiety or stress, and may leave the trolling individual with regrets. Third, these different trolling motivations can be grouped into two different categories (trolling for fun and defensive/reactive trolling), each of which shows different patterns of correlation (thus, defensive trolling motivations appear to have no connection to the enjoyment of trolling).

Nevertheless, our study has several limitations. Although we were able to show some different underlying motivations for trolling behavior, the main dependent variables used in our analyses still differentiate too poorly between trolling for fun and defensive/reactive trolling. We were only able to distinguish enjoyment of trolling from general trolling behavior—not from defensive trolling in particular. This might also help explain why our regression results only hint at how these different ways of trolling present different patterns of association with respect to social, political, personality, and situational factors, without offering unequivocal evidence for different explanatory patterns. The second reason why some socio-structural variables underperformed in our study is that the survey used for this study was geared towards psychology rather than the social sciences. Hence, some of the crucial socioeconomic variables such as education, class, and income—all of which are standard in sociological or political science research—are missing. This means that we would profit from further analyses with better data in order to solidify these currently preliminary results.

The trolling research would also benefit from a similar study that controlled not for Big5 personality traits, but for dark ones. This would yield far less ambivalent results with respect to trolling behavior. Further research could also attempt to shed more light on how the Attacked and Annoyed motivation relates to different

sorts of trolling, and disentangle the complex and counterintuitive findings of this study. It merits noting that our initial hypothesis—that anger and the experience of being attacked have a stronger association with defensive/reactive trolling than with trolling for fun—turns out not to be the case. Rather, they were negatively related to likelihood to troll (even after controlling for other motivations). By contrast, this factor becomes positively associated with the enjoyment of trolling when analyzed in regression models with more controls. Although we have provided a few possible explanations of these results, a more thorough inquiry would be very useful.

The last limitation we want to mention is how we measured motivation. Our survey asks how one "feels/is" while trolling, not exactly why one is trolling. While the omission of the direct question could be seen as counterproductive, we believe that this slightly concealing formulation is merited due to the risk of rationalization or dishonesty on the part of the respondent if asked directly. Also, the survey makes it nearly impossible to disentangle the causal relationship between the reported emotional state and the action. Did the trolling person feel a certain way (for example, powerful) before starting the trolling behavior, or rather start to feel this way unintentionally while trolling, or begin to troll with the objective to feel powerful? We need research that seeks to separate out these processes, especially since part of our argument distinguishes between trolling as a way of making oneself feel less powerless (defensive/reactive trolling) and trolling for fun from a place of power.

Conflict of Interests

The authors declare no conflict of interests.

Supplementary Material

Supplementary material for this article is available online in the format provided by the authors (unedited).

References

Abts, K., & Baute, S. (2021). Social resentment, blame attribution and Euroscepticism: The role of status insecurity, relative deprivation and powerlessness. *Innovation: The European Journal of Social Science Research*, *35*(1), 39–64.

Adamczyk, A. (2012). Understanding delinquency with friendship group religious context: Delinquency with friendship group religious context. *Social Science Quarterly*, *93*(2), 482–505.

Agnew, R. (1995). Testing the leading crime theories: An alternative strategy focusing on motivational processes. *Journal of Research in Crime and Delinquency*, *32*(4), 363–398.

Andersen, I. V. (2021). Hostility online: Flaming, trolling, and the public debate. *First Monday*, *26*(3). https://doi.org/10.5210/fm.v26i3.11547

Aseltine, R. H., Gore, S., & Gordon, J. (2000). Life stress, anger and anxiety, and delinquency: An empirical test of general strain theory. *Journal of Health and Social Behavior*, *41*(3), 256–275.

Bail, C. A. (2021). *Breaking the social media prism: How to make our platforms less polarizing*. Princeton University Press.

Bonnette, A., Robinson, A., Dailey, S., Ceballos, N. A., & Howard, K. (2019). *Upward social comparisons and posting under the influence: Investigating social media behaviors of U.S. adults with generalized anxiety disorder*. Spotlight on Research. https://www.spotlightonresearch.com/mental-health-research/sormhbonnette19

Bor, A., & Petersen, M. B. (2021). The psychology of online political hostility: A comprehensive, cross-national test of the mismatch hypothesis. *American Political Science Review*, *116*(1), 1–18.

Buckels, E. E., Trapnell, P. D., Andjelovic, T., & Paulhus, D. L. (2019). Internet trolling and everyday sadism: Parallel effects on pain perception and moral judgment. *Journal of Personality*, *87*(2), 328–340.

Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, *67*, 97–102.

Bulut, E., & Yörük, E. (2017). Digital populism: Trolls and political polarization of Twitter in Turkey ERG. *International Journal of Communication*, *11*, 4093–411.

Campbell, W. K., & Miller, J. D. (Eds.). (2011). *The handbook of narcissism and narcissistic personality disorder: Theoretical approaches, empirical findings, and treatments*. Wiley.

Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). Anyone can become a troll: Causes of trolling behavior in online discussions. In C. P. Lee & S. Poltrock (Eds.), *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 1217–1230). Association for Computing Machinery.

Cioban, S., Lazăr, A. R., Bacter, C., & Hatos, A. (2021). Adolescent deviance and cyber-deviance: A systematic literature review. *Frontiers in Psychology*, *12*, Article 4416.

Cook, C., Schaafsma, J., & Antheunis, M. (2018). Under

the bridge: An in-depth examination of online trolling in the gaming context. *New Media & Society*, *20*(9), 3323–3340.

Craker, N., & March, E. (2016). The dark side of Facebook®: The dark tetrad, negative social potency, and trolling behaviours. *Personality and Individual Differences*, *102*, 79–84.

Festl, R., & Quandt, T. (2013). Social relations and cyberbullying: The influence of individual and structural attributes on victimization and perpetration via the internet. *Human Communication Research*, *39*(1), 101–126.

Fichman, P., & Sanfilippo, M. R. (2015). The bad boys and girls of cyberspace: How gender and context impact perception of and reaction to trolling. *Social Science Computer Review*, *33*(2), 163–180.

Fichman, P., & Sanfilippo, M. R. (2016). *Online trolling and its perpetrators: Under the cyberbridge*. Rowman & Littlefield.

Flores-Saviaga, C., Keegan, B., & Savage, S. (2018). Mobilizing the trump train: Understanding collective action in a political trolling community. In J. Hancock (Ed.), *Proceedings of the Twelfth International AAAI Conference on Web and Social Media* (pp. 82–91). AAAI Press.

Foster, J. D., & Trimm, R. F. (2008). On being eager and uninhibited: Narcissism and approach–avoidance motivation. *Personality and Social Psychology Bulletin*, *34*(7), 1004–1017.

Frischlich, L., Schatto-Eckrodt, T., Boberg, S., & Wintterlin, F. (2021). Roots of incivility: How personality, media use, and online experiences shape uncivil participation. *Media and Communication*, *9*(1), 195–208.

Gidron, N., Adams, J., & Horne, W. (2019). Toward a comparative research agenda on affective polarization in mass publics. *APSA Comparative Politics Newsletter*, *29*, 30–36.

Gidron, N., & Hall, P. A. (2017). The politics of social status: Economic and cultural roots of the populist right. *The British Journal of Sociology*, *68*(S1), 57–84.

Gylfason, H. F., Sveinsdottir, A. H., Vésteinsdóttir, V., & Sigurvinsdottir, R. (2021). Haters gonna hate, trolls gonna troll: The personality profile of a Facebook troll. *International Journal of Environmental Research and Public Health*, *18*(11), Article 5722.

Hagan, J., & Foster, H. (2001). Youth violence and the end of adolescence. *American Sociological Review*, *66*(6), 874–899.

Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research. Language, Behaviour, Culture*, *6*(2), 215–242.

Hare, R. D. (2006). Psychopathy: A clinical and forensic overview. *Psychiatric Clinics of North America*, *29*(3), 709–724.

Harteveld, E., Mendoza, P., & Rooduijn, M. (2021). Affective polarization and the populist radical right: Creating the hating? *Government and Opposition*, *57*(4),

703–727. https://doi.org/10.1017/gov.2021.31

Herring, S., Job-Sluder, K., Scheckler, R., & Barab, S. (2002). Searching for safety online: Managing "trolling" in a feminist forum. *The Information Society*, *18*(5), 371–384.

Hochschild, A. R. (2016). *Strangers in their own land: Anger and mourning on the American right*. New Press.

Hopkinson, C. (2013). Trolling in online discussions: From provocation to community-building. *Brno Studies in English*, *39*(1), 5–25.

Howard, K., Zolnierek, K. H., Critz, K., Dailey, S., & Ceballos, N. (2019). An examination of psychosocial factors associated with malicious online trolling behaviors. *Personality and Individual Differences*, *149*, 309–314.

Jakobwitz, S., & Egan, V. (2006). The dark triad and normal personality traits. *Personality and Individual Differences*, *40*(2), 331–339.

Jones, D. N., & Paulhus, D. L. (2011). The role of impulsivity in the dark triad of personality. *Personality and Individual Differences*, *51*(5), 679–682.

Jones, S., Miller, J., & Lynam, D. (2011). Personality, antisocial behavior, and aggression: A meta-analytic review. *Journal of Criminal Justice*, *39*(4), 329–337.

Juvonen, J., & Graham, S. (2014). Bullying in schools: The power of bullies and the plight of victims. *Annual Review of Psychology*, *65*(1), 159–185.

Lee, B. H. (2018). Explaining cyber deviance among school-aged youth. *Child Indicators Research*, *11*(2), 563–584.

Li, Y.-J., Cheung, C. M. K., Shen, X.-L., & Lee, M. K. O. (2022). When socialization goes wrong: Understanding the we-intention to participate in collective trolling in virtual communities. *Journal of the Association for Information Systems*, *23*(3), 678–706. https://doi.org/10.17705/1jais.00737

Liu, P., Guberman, J., Hemphill, L., & Culotta, A. (2018). Forecasting the presence and intensity of hostility on Instagram using linguistic and social features. In J. Hancock (Ed.), *Proceedings of the Twelfth International AAAI Conference on Web and Social Media* (pp. 181–190). AAAI Press.

Lopes, B., & Yu, H. (2017). Who do you troll and why: An investigation into the relationship between the dark triad personalities and online trolling behaviours towards popular and less popular Facebook profiles. *Computers in Human Behavior*, *77*, 69–76.

Maltby, J., Day, L., Hatcher, R. M., Tazzyman, S., Flowe, H. D., Palmer, E. J., Frosch, C. A., O'Reilly, M., Jones, C., Buckley, C., Knieps, M., & Cutts, K. (2016). Implicit theories of online trolling: Evidence that attention-seeking conceptions are associated with increased psychological resilience. *British Journal of Psychology*, *107*(3), 448–466.

Marchal, N. (2021). "Be nice or leave me alone": An intergroup perspective on affective polarization in online political discussions. *Communication Research*, *49*(3), 376–398.

Matthews, J., & Goerzen, M. (2019). Black hat trolling, white hat trolling, and hacking the attention landscape. In L. Liu & R. White (Eds.), *Companion proceedings of the 2019 World Wide Web Conference* (pp. 523–528). Association for Computing Machinery.

May, D., Bossler, A., & Holt, T. (2012). Predicting online harassment victimization among a juvenile Population. *Youth & Society*, *44*(4), 500–523.

Nai, A. (2021). Fear and loathing in populist campaigns? Comparing the communication style of populists and non-populists in elections worldwide. *Journal of Political Marketing*, *20*(2), 219–250.

Nai, A., & Maier, J. (2018). Perceived personality and campaign style of Hillary Clinton and Donald Trump. *Personality and Individual Differences*, *121*(8), 80–83.

Nai, A., Maier, J., & Vranić, J. (2021). Personality goes a long way (for some): An experimental investigation into candidate personality traits, voters' profile, and perceived likeability. *Frontiers in Political Science*, *3*, Article 636745.

Nevin, A. (2015). *Cyber-psychopathy: Examining the relationship between dark e-personality and online misconduct* [Master's thesis, The University of Western Ontario]. Scholarship@Western. https://ir.lib.uwo.ca/cgi/viewcontent.cgi?referer=&httpsredir=1&article=4531&context=etd

Nyhuis, D., Song, H., & Boomgaarden, H. (2020). Disproportionality in media representations of campaign negativity. *Political Science Research and Methods*, *9*(3), 519–531.

Paasch-Colberg, S., Strippel, C., Trebbe, J., & Emmer, M. (2021). From insult to hate speech: Mapping offensive language in German user comments on immigration. *Media and Communication*, *9*(1), 171–180.

Pao, E. (2015, July 16). Former Reddit CEO Ellen Pao: The trolls are winning the battle for the Internet. *Washington Post*. https://www.washingtonpost.com/opinions/we-cannot-let-the-internet-trolls-win/2015/07/16/91b1a2d2-2b17-11e5-bd33-395c05608059_story.html

Pew Research Center. (2016). *Research in the crowdsourcing age, a case study: How scholars, companies and workers are using Mechanical Turk, a "gig economy" platform, for tasks computers can't handle*. https://www.pewresearch.org/internet/2016/07/11/research-in-the-crowdsourcing-age-a-case-study

Quandt, T., Klapproth, J., & Frischlich, L. (2022). Dark social media participation and well-being. *Current Opinion in Psychology*, *45*, Article 101284. https://doi.org/10.1016/j.copsyc.2021.11.004

Rafalow, M. H. (2015). n00bs, trolls, and idols: Boundary-making among digital youth. In S. L. Blair, P. N. Claster, & S. M. Claster (Eds.), *Sociological studies of children and youth* (Vol. 19, pp. 243–266). Emerald Group.

Reiljan, A. (2020). "Fear and loathing across party lines" (also) in Europe: Affective polarisation in European party systems. *European Journal of Political Research*, *59*(2), 376–396.

Salmela, M., & Capelos, T. (2021). Ressentiment: A complex emotion or an emotional mechanism of psychic defences? *Politics and Governance*, *9*(3), 191–203.

Salmela, M., & von Scheve, C. (2018). Emotional dynamics of right- and left-wing political populism. *Humanity & Society*, *42*(4), 434–454.

Sanfilippo, M. R., Yang, S., & Fichman, P. (2017). *Managing online trolling: From deviant to social and political trolls*. In T. X. Bui & R. Sprague, Jr. (Eds.), *Proceedings of the 50th Hawaii International Conference on System Sciences* (pp. 1802–1811). HICSS.

Seidman, G. (2013). Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and Individual Differences*, *54*(3), 402–407.

Sentse, M., Scholte, R., Salmivalli, C., & Voeten, M. (2007). Person–group dissimilarity in Involvement in bullying and its relation with social status. *Journal of Abnormal Child Psychology*, *35*(6), 1009–1019.

Sest, N., & March, E. (2017). Constructing the cyber-troll: Psychopathy, sadism, and empathy. *Personality and Individual Differences*, *119*, 69–72.

Shachaf, P., & Hara, N. (2010). Beyond vandalism: Wikipedia trolls. *Journal of Information Science*, *36*(3), 357–370.

Spears, R. (2017). Social identity model of deindividuation effects. In P. Rössler (Ed.), *The international encyclopedia of media effects*. https://doi.org/10.1002/9781118783764.wbieme0091

Toro, R., García-García, J., & Zaldívar-Basurto, F. (2020). Antisocial disorders in adolescence and youth, according to structural, emotional, and cognitive transdiagnostic variables: A systematic review. *International Journal of Environmental Research and Public Health*, *17*(9), Article 3036.

Udupa, S. (2018). Enterprise Hindutva and social media in urban India. *Contemporary South Asia*, *26*(4), 453–467.

About the Authors

Monika Verbalyte holds a BA in political science from the University of Vilnius (2007) and an MA in European sociology from Freie Universität Berlin (2010). Since September 2020, Monika is a research fellow at Europa-Universität Flensburg in the project "Value Conflicts in a Differentiated Europe" (ValCon). Monika's research interests are emotion sociology, political communication, political attitudes and behavior, European sociology, political scandals, values and norms, affective and political polarization, social network analysis, comparative survey research, discourse methodology, and qualitative interviews.

Christoph Keitel holds a master's degree in European Studies. During his graduate studies, he worked as a student assistant in the research project "Value Conflicts in a Differentiated Europe: The Impact of Digital Media on Value Polarisation in Europe." Following a tutoring position, Christoph is currently on a break from academia, working as a junior project manager for online market research at the Hamburg-based Trend Research GmbH. Christoph's research interests include media studies, political polarization, and behavioural science.

Krista Howard earned her PhD in health psychology and neuroscience in 2010 from the University of Texas at Arlington. She is a tenured professor in the Psychology Department at Texas State University. Dr Howard studies the relationship between mental health and health behavior outcomes, with a focus on social media behaviors, occupational health, chronic illnesses, and maternal health.