@NTIDOTE 2.0 - Hate speech among Belgian youth aged 15 to 25
"50 Shades of Hate Speech"

In recent years, the digital world has witnessed various forms of cyber violence. However, little attention has been given to understanding these behaviors, including their definitions, prevalence within the Belgian population, relevant explanatory models, or the online platforms that host them. Thus, the @ntidote 2.0 project specifically focuses on two forms of cyber violence: online hate speech and non-consensual dissemination of intimate images (NCII), particularly among young people aged 15 to 25. This note addresses one of the significant outcomes of this project, namely the definition of online hate speech as expressed by this audience. While our initial research questions were: **"How do young people aged 15 to 25 perceive online hate speech? What are their definitions of online hate speech, specifically within a diverse sample (age, gender, sexual orientation, cultural background, and self-reported status[1])?"** *We encountered complex results that broadly question the perception of young people regarding aggressions, even micro-aggressions, and their modes of expression in the digital space.*

Context and Research Question(s)

The qualification of online hate speech varies across disciplines and affiliated authors. A commonly accepted definition stipulates that online hate speech encompasses any form of expression (texts, videos, audios, photos, images, games, and others) through the use of the Internet (digital platforms, social networks, and others) that is motivated by prejudice, based on intolerance or discrimination, and targets a group of individuals (or an individual within that group) sharing a common, innate or acquired, current or perceived characteristic, such as cultural origin, beliefs, disabilities, gender, or sexual orientation (Almagor, 2011 and Council of Europe, 1997 cited by Chetty & Alathur, 2018; al Serhan & Elareshi, 2019; Awan, 2016 cited by al Serhan & Elareshi, 2019; Castano-Pulgarín et al., 2021; Council of Europe, 2013 cited by Keipi et al., 2017; Hawdon et al., 2017; Keipi et al., 2017; Simpson, 2013 cited by Burch, 2018). Furthermore, the age group between adolescence and early adulthood appears to be particularly exposed to online hate speech (Bautista-Ortuño et al., 2018; Hawdon et al., 2017; al Serhan & Elareshi, 2019). Indeed, this audience possesses two characteristics that make it specific: their substantial presence on social media and their search for identity (Costello et Hawdon, 2020; Hawdon et al., 2017; Keipi et al., 2017). Therefore, if we want to communicate about online hate speech among young people aged 15 to 25, whether in terms of prevention or intervention, it becomes necessary to **establish a common foundation of understanding**. **This is why we have chosen to investigate the understanding of young people aged 15 to 25 regarding what constitutes online hate speech**, among diverse profiles in terms of gender, sexual orientation, cultural background, and self-reported status (perpetrator, victim, and/or bystander), following the guidelines of the literature (Bautista-Ortuño et al., 2018; Costello et al., 2019). Several questions have guided our research, such as: 1) is the mode of expression in the form of speech necessary? 2) is the mode of dissemination within the digital realm predominant? 3) what is the content of the expressed statements? 4) what behaviors are identified, and with what intentions are they committed? To do this, we conducted and subsequently analyzed twenty-three semi-structured interviews with young Belgians (both French-speaking and Dutch-speaking).

---

[1] The self-reported status is the status identified by the young person themselves, whether they are a bystander, victim, and/or perpetrator.

Principal results

Diversity of Participant Profiles

All 23 participants consider themselves witnesses of online hate speech. Among them, 13 also identified themselves as victims, and 3 as perpetrators, with 2 combining both statuses (perpetrator-victim). Regarding other characteristics, 10 individuals are female, and 13 are male. 13 define their sexual orientation as heterosexual, 3 as homosexual, 3 as bisexual, 1 as pansexual, and 3 as "other." Finally, 10 individuals are exclusively of Belgian origin, 1 person is of Belgian and European origin, 1 person is of Belgian and African origin, 4 are of European origin, and 7 are of African origin[2]. For the presentation of the results, similar to the gender variable, we have chosen to dichotomize certain variables, such as sexual orientation (heterosexual, n=13 and non-heterosexual, n=10), cultural origin (exclusively of European origin, n=15 and non-European, n=8), and age (15-17 years; n=9 and 18-25 years, n=14).

Different nuances of online hate speech

During the analysis of the definitions provided by the young individuals interviewed, two categories of definitions emerged. On one hand, there was a definition, which we consider quite broad, referring to behaviors such as "insults," "jokes," "harassment," "verbal aggression," or "cyber violence," associated with a varied intention (joking, intent to harm, etc.). On the other hand, there was a definition perceived as stricter, encompassing behaviors such as "incitement to violence," "exclusion," "promotion of racism," or "discrimination," combined with a more pronounced intention against the other (pure hatred). The first category of definition, which we named "**aggressive message**," was addressed by thirteen participants and is characterized by aggressive expression towards an individual or a representative of a group. The second category, named "**hateful message**," was mentioned by ten participants and aims at hateful incitement towards a representative of a group or an entire group. Within these two categories, explicit mention of the digital realm (social networks, cyber violence, online, etc.) was found, although more prominently within the "aggressive message" category.

	Aggressive message (n=13)	Hateful message (n=10)
Words Reiterated Within Participants' Formulated Definitions **(n=23)**	Conscious, deliberate, or impulsive actions; Insults; Harmful discourse; Cyber violence; Harassment; Feeling less safe; Contrary to freedom of speech; Inflicting harm, insulting; Something more prevalent on social media; Online hateful content spilling into offline interactions; Cyber violence, innocent jokes, undertones; Form of harassment; Harassment, aggression; Unpleasant, intent to hurt.	Clear hateful thoughts against someone; Discrimination; Expressing their hatred, attempting to persuade people with hatred; Calling for racism; Holding hatred towards others and displaying it on social media; Speech that incites hatred; Inciting violence, whether physical or verbal; Speech containing racism or gender-based negativity; Excluding individuals from minority groups; Pure hatred, non-acceptance, discrimination, ignorance.

Regarding the diversity within the sample, we observe that those with non-heterosexual orientation, female gender, and the age group 18-25 are more inclined to formulate a definition falling under the "aggressive message" category. Within this group, we have 12 participants who reported a victim status. It is also notable that for gender and sexual orientation, their counterparts (male and/or heterosexual) were predominantly found in the "hateful message" category of definition. Cultural background is the only variable where respondents appear to be fairly evenly distributed between these two categories of online hate speech, both for young Europeans and non-Europeans.

Conclusion and recommendations

Given the diversity within our sample of Belgian youth aged 15 to 25, various definitions of online hate speech have been proposed and classified into two categories: "aggressive message" and "hateful message." Examining our research questions, it is evident that the **content of the discourse** characterized by aggression is present among our participants. The presence of **aggression (whether intentional or not),** appears to be a distinguishing factor for both categories and could be considered a defining characteristic of online hate speech. Interestingly, distinct attributes of online hate speech emerged within both categories. In the "hateful message" category, the **use of speech as a means of expression** is emphasized, albeit with potent consequences

---

[2] The birthplace of the parents was selected by the research team to identify the respondents' origin.

(exclusion, discrimination, incitement to violence). Additionally, **the digital context** is mentioned, albeit in a secondary manner, with the focus being on behaviors and intentions. This perception of online hate speech seems to be more prevalent among male and/or heterosexual participants in our sample. These **individual characteristics** align with those of online hate speech perpetrators (Bernatzky et al., 2022), even though only one participant in our study identified as a perpetrator within the "hateful message" category.

Within the "aggressive message" category, definitions focus on **verbal violence** more broadly, incorporating the necessary digital factor and potential repetition, akin to **harassment and cyber-harassment behaviors**. This category is primarily composed of females, individuals aged 18 to 25, and/or those identifying as non-heterosexual; characteristics often associated with individuals experiencing online victimization (Costello & Hawdon, 2020; Reichelmann et al., 2021).

To better illustrate these definitions and developed categories, we have established **a continuum of behaviors associated with the various nuances of online hate speech**, ranging from micro-aggressions (online or offline). Micro-aggressions refer to intentional and unintentional daily behaviors and language experienced by minority and/or oppressed individuals (Clark et al., 2011; Constantine, 2007; Sue et al., 2008), extending to violent acts committed in public spaces (see Figure 1). Our results underscore the importance of not overlooking micro-aggressive and harassment behaviors, as they can serve as fertile ground for increasingly violent behaviors, both online and offline (Cramer et al., 2022; Schweppe & Perry, 2022; al Serhan & Elareshi, 2019; Anti-Defamation League, 2018).

Connected to these initial findings, we can make several **recommendations**:

- The necessity to **communicate the various forms and nuances of online hate speech**, particularly to the adolescent and emerging adult population (15-25 years old). Our results illustrate that all 23 participants are at least witnesses to online hate speech. Therefore, it seems essential to explain and inform about the presence of aggressive and hateful messages online, independent of their classification. Indeed, researchers advise against adhering to an overly restrictive vision of online hate speech (Perry, 2001; Schweppe & Perry, 2022), which is also reflected in the definitions provided by our respondents. Hence, continuing to **hear the voices of individuals aged 15 to 25** on this subject, incorporating a **diversity of profiles** as established in this research, is important.

- Within the @ntidote project, understanding online hate speech can be facilitated through an understanding of **micro-aggressions**. Legally, micro-aggressions can sometimes be considered by the law, depending on the harm and context (e.g., workplace), but are often seen as non-criminal. However, some international groups advocate for their criminalization (Schweppe & Perry, 2022). The need for reflection on the treatment of these micro-aggressions seems undeniable, both for prevention and education, as well as for a potential qualification within the Criminal code and the possibility for victims to file complaints. For example, **better communication about these micro-aggressions** and the current qualifications within the criminal code would be necessary to disseminate among this audience, especially when **preventive actions** are taken regarding online hate speech. Indeed, daily online victimization could contribute to the choice of a broader formulation of online hate speech (qualified as a message with an aggressive character), contributing to a form of amplification of the phenomenon. Regarding victimization, we know that **subjective perception** is an essential factor in understanding the harm caused, beyond more objective indicators (Chetty & Alathur, 2018; Tontodimamma et al., 2021). However, the legal approach prefers predictability and normativity over considering subjectivity. Therefore, a balance seems to be found and, above all, communicated to enable victims to be heard and supported at the same time.

- It is intriguing to delve more deeply into **the role of the digital** realm in understanding cyber violence, such as online hate speech. Is the digital realm a hosting platform that accommodates aggressive behaviors present in public spaces, or does it amplify or even initiate these aggressive behaviors? Our results regarding the different nuances of online hate speech suggest that both possibilities exist. Therefore, delving into scientific research on the presence of moderation norms regarding cyber violence and the various nuances of online hate speech, including micro-aggressions, would be relevant.

Figure 1 : Illustration du continuum du discours de haine en ligne selon nos répondants
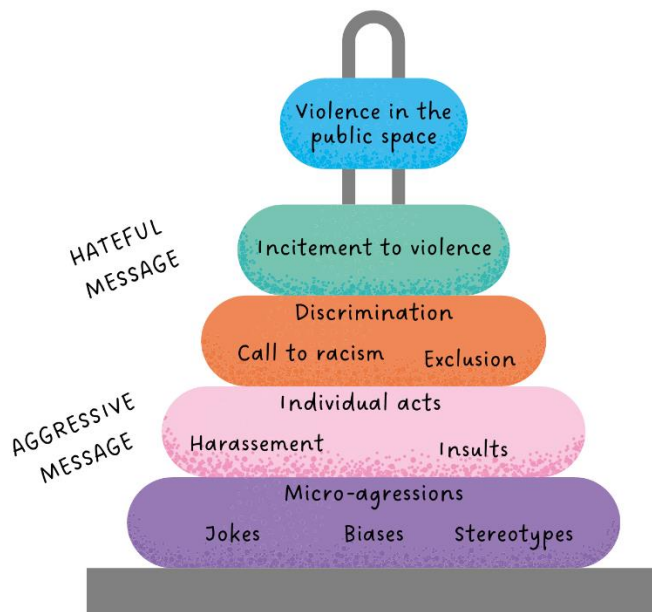
En savoir plus

al Serhan, F. & Elareshi, M. (2019). University Students' Awareness of Social Media Use and Hate Speech in Jordan. *International Journal of Cyber Criminology*, *13*(2), 548–563. https://doi.org/10.5281/zenodo.3709236

Anti-Defamation League (2018). *The Pyramid of hate*. Retrieved from https://www.adl.org/sites/default/files/documents/pyramid-of-hate.pdf .

Bautista-Ortuño, R., Perea García, J., Rodríguez Gómez, N. & Castro Toledo, F. (2018). "May I offend you?" An experimental study on perceived offensiveness in online violent communication and hate speech. *International e-journal of criminal sciences*, 12.

Bernatzky, C., Costello, M., & Hawdon, J. (2022). Who Produces Online Hate?: An Examination of the Effects of Self-Control, Social Structure, & Social Learning. *American Journal of Criminal Justice, 47*(3), 421–440. https://doi.org/10.1007/s12103-020-09597-3

Burch, L. (2018). "You are a parasite on the productive classes": online disablist hate speech in austere times. *Disability & Society,* 33(3), 392–415. https://doi.org/10.1080/09687599.2017.1411250

Castaño-Pulgarín, S., Suárez-Betancur, N., Vega, L. & López, H. (2021). Internet, social media and online hate speech. Systematic review. *Aggression and Violent Behavior*, *58*, 101608. https://doi.org/10.1016/j.avb.2021.101608

Chetty, N. & Alathur, S. (2018). Hate speech review in the context of online social networks. *Aggression and Violent Behavior*, *40*, 108–118. https://doi.org/10.1016/j.avb.2018.05.003

Costello, M. & Hawdon, J. (2020). Hate Speech in Online Spaces. In *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (p.1397–1416). *Springer International Publishing*. https://doi.org/10.1007/978-3-319-78440-3_60

Cramer, R. J., Cacace, S. C., Sorby, M., Adrian, M. E., Kehn, A., & Wilsey, C. N. (2022). A Psychometric Investigation of the Hate-Motivated Behavior Checklist. *Journal of Interpersonal Violence, 38*(7-8), 5638-5660. https://doi.org/10.1177/08862605221127196

Hawdon, J., Oksanen, A. & Räsänen, P. (2017). *Exposure to Online Hate in Four Nations: A Cross-National Consideration. Deviant Behavior*, 38(3), 254–266. https://doi.org/10.1080/01639625.2016.1196985

Keipi, T., Näsi, M., Oksanen, A., & Räsänen, P. (2017). *Online hate and harmful content: Cross-national perspectives*. London: Routledge.

Reichelmann, A., Hawdon, J., Costello, M., Ryan, J., Blaya, C., Llorent, V., Oksanen, A., Räsänen, P., & Zych, I. (2021). Hate Knows No Boundaries: Online Hate in Six Nations. *Deviant Behavior, 42*(9), 1100–1111. https://doi.org/10.1080/01639625.2020.1722337

Schweppe, J., & Perry, B. (2022). A continuum of hate: delimiting the field of hate studies. *Crime, Law, and Social Change, 77*(5), 503–528. https://doi.org/10.1007/s10611-021-09978-7

Tontodimamma, A., Nissi, E., Sarra, A., & Fontanella, L. (2021). Thirty years of research into hate speech: topics of interest and their evolution. Scientometrics, 126(1), 157–179. https://doi.org/10.1007/s11192-020-03737-6

Contact details

Contact

Name, First Name: Gangi, Océane and Mathys, Cécile

Institution/Department: Department of Criminology, ULiege

Email: oceane.gangi@uliege.be cecile.mathys@uliege.be

The @ntidote team:

Michel Walrave, University of Antwerp

Vanessa Franssen, University of Liège.

Cécile Mathys, University of Liège.

Catherine Van de Heyning, University of Antwerp.

Jogchum Vrielink, Saint-Louis University Brussels

Mona Giacometti, University of Antwerp

Aurélie Gilen, University of Antwerp

Océane Gangi, University of Liège