

# Online aggression from a sociological perspective: An integrative view on determinants and possible countermeasures

**Lea Stahel**

University of Zurich  
Department of Sociology  
Andreasstrasse 15, CH-8003 Zurich  
stahel@soziologie.uzh.ch

**Sebastian Weingartner**

University of Zurich  
Department of Sociology  
Andreasstrasse 15, CH-8003 Zurich  
weingartner@soziologie.uzh.ch

## Abstract

The present paper introduces a theoretical model for explaining aggressive online comments from a sociological perspective. It is innovative as it combines individual, situational, and social-structural determinants of online aggression and tries to theoretically derive their interplay. Moreover, the paper suggests an empirical strategy for testing the model. The main contribution will be to match online commenting data with survey data containing rich background data of non-/aggressive online commentators.

## 1 Introduction

In the past years, online aggression in social media has attracted a lot of attention not only in the broader public but also in academia (e.g. Cicchirillo et al. 2015; Sydnor 2018). Studies show that offending, defaming, or threatening online comments posted by Internet users fundamentally negatively affect the targeted persons' well-being, social harmony, and democratic outcomes (e.g. Anderson et al., 2014; Bauman, 2013; Kwon and Gruzd, 2017). Accordingly, knowing why people aggress online is the first step to counter it. Although previous research on online aggression has been successful in suggesting and explaining single determinants driving aggressive online commenting (see studies in the *State of Research* below), (1) their interplay has hardly been studied due to the lack of an overarching theoretical framework and (2) socio-structural determinants have been largely ignored so far. Moreover, from a methods point of view, (3) there are no studies that systematically link digital commenting data to offline information on adult aggressors from the wider population.

Hence, the present paper introduces a theoretical model that relates several determinants of online

aggression to each other in a more general framework of sociological explanation. Based on the model, we aim to answer the following research questions: (1) Which individual determinants, situational determinants, and social-structural determinants drive online aggression? (2) How do various determinants relate to each other when producing aggressive online behavior? (3) Are there differences in online aggression between social-structural groups?

Answering such questions requires a specific empirical strategy. We intend to conduct a large-scale quantitative survey in German-speaking Switzerland, including aggressive and non-aggressive online commentators. They are drawn from a large population of commentators having submitted to online commentary sections of a large Swiss media organization. We match their survey information with their commenting behavior, ranging from non-aggressive to frequently aggressive (this classification emerges from human/automated content analysis).

We will elaborate on the theoretical model and the planned empirical strategy in the following sections. First, however, we will describe in more detail the current state of online aggression (OA in the remainder of the paper) research.

## 2 State of research

In the literature so far, determinants of OA are explored primarily from three different perspectives: the individual-psychological, the situational, and the social-structural. All three perspectives are shortly reviewed here, from the fields of psychology, political science, and communication.

### 2.1 Psychological-individual determinants

From a psychological-individual perspective, OA can on the one hand be motivated by relatively stable psychological traits ("aggressors as

antisocial individuals”). The underlying theory proposes that each individual has a unique personality and that associated traits motivate behavior and thus (online) aggression. For example, online aggressors score relatively higher in narcissism, psychopathy, and Machiavellianism (e.g. Abell and Brewer, 2014), might lack empathy (Steffgen et al., 2011), may be less open, low in self-control, and impulsive (Peterson and Densley, 2017), but also more depressive and shy (Bauman, 2013).

On the other hand, OA can be motivated by less stable individual emotions, beliefs, and goals (“aggressors as venting, convinced Internet activists”). For example, people in negative mood may troll (Cheng et al., 2017), being angry at unfair negotiators motivates to digitally aggress (Johnson et al., 2009), and car drivers vent their rage (Stephens et al., 2016). Also, online aggressors believe that they do not get caught and that their online content is not permanently stored (Wright, 2013). Further, people participating in collective online outrage are motivated by moral heuristics and moral beliefs (e.g. based on moral disengagement theory by Faulkner and Bliuc, 2016) and punishing violators of social norms (based on social norm theory; Rost et al., 2016). Finally, online aggressors have goals. They spread political ideologies, seek thrill and fun, draw attention to social injustice (Erjavec and Kovačič, 2012), or seek social standing, status, and recognition (e.g. Ballard and Welch, 2017).

## 2.2 Situational determinants

Research on situational determinants suggests that online aggressive individuals are influenced by properties of the digital media environment and the surrounding social and situational context (“aggressors as ordinary people, but situationally-driven”). The psychological-communicative Reduced cues approach (Sproull and Kiesler, 1986) argues that properties of online environments may cause toxic online disinhibition (Suler, 2004): people feel less restraint because of the absence of social-context cues, anonymity, invisibility, asynchronicity, or minimization of authority. This is explained either by deindividuation theories (Diener, 1980) or by the social identity model of deindividuation effects (SIDE) which argues that deindividuation triggered by reduced social cues and anonymity in online settings boosts the salience of individuals’

social identity relative to their personal identity. Thus, if a group norm is salient (e.g. in an online forum), commentators will conform to it rather than engage in uncontrolled aggressive behavior (Reicher et al., 1995). SIDE is empirically supported in several settings (e.g. Hmielowski et al., 2014).

OA is also explained by social learning theories and situational social control. For example, perceiving flaming norms socializes people into flaming (Cheng et al., 2017). Also, people more likely aggress online if informal social controls from an effective community policy and peer pressure are lacking, predicted by routine activity theories of crime (Navarro and Jasinski, 2012), deterrence theory (Xu et al., 2016), or social norms (Álvarez-Benjumea and Winter, 2018). Similarly, people more likely aggress if they have become cyber-victims themselves (Quintana-Orts and Rey, 2018), receive comments challenging their beliefs (Hutchens et al., 2015) or threatening their face (Masullo Chen and Lu, 2017), or if public actors misbehave (Johnen et al., 2017; Rost et al., 2016). Finally, legal frameworks, ethical guidelines, and moderation strategies set up by online (news) platforms may be situationally influential (Ksiazek, 2015).

## 2.3 Social-structural determinants

Research on social-structural determinants is very scarce. It includes socio-demographics, social group memberships, and structural positions and relations. Accordingly, OA may differ by cultural and national backgrounds (Shapka et al., 2018), gender (Ballard and Welch, 2017; Bauman, 2013; Shapka et al., 2018), and age (Bauman, 2013; Shapka et al., 2018). Also, incivility on Twitter is higher in areas of low socioeconomic status (SES), low social capital potential (i.e. potential for interconnected citizen networks), and low in-district partisan polarization (Vargo and Hopp, 2017). Finally, (few) structural and sociodemographic factors are considered in the social media cyberbullying model (SMCBM) model by (Lowry et al., 2016).

## 2.4 Gaps

Reviewing the literature on OA, several gaps emerge. Theoretically, there is, first, no overarching theoretical framework integrating the determinants suggested. Accordingly, theoretical approaches to cyberbullying are “sparse and

piecemeal” (Espelage et al., 2012: 49) and „have received scant conceptual development” (Runions, 2013: 751). Hence, a major task of future research is to develop “a comprehensive theoretical model that might ground the conversation about cyber aggression and violence” (Peterson and Densley, 2017: 197). At best, such a model addresses the “interaction between micro, meso, and macro levels of explanation” in order to overcome research’s current lack of “continuity and coherence” (Peterson and Densley, 2017: 197).

Second, there is a need to relate OA more systematically to social-structural factors. Up to now, information on aggressors and their aggression-benefiting circumstances is limited (Coe et al., 2014: 675; Peterson and Densley, 2017:195). Especially with regard to potential “civility divides” (Vargo and Hopp, 2017: 26), exploring socio-demographic and socio-economic determinants (such as gender, age, education, or prestige) enables to empirically test whether “those equipped with economic and social privilege in the off-line realm may disproportionately gain value from online deliberation, while those with diminished economic and social resources may interact in a hostile, uncivil, (...) strata of the Internet” (Vargo and Hopp, 2017: 24; also see Cicchirillo et al., 2015).

Third, there are no studies that systematically link digital commenting data to offline information in a large sample of adult aggressors. Most studies only use natively online data. If offline information is collected at all, then it is linked to OA intentions, self-reports, or experimental triggers, at best.

### 3 Theoretical model

Here, we introduce an integrative model that relates a multitude of determinants to each other in a general framework of sociological explanation, also explicitly theorizing social-structural determinants. This model builds on the ideas of structural individualism (Coleman, 1994) and the model of frame-selection (Esser, 2001; Kroneberg, 2011).

Basically, structural individualism aims at dissecting social phenomena into its constitutive parts, that is meaningful decisions of individual actors. These decisions, however, are embedded in a configuration of social structures and institutions. This social context, in turn, affects (if correctly perceived) actors’ goals, beliefs, and opportunities,

which then guide their behavior (Maurer and Schmid, 2010; Udehn 2001). From this perspective, OA comments are defined as individual decisions (actions) which are in a first step explained by both characteristics of the individual (e.g. beliefs) and situational parameters (e.g. others’ behaviors). In a second step, individual determinants are related to social-structural background. The relationship between these two sets of determinants can be thought of in several ways: social context conditions may structure the set of behavioral alternatives available, the behavioral costs, and an individual’s preferences, attitudes, and body of knowledge. Theoretically, this can be explained by learning theories (Bandura 1977) or social production function theory (Ormel et al. 1999).

It needs to be specified, then, how individual decisions come about. This is important because the theory of action chosen has an impact on which individual and situational determinants can be taken into account. Instead of relying on a rather simple rational-choice approach for explaining individual decisions, we opt for the more elaborate model of frame-selection (MFS) as introduced by Kroneberg (2011, 2014). In classical rational-choice theory (Opp, 1999), it is assumed that actors choose those behavioral alternatives which they expect to best fulfill their preferences given certain behavioral constraints. Thus, behavior is a function of individual goals (evaluative beliefs, including egoistic just as prosocial goals), beliefs about the consequences of decisions, and behavioral constraints (the latter two are often summarized as descriptive beliefs). However, rational-choice theory is silent about which descriptive and evaluative beliefs are active in a specific decision situation. Therefore, MFS explicitly incorporates the process of the definition of the situation (Esser, 1996). In this process, actors subjectively define which kind of situation they are actually facing (which may – in contrast to rational-choice theory – deviate from “objective” situational requirements). They do so by synchronizing given situational cues with internalized knowledge about typical situations (frames). Hence, descriptive and evaluative beliefs guiding behavior are not taken for granted but depend on actors’ subjective perceptions of the situation. This means that behavioral differences between (groups of) actors do not simply result from individual or situational

differences, but from interactions between individual and situational characteristics.

Based on these theoretical considerations, we propose the following explanatory model of OA (Figure 1):

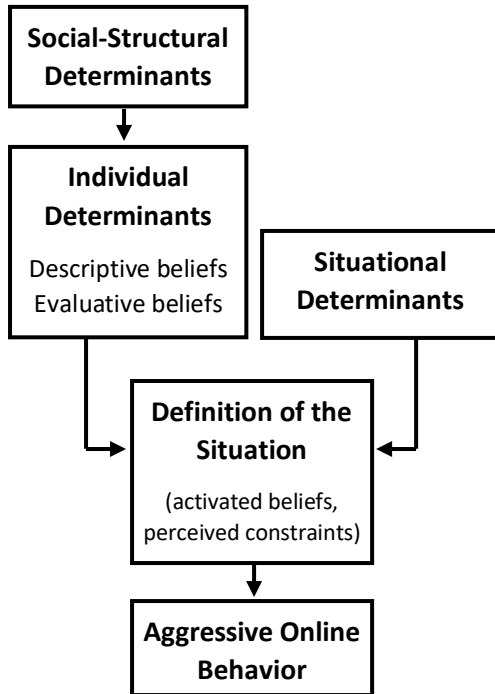


Fig 1: Explanatory model of aggressive online behavior.

In this model, OA behavior results from individuals' definitions of potential online commenting situations. Such definitions represent a situation's general meaning and thus determine which individual beliefs are activated and which situational constraints are perceived by the actor. How situations are defined depends on two sets of factors: (1) situational determinants comprise all relevant characteristics of the situational context and thus are in principle identical for all actors in the same situation (but still differently perceived). (2) individual determinants comprise all descriptive (representations of current states of the world) and evaluative beliefs (representations of desired states of the world) of an individual and thus do not vary across situations for a specific actor. The interactive relationship between individual and situational determinants can be understood in two ways. Straightforwardly, it means that those individual beliefs (and opportunities) guide behavior which are activated by certain situational cues. This differs according to the overall set of beliefs internalized by the individual. However, if some descriptive or

evaluative beliefs are strongly internalized and thus chronically active, they can prompt a certain definition of the situation (and thus action) irrespective of the situational conditions given (possible misperception). As mentioned above, we assume descriptive and evaluative beliefs to be tied to social-structural determinants. In accordance with structural individualism, sociological factors such as socio-economic or demographic attributes are reflected in individual determinants. Hence, social-structural groups are expected to be similar in terms of certain beliefs. Overall, the model emphasizes that OA does neither result from characteristics of the individual, nor from characteristics of the situation, but rather from the interplay of these two.

#### 4 Empirical approach

The empirical study seeks to collect data on individual, situational, and social-structural determinants of OA behavior. Therefore, we intend to conduct an online survey in German-speaking Switzerland with four different groups: frequent OA commentators, occasional OA commentators, non-OA commentators, and non-commentators. Group-differences in determinants, then, allow to assess determinants' relative effect on OA behavior. However, sampling OA commentators is not easy because it is a relatively rare behavior. Thus, we apply an elaborate, two-step sampling strategy: First, in order to sample OA and non-OA commentators, we use the unique opportunity to collaborate with a large Swiss media corporation. We will use a large dataset of news comments submitted to its website (including meta-data such as time of submission). The dataset includes moderated comments: comments considered as being non-aggressive by moderators (and were published in the commentary section) and comments considered as aggressive (and were not published). By employing human/automated content analysis of all comments, we identify the following groups and assign all commentators to one of them: frequently aggressive commentators, occasionally aggressive commentators, and non-aggressive commentators. From each group, we invite around 1500 people to participate in the survey. Second, in order to sample persons who do not engage in online commenting at all (non-commentators), we use a random sample of the resident population of German-speaking Switzerland.

Particular attention is given to data protection and the ethics of recruiting. First, all the comments and meta data received by the Swiss media corporation is principally public data, thus principally searchable and retrievable. This is because commentators submit their comments to news platforms in the knowledge that their comments get principally published (even in cases where comments are ultimately not published by moderators). Beyond, this data set is given to us in an anonymized form. Thus, privacy concerns can be excluded. Second, not the authors but the Swiss media corporation invites the commentators to participate in the survey (as the e-mail addresses of commentators are only available to the corporation but not to us). Third, by forming groups of commentators (see above) whereby individuals in each group receive group-specific online surveys, the survey data of individuals will only be connected to the affiliation to these groups but at no time to individual comments or commentators. This makes it impossible to identify single individuals in the resulting data set. Fourth, an ethics approval will be sought in the process of designing the survey.

Our approach of matching online data with survey data allows to combine behavioral data with a broad range of – so far scarcely collected – individual, social-structural, and situational determinants of OA. While individual and social-structural determinants will mainly be measured in the survey, most situational determinants will be measured through aggregating user-generated comments and meta-data.

## 5 Conclusion

The preceding paper introduced a novel, sociologically informed theoretical framework integrating a broad set of determinants of aggressive online commenting behavior. Furthermore, it suggested an empirical strategy allowing to disentangle the effects single determinants by matching online data with survey data.

## References

Abell, L., Brewer, G., 2014. Machiavellianism, self-monitoring, self-promotion and relational aggression on Facebook. *Computers in Human Behavior* 36, 258–262.

- Álvarez-Benjumea, A., Winter, F., 2018. Normative Change and Culture of Hate: An Experiment in Online Environments. *European Sociological Review* 29, 1–15.
- Anderson, A.A., Brossard, D., Scheufele, D.A., Xenos, M.A., Ladwig, P., 2014. The “nasty effect:” Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication* 19 (3), 373–387.
- Ballard, M.E., Welch, K.M., 2017. Virtual warfare: Cyberbullying and cyber-victimization in MMOG play. *Games and Culture* 12 (5), 466–491.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs NJ: Prentice Hall.
- Bauman, S., 2013. Cyberbullying: What does research tell us? *Theory into practice* 52 (4), 249–256.
- Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., Leskovec, J., 2017. Anyone can become a troll: Causes of trolling behavior in online discussions. arXiv preprint arXiv:1702.01119.
- Cicchirillo, V., Hmielowski, J., Hutchens, M., 2015. The mainstreaming of verbally aggressive online political behaviors. *Cyberpsychology, Behavior, and Social Networking* 18 (5), 253–259.
- Coe, K., Kenski, K., Rains, S.A., 2014. Online and uncivil?: Patterns and determinants of incivility in newspaper website comments. *Journal of Communication* 64 (4), 658–679.
- Coleman, J.S., 1994. *Foundations of social theory*. Harvard University Press, Cambridge.
- Diener, E., 1980. Deindividuation: The absence of self-awareness and self-regulation in group members. *The psychology of group influence* 209242.
- Erjavec, K., Kovačič, M.P., 2012. “You Don't Understand, This is a New War!” Analysis of Hate Speech in News Web Sites' Comments. *Mass Communication and Society* 15 (6), 899–920.
- Espelage, D.L., Rao, M.A., Craven, R.G., 2012. Theories of cyberbullying. Principles of cyberbullying research: Definitions, measures, and methodology, 49–67.
- Esser, H., 2001. *Soziologie. Spezielle Grundlagen*. Band 6: Sinn und Kultur. Campus, Frankfurt a. M.
- Esser, H., 1996. Die Definition der Situation. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 48 (1), 1–34.
- Faulkner, N., Bliuc, A.-M., 2016. ‘It’s okay to be racist’: moral disengagement in online discussions of racist incidents in Australia. *Ethnic and Racial Studies* 39 (14), 2545–2563.

- Hmielowski, J.D., Hutchens, M.J., Cicchirillo, V.J., 2014. Living in an age of online incivility: Examining the conditional indirect effects of online discussion on political flaming. *Information, Communication & Society* 17 (10), 1196–1211.
- Hutchens, M.J., Cicchirillo, V.J., Hmielowski, J.D., 2015. How could you think that?: Understanding intentions to engage in political flaming. *New Media & Society* 17 (8), 1201–1219.
- Johnen, M., Jungblut, M., Ziegele, M., 2017. The digital outcry: What incites participation behavior in an online firestorm? *New Media & Society*, 1461444817741883.
- Johnson, N.A., Cooper, R.B., Chin, W.W., 2009. Anger and flaming in computer-mediated negotiation among strangers. *Decision Support Systems* 46 (3), 660–672.
- Kroneberg, C., 2011. *Die Erklärung sozialen Handelns: Grundlagen und Anwendung einer integrativen Theorie*. VS-Verlag, Wiesbaden.
- Ksiazek, T.B., 2015. Civil interactivity: How news organizations' commenting policies explain civility and hostility in user comments. *Journal of Broadcasting & Electronic Media* 59 (4), 556–573.
- Kwon, K.H., Gruzd, A., 2017. Is aggression contagious online?: a case of swearing on donald trump's campaign videos on youtube. Proceedings of the 50th Hawaii International Conference on System Sciences, 2165–2174.
- Lowry, P.B., Zhang, J., Wang, C., Siponen, M., 2016. Why do adults engage in cyberbullying on social media?: An integration of online disinhibition and deindividuation effects with the social structure and social learning model. *Information Systems Research* 27 (4), 962–986.
- Masullo Chen, G., Lu, S., 2017. Online political discourse: Exploring differences in effects of civil and uncivil disagreement in news website comments. *Journal of Broadcasting & Electronic Media* 61 (1), 108–125.
- Maurer, A., Schmid, M., 2010. *Erklärende Soziologie: Grundlagen, Vertreter und Anwendungsfelder eines soziologischen Forschungsprogramms*. VS-Verlag, Wiesbaden.
- Navarro, J.N., Jasinski, J.L., 2012. Going cyber: Using routine activities theory to predict cyberbullying experiences. *Sociological Spectrum* 32 (1), 81–94.
- Opp, K.-D., 1999. Contending conceptions of the theory of rational action. *Journal of Theoretical Politics* 11 (2), 171–202.
- Ormel, J., S. Lindenberg, N. Steverink, et al. (1999). Subjective well-being and social production functions. *Social Indicators Research* 46(1): 613–90.
- Peterson, J., Densley, J., 2017. Cyber violence: What do we know and where do we go from here? *Aggression and violent behavior* 34, 193–200.
- Quintana-Orts, C., Rey, L., 2018. Forgiveness and cyberbullying in adolescence: Does willingness to forgive help minimize the risk of becoming a cyberbully? *Computers in Human Behavior* 81, 209–214.
- Reicher, S.D., Spears, R., Postmes, T., 1995. A social identity model of deindividuation phenomena. *European review of social psychology* 6 (1), 161–198.
- Rost, K., Stahel, L., Frey, B.S., 2016. Digital social norm enforcement: Online firestorms in social media. *PLoS One* 11 (6), e0155923.
- Runions, K.C., 2013. Toward a conceptual model of motive and self-control in cyber-aggression: Rage, revenge, reward, and recreation. *Journal of youth and adolescence* 42 (5), 751–771.
- Shapka, J.D., Onditi, H.Z., Collie, R.J., Lapidot-Lefler, N., 2018. Cyberbullying and Cybervictimization Within a Cross-Cultural Context: A Study of Canadian and Tanzanian Adolescents. *Child development* 89 (1), 89–99.
- Sproull, L., Kiesler, S., 1986. Reducing social context cues: Electronic mail in organizational communication. *Management science* 32 (11), 1492–1512.
- Steffgen, G., König, A., Pfetsch, J., Melzer, A., 2011. Are cyberbullies less empathic?: Adolescents' cyberbullying behavior and empathic responsiveness. *Cyberpsychology, Behavior, and Social Networking* 14 (11), 643–648.
- Stephens, A.N., Trawley, S.L., Ohtsuka, K., 2016. Venting anger in cyberspace: Self-entitlement versus self-preservation in# roadrage tweets. *Transportation research part F: traffic psychology and behaviour* 42, 400–410.
- Suler, J., 2004. The online disinhibition effect. *Cyberpsychology & behavior* 7 (3), 321–326.
- Sydnor, E. (2018). Platforms for incivility: Examining perceptions across different media formats. *Political Communication* 35(1): 97–116.
- Udehn, L. (2001). *Methodological individualism: Background, history and meaning*. London: Routledge.
- Vargo, C.J., Hopp, T., 2017. Socioeconomic status, social capital, and partisan polarity as predictors of political incivility on Twitter: a congressional

district-level analysis. *Social Science Computer Review* 35 (1), 10–32.

Wright, M.F., 2013. The relationship between young adults' beliefs about anonymity and subsequent cyber aggression. *Cyberpsychology, Behavior, and Social Networking* 16 (12), 858–862.

Xu, B., Xu, Z., Li, D., 2016. Internet aggression in online communities: a contemporary deterrence perspective. *Information Systems Journal* 26 (6), 641–667.