Postmodern Openings
ISSN: 2068-0236 | e-ISSN: 2069-9387
Covered in: Web of Science (WOS); EBSCO; ERIH+; Google Scholar; Index Copernicus; Ideas RePeC; Econpapers; Socionet; CEEOL; Ulrich ProQuest; Cabell, Journalseek; Scipio; Philpapers; SHERPA/RoMEO repositories; KVK; WorldCat; CrossRef; CrossCheck

2020, Volume 11, Issue 2 Supl.1, pages: 107-124 | https://doi.org/10.18662/po/11.2Sup1/182

A Moderated Mediation Effect of Online Time Spent on Internet Content Awareness, Perceived Online Hate Speech and Helping Attitudes Disposal of Bystanders

Dana RAD[1],
Edgar DEMETER[2]

[1] Assoc.Prof.PhD, Aurel Vlaicu University of Arad, Arad, Romania, dana@xhouse.ro

[2] Assist.Prof.PhD, Aurel Vlaicu University of Arad, Arad, Romania, eddemeter@yahoo.com

Abstract: Digital counter narratives have proved to be a consistent support to victims of online speech as they are defined as online narrative activities and actions, mostly verbal that have as effect the lowering of the negative psychological distress caused by hate speech (HS) over their victims. Internet content awareness (ICA) has an indirect significant effect on the relationship between perceived online HS and helping attitudes disposal towards the victim of HS, moderated by the online time spent by youth. Does internet content awareness (ICA) influence the helping attitudes disposal of bystanders in an online hate speech situation? Is the indirect effect of ICA mediation significant in triggering the helping attitudes towards the victims? Does the online time spent by youth moderates this relationship? These are the questions proposed by this research funded by Erasmus+ project Hate's Journey. Our research team has designed an online questionnaire addressing 206 youth from Latvia, Turkey, Spain and Romania. We have used a moderated mediation model (Model 7 from Process Hayes). The key finding of this research is that in an online hate speech situation, the more aware a person is regarding the online content, the more helping attitudes will show towards the victim of HS, under the conditioning of less and medium time spent online. Conclusions and discussions will focus on the argumentation of results with regards to emergent behaviour paradigm.

Keywords: *perceived online hate speech; moderated mediation; emergent behaviour; online time spent; internet content awareness; helping attitudes disposal of bystanders.*

How to cite: Rad, D., & Demeter, E. (2020). A Moderated Mediation Effect of Online Time Spent on Internet Content Awareness, Perceived Online Hate Speech and Helping Attitudes Disposal of Bystanders. *Postmodern Openings, 11*(2Sup1), 107-124. https://doi.org/10.18662/po/11.2Sup1/182

1. Introduction

1.1. Defining Emergent Behaviour

Emergent behavior or emerging property may arise when a variety of basic entities (agents) function in an environment that shapes more complex behaviors than a whole (Anthony, 2004). This complex behavior is not the sum of the individual behaviors and cannot be deduced or predicted based on these individual behaviors of the low level entities. Examples of emergent behavior include bird flocks or fish banks.

The basic principle of emergent behavior is the one in which the simple behavior of the individual components creates a behavior that is much more complex than can be estimated of the system. These emergent properties are completely unpredictable, without precedent. In many cases the manifestation of emergent behavior represents a new phase of evolution of the system of entities. Another explanation that may explain the proliferation of emerging behaviors is that the number of associations among individual entities grows with their number exponentially. This fact gives rise to the possibility of new subtle behaviors that cannot be deduced or predicted based on a rigorous mathematical apparatus due to the extremely large number of possibilities (Anthony, 2004). However, a large number of interactions between entities does not individually explain the emergency. There were identified situations in which the large number of entities did not lead to an emergency but, on the contrary, resulted in the complete elimination of the emergent behavior from the system. At the same time, the large number of entities produce a proportional noise level that hides the useful signal until it is eliminated. This shows that not only the large number of entities in the system or the large number of interconnections made between them matters, but also the way in which these connections are organized: hierarchically, centralized, decentralized, etc. It seems, however, that emergent behavior appears especially in the case of highly decentralized structures.

Emergent behaviors may be either helpful, positive or potentially detrimental, but they are very difficult to anticipate in either case before they manifest themselves (Macaulay, 2017). Emergent behaviors are often commonly viewed as more dynamic processes than simply the sum of components. An emergent behavior is inherently new and created by combining of two or more distinct events, which taken independently to not show the behavior (Macaulay, 2017).

Although research on emergent norm theory flourished since 1987, not much attention has been given to internal, intra-individual emergent

norms that might occur under tensed situations and guide the person towards automatic and unaware behaviors, especially in the digital realm.

Emergent norm theory hypothesizes that as a consequence of the development of new behavioral norms in reaction to a rushing crisis, non-traditional conduct like the one correlated with collective conduct emerges in groups. Collective action encompasses all sorts of social actions for advocates of emergent norm theory, in which traditional norms fail to work as maps of social conduct and likewise individuals collectively abolish or move beyond the usual hierarchical processes and structures of society (Turner & Killian, 1987), and new protocols will then be part of collective conduct. The fundamental principles of emerging norm theory are that collective conduct is logical, that collective conduct is an outcome to a rushing uncertain occurrence, and that new behavioral standards relevant to the collective action scenario are emerging by group mechanisms without adequate planning or training. The emerging norm theory developed out of two major patterns, initially introduced by Turner and Killian (1972). The Le Bonian practice of envisaging about groups as individuals without rules and collective behavior as arbitrary acts specifically triggered Turner and Killian to reflect on how rules are enforced in crowds. Secondly, the symbolic interactionist methods and analysis of small-groups led to the development of an interaction-designed model of norms.

Emergent norm theory means that as a problem occurs, masses gather together that force people to give up their former expectations of permissible behavior and find new ways of behaving. There is no simple law for controlling crowd behavior when a crowd is formed, and there is no leader. Yet the audience is concentrated on those who behave differently, and this differentiation is further taken on as the modern norm for group behavior. Like the new standard continues to be enforced within the crowd, the pressure for obedience and deviance amongst the public is increasing and opposition is suppressed (Aguirre et al., 1998). The silencing of opposing opinions leads to a sense of unanimity on the community.

The norms set in crowds aren't strict codes of behavior. Rather, they are more like hierarchical behavioral mechanisms that impose boundaries on reasonable issues (Turner & Killian, 1987). Such standards are borne through either emergent or preexisting social connections. Researchers Turner and Killian suggest that whatever enhances contact among group participants is fostering the development of standards, naming it milling. However, although the current viewpoint of emergent norm theory opposes the notion that crowd activity is especially irrational, it suggests that certain

group participants are predictive and contribute to the propagation of new emerging norms.

Literature depicts two major sources of criticism that challenged the emergent norm theory. Firstly, Reicher (1987) concludes that crowds when coming together they carry along standards with them, current standards therefore don't need to arise. Such norms differ in dependence with group characteristics, such differences reflecting the various ways crowds behave, while still being norms.

The second direction of critique suggests, firstly, that most of social interactions result in the mitigation of social norms and, secondly, that creativity in the development and conduct of norms that developed as a norm is created usually not by activity, but involving long-term logical planning structures or based on slight differences in established repertoires (Tilly 1993).

Couch (1968) is often cited for supporting this criticism while concluding before the designing of emergent norm theory. Some researchers also suggest that following the emergence of a norm in a group presents significant methodological difficulties.

Though emergent norm theory has been initially employed to a myriad of types of social behavior, it is more generally used to facilitate the understanding of group and crowd behavioural trends. Particularly, emergent norm theory has played a central role in catastrophe science, being employed in understanding of the actions of societies facing a rushing crisis, a tragedy, and then being compelled to discover novel pathways of responding that can facilitate ensure that as many people as possible are safe and alive.

Johnson (1987) concluded that emergent norm theory may bring evidence to the often seen violent and selfish actions in mass panics. Johnson suggests the collapse of social order in some cases contributes to such forms of actions as reasonable reactions to the current social circumstances.

1.2. Digital correlates

The Internet of Things describes a network of "smart" physical interconnected objects, which have embedded the technology needed to be able to notify and communicate data about their internal status as well as interact with it and with the external environment. The current IoT explosion has the potential to substantially change the industrial activity and of course, impact the way people work, interact and communicate (Citron & Norton, 2011).

Digital literacy is defined as the skills, knowledge and understanding needed to use new technologies and media to create and distribute meaning, being the true language of the 21st century (Kiesler et al., 1984).

In the current information society, adolescents grow up in a media saturated environment, the presence of computers, the internet and others digital technologies in homes being commonplace, mediated communication thus becoming part of daily activities (Buckingham, 2003). Youth develop different lifestyles from those of parents, with specific habits regarding the use of social media platform and in general the digital environment, being often more efficient in their use than adults.

At the same time, researchers are drawing attention to digital gaps existing between youth of different ages (Neuman & Celano, 2006). If at first phase these inequalities were thought in terms of access to the respective technology, in the second phase, with the proliferation of technologies and improving access, the emphasis has been shifted to inequalities in terms of skills (Goldin & Katz, 2008; Johnson, 1997). Buckingham (2008) appreciates the use of new ones media technologies by young people are no longer in the field of access, but more than that of social and cultural practices, because digital media offers new forms of mediation and representation of the world, respectively alternative channels for communication and interplay.

Livingstone (2004) considers that the term literacy applies to the whole media and refers to the understanding and interpretation of all texts, images and of video materials, whatever their level of complexity and whatever the channel through which these materials are broadcast. At the same time, Livingstone points out that talking about literacy involves understanding the relationship between three processes that are culturally and historically conditioned:

1. the process of symbolic meaning of culture, knowledge, and values,

2. the stratified distribution of interpretative competencies in population,

3. institutional and ideological use of competences to support social control. When applying the concept of literacy on digital media, we are witnessing the emergence of new still unanswered questions. Thus, it is difficult to define the processes of symbolic representation when the scope of digital media is unclear.

Also, competences and skills do not develop independently, but are linked to technology and media characteristics, being the result of interactions between user and media concerned. The third process, that of

institutional use of digital skills, also cause controversy: optimists see in the presence of digital competences the way of democratization and empowerment of people, while pessimists see in the same process another source of socio-cultural inequality (Potter, 2004).

Researchers Leets and Giles (1999) find that hate speech is a sub-category of aggressive discourse, the last concept being conceptualized as utterances intended to inflict damage, and regardless of intention, which their target group perceive to result in harm (Leets & Giles, 1999). Digital hate speech is disseminated via sites, social media, video gaming and other electronic networks.

Over the internet, people can more quickly scan for possible audience members who intend to express hate speech, raising their impact in the media and physical world (Duffy, 2003). Therefore, online hate speech may have tangible and indirect impacts on the psychological well-being of people in the short and long term, with the amount of harm being significantly higher for victims than for bystanders (Oksanen et al., 2014).

Bystanders are the ones hearing cases of hate speech. They can neglect or prefer not to get involved which may reinforce the victim's feeling of helplessness and alienation, or motivate the attacker. Therefore, numerous efforts have been aimed at motivating people to take a stance, for example by counter narratives or otherwise defending the victims.

It promotes hate speech normalization, with victims and bystanders adopting a passive mentality, not involving or acting. In this context, an important observation was raised by one group of researchers regarding bystanders not taking action because they feel strongly related to the target of hate speech (Bellmore et al., 2012; Barchia & Bussey, 2011), alluding to the influence empathy may play in affect (Rad et al., 2019).

There have been widely described three diverse forms of reaction to digital hate speech. One of the most commonly discussed remedy was legislation, calling the monitoring of hate speech on websites and social media channels, thus providing consumers the right to report offenses to law enforcement. Whereas, the various forms of internet and social platforms providers have a central role, setting specific standards and protocols, appropriate monitoring and notification processes, and rethinking some of their network functionality, according to youth. Another approach is to build centralized and informal environments of assistance for victims. It may look like of online educational programs, interpersonal counselling, social environments and youth groups, for example, by even giving help to haters, seeking to better understand their motivations and resolve them. There is a clear need to understand the importance of improving human ties

within the society and ensuring that people respond when acts of hate speech occur. Ultimately, awareness raising and educational campaigns are relevant, preferably from early childhood on, while at the same time referring to the position of parents. As we discussed, a lot of work is required to fully grasp the essence of online hate speech, and how this impacts the everyday online media experience of teenagers.

2. Research methodology

2.1. Objectives and hypothesis

Does the perceived online hate speech influence the helping attitudes disposal (HAS) of bystanders in an online hate speech situation? Is the indirect effect of ICA mediation significant in triggering the helping attitudes? Does the online time spent by youth moderates this relationship, meaning that decreased online time spent positively moderates the relationship, either average and increased online time spent positively moderates the relationship?

A moderated mediation appears when the mediating process is moderated because W (online time spent) alters the relationship between IV (perceived online hate speech) and M (internet content awareness); conceptually, moderated mediation is when an indirect effect is moderated, that is, the indirect effect varies across the three levels of the moderator: low, medium, and high values. Hence how is the moderating effect manifested? In order to find out, we will analyze the level of helping attitudes disposal in relation to the higher and lower values of online time spent.

2.2. Participants

Our research targeted a pool of 206 participants from Romania (24.8%), Latvia (24.8%), Spain (24.8%), and Turkey (25.7%), with an average age of 30 years, male respondents (39.8%) and female respondents (60.2%), with an educational level of 3.9% - elementary school, 1.9% - technical school, 29.1% - high school, 32% - Bachelor, 29.1% - Master and 3.9% - PhD level. Unemployed respondents represent 5.8%, students represent 43.7%, volunteers represent 1% and employed are 49.5% (Rad, D., Dixon, D., & Rad, G., 2020).

Our research utilized convenience sampling technique, as the purpose of this investigation is explorative. The total of participants were consecutively selected according to the order of appearance when completing the online questionnaire shared on social media platforms by

each of the 4 project partner countries, each country targeting at least 50 respondents, according to the convenient accessibility principle. Data collection procedure was coordinated by 4 partner organizations: Asociación Cultural Social y Educativa Segundas Oportunidades (Spain), Aurel Vlaicu University of Arad (Romania), Ucarli Genclik Dernegi (Turkey) and Young Folks (Latvia). The sampling process ended by the time each of the 4 project partner countries reached their sample saturation of 50 valid respondents and the 3 month time saturation, during June – August 2019 (Rad et al., 2020).

2.3. Instruments

For the purpose of this research, we have included in our online investigation the following instruments. For assessing youth's helping attitudes disposal towards the victims of hate speech, we have used the Helping Attitudes Scale (HAS), as presented by Nickell (1998). HAS consists of 20-item measure of research participants' beliefs, emotions, and conducts correlated with helping, interactions with others. Each of the 20 items are answered using a 5 points Likert scale, from 1 - strongly disagree to 5 - strongly agree. Items 1, 5, 8, 11, 18, 19 must be scored in reverse, and all 20 scores are then added in the overall score, with ranges between 20 and 100. A 60 is, as the researcher suggests, a neutral score (Nickell, 1998). As resulted from research on giving, the generosity characteristic increased with age, education, trust, income, and prosocial value orientation (Bekkers, 2007) and most conclusions were given according to the integrative model of social value orientation (De Cremer & Van Lange, 2001). Referring to HAS scale statistics, we have obtained a mean m=76.49, a variance of 138.573 and a standard deviation of SD=11.77. Scale's reliability was also tested, and Cronbach's Alpha coefficient equals 0.894, assuring a reliability of 89% of the HAS scale.

For internet content awareness ICA we have used a four item measure – Item 25 On a one to five scale where 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree, thinking about when you use social media websites or apps such as Facebook, Instagram, Snapchat, WhatsApp, etc…, we would like to know how much you agree with the following statements:

25.1. I'm able to access the information and content I want on social media.

25.2. I understand how people create and spread messages on social media.

25.3. I understand the role social media websites/apps play in shaping the information and content I see.

25.4. I'm confident creating and sharing my own social media messages.

Referring to ICA scale statistics, we have obtained a mean m=14.89, a variance of 11.628 and a standard deviation of SD=3.41. Scale's reliability was also tested, and Cronbach's Alpha coefficient equals 0.877, assuring a reliability of 87% of the ICA scale.

The following single research items were used:

- for online time spent (M=5.82, SD=1.86 assessment this research used a single item measure – Item 6. How much time you spend online. Answers were coded as follows: Never or hardly ever=2, At least every month=3, Every week=4, Daily or almost daily=5, Several times each day=6, Almost all the time=7, I don't know/ I prefer to not answer=1. As presented in Table 1, 1% of participants declared - never or hardly ever, 8.7% - every week, 20.4% - almost daily, 46.6% - several times per day, 23.3% - and almost all the time.

Table 1. Frequencies for online time spent
Source: Authors own conception

Online time	Frequency	Percent	Valid Percent	Cumulative Percent
2	2	1.0	1.0	1.0
4	18	8.7	8.7	9.7
5	42	20.4	20.4	30.1
6	96	46.6	46.6	76.7
7	48	23.3	23.3	100.0
Total	206	100.0	100.0	

- for perceived online hate speech (M=3.88, SD=1.00) assessment this research used a single item measure – Item 13. Have you ever seen, heard or read about online hate speech situations? Answers were coded as follows: Yes=1, No=2. As we can see from Table 2, 13.6% of the respondents declared that they have perceived online hate speech and 86.4% declared that they have not perceived online hate speech.

Table 2. Frequencies for perceived online hate speech

Source: Authors own conception

Have you ever seen, heard or read about online hate speech situations?	Frequency	Percent	Valid Percent	Cumulative Percent
1 - Yes	28	13.6	13.6	13.6
2 - No	178	86.4	86.4	100.0
Total	206	100.0	100.0	

2.4. Research design

Mediating variables have an important role in contributing to explaining the psychological phenomena. Mediation represents a process in which one independent variable (X) has an influence over one dependent variable (Y). Thus, the variability of X determines a variation in one or more mediating variables (M), which causes, further on, the variation of variable Y. The simplest mediation model consists of an independent variable (X), a mediator (M) and a dependent variable (Y). Thus, we understand that the mediator, represented by a behavioral, biological, psychological or social construct that transmits to the dependent variable the effect it receives from the independent variable.

The mediation model contains: a) two dependent variables (M and Y), which receive the effect of variable X; b) two independent variables (X and M) that have an effect on the variable Y. Thus, the independent variable exerts an effect on the dependent variable through the mediating variable. The best known method of testing the mediation effect was proposed by Baron and Kenny (1986) and is based on the principles of regression analysis. The method proposed by them is known as the method of causal steps approach and aims to determine if M can be considered as an average variable of the effect of X on Y. As Baron and Kenny (1986) stated, a moderator is a quantitative variable (e.g., time spent online) that affects the direction or intensity of the association between the independent variable (VI) and the dependent variable (VD).

Moderate mediation, known as conditional indirect effects, happens where the influence of an independent variable X on an outcomes variable Y through a mediator variable M varies depending on different levels of a moderator W. More specific, either the effect of X on the Y, and/or the effect of Y on M depends on the level of the moderator W.

Thus, a moderated mediation appears when the mediating process is moderated because W alters the relationship between IV and M; conceptually, moderated mediation is when an indirect effect is moderated, that is, the indirect effect varies across the three levels of the moderator: low, medium, and high values. Hence how is the moderating effect manifested? In order to find out, we will analyze the level of helping attitudes disposal in relation to the higher and lower values of online time spent.

In order to test our moderated mediation hypothesis, we have computed in SPSS's PROCESS macro Version 3 (Hayes, 2017) the model 7, where helping attitudes disposal HAS is our independent variable Y, perceived online hate speech Item 13 is the dependent variable X, internet content awareness ICA is the mediator M and online time spent is the moderator variable W.

3. Results

According to the moderated mediation model, following Preacher et al. (2007), online time has an effect on internet content awareness which, in turn, mediates the relationship between perceived online hate speech on helping attitudes disposal. In other words, we expect that a decreased and medium online time spent will result in an increased level of internet content awareness, and the latter will lead to higher helping attitudes disposal. Aiken and West (1991) stated high and low values as +/- 1 SD standard deviation from the mean for a continuous moderating variable. Thus the impact of the mediator variable internet content awareness on the relationship between VI (perceived online hate speech) and VD (helping attitudes disposal) is different depending on the level of the moderating variable (online time spent) for the low level of online time, the effect is increased, the effect is also increased for average online time, and for increased online time, the effect is no longer significant.

Table 3. Interaction effect

Source: Authors own conception

```
OUTCOME VARIABLE: ICA internet content awareness

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .5239      .2744     8.5619    25.4671     3.0000   202.0000     .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant    15.0129      .2080    72.1846     .0000    14.6028    15.4230
Item13       2.3965      .6591     3.6362     .0004     1.0970     3.6961
Online_time  1.1745      .2181     5.3847     .0000      .7444     1.6046
```

```
Int_1       -1.9210      .6603    -2.9092      .0040    -3.2230     -.6190

Product terms key:
 Int_1    :        Item13 (perceived online hate speech)   x      Online_time

Test(s) of highest order unconditional interaction(s):
       R2-chng          F        df1        df2          p
X*W      .0304     8.4636     1.0000   202.0000      .0040
----------
    Focal predict: Item13   (X)
          Mod var: Online_t (W)

Conditional effects of the focal predictor at values of the moderator(s):

   Online_t     Effect         se          t          p       LLCI       ULCI
     -.8155     3.9632      .6673     5.9388      .0000     2.6473     5.2790
      .1845     2.0422      .7158     2.8531      .0048      .6308     3.4535
     1.1845      .1211     1.2047      .1006      .9200    -2.2543     2.4966
```

As seen in Table 3, the value of R2 change for the model with the interaction was 0.030, statistically significant [F (1,202) = 8.46; p = 0.004]. This result indicates that online time spent moderates the relationship between perceived online hate speech and helping attitudes disposal. The moderating effect is manifested by diminishing the relationship between perceived online hate speech and helping attitudes disposal as the time spent online increases, becoming no longer significant at the increased values of the online time spent moderator.

Table 4. Mediation effect

Source: Authors own conception

```
OUTCOME VARIABLE: HAS (helping attitudes disposal)

Model Summary
         R       R-sq        MSE          F        df1        df2          p
      .5520      .3047    97.2925    44.4902     2.0000   203.0000      .0000

Model
             coeff         se          t          p       LLCI       ULCI
constant   48.8392     3.3233    14.6962      .0000    42.2867    55.3918
Item13      1.1930     2.1670      .5505      .5826    -3.0796     5.4656
ICA         1.8569      .2183     8.5058      .0000     1.4265     2.2874

****************** DIRECT AND INDIRECT EFFECTS OF X ON Y *****************

Direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
     1.1930     2.1670      .5505      .5826    -3.0796     5.4656

Conditional indirect effects of X on Y:

INDIRECT EFFECT:
 Item13      ->    ICA         ->    HAS

   Online_t     Effect     BootSE   BootLLCI   BootULCI
     -.8155     7.3594     1.3728     4.6432     9.9610
      .1845     3.7922     1.7436      .2286     7.0854
     1.1845      .2250     2.8754    -5.9041     5.2651
```

```
      Index of moderated mediation:
              Index     BootSE   BootLLCI   BootULCI
Online_t    -3.5672     1.4269    -6.5893    -1.0272
```

We have then fitted a series of regression models, in the first step predicting the mediator variable, internet content awareness (ICA) using the independent variable, perceived online hate speech (step 2); then the dependent variable, helping attitudes disposal HAS using both the independent variable, perceived online hate speech and the mediator internet content awareness ICA (steps 3 and 4); and in the last step the dependent variable helping attitudes disposal HAS, using the independent variable perceived online hate speech (step 1). Thus, while the independent variable, perceived online hate speech was a significant predictor for both the dependent helping attitudes disposal HAS and the mediator variables internet content awareness ICA, it is no longer a significant predictor in the presence of the mediation variable, bringing confirmation to our mediation hypothesis. It can be concluded that internet content awareness fully mediated the relationship between perceived online hate speech and helping attitudes disposal towards the victims of hate speech.

After the regression models an estimate is also given for the indirect effect of X on Y. The effect size for decreased online time spent was 7.3594, with a 95% confidence interval which did not include zero, the effect size for medium online time spent was 3.7922, with a 95% confidence interval which did not include zero, thus the effect was significantly greater that zero at $p < .05$. However this was not accomplished when computing the effect size for increased online time spent. The effect value of .2250, with a 95% confidence interval did include zero, thus the effect was not significantly greater that zero.

Regression analysis was employed to investigate if online time spent moderated the internet content awareness mediated the relationship between perceived online hate speech and helping attitudes disposal of bystanders in hate speech situations. Table 1 and Table 2 indicate that perceived online hate speech was a significant predictor of helping attitudes disposal and internet content awareness was also a significant predictor of helping attitudes disposal, results supporting the mediational hypothesis. Perceived online hate speech (Item 13) was no longer a significant predictor of helping attitudes disposal after controlling for the mediator, internet content awareness, B = .5505, SE = 2.1670, consistent with full mediation. The indirect effect was tested using a percentile bootstrap estimation method

with 10000 samples (Shrout & Bolger, 2002), introduced with Version 3 of the PROCESS macro (Hayes, 2017).

Further online time spent moderates the relationship between perceived online hate speech and helping attitudes disposal. The moderating effect is manifested by diminishing the relationship between perceived online hate speech and helping attitudes disposal as the time spent online increases, becoming no longer significant at the increased values of the online time spent moderator.

This analysis also had weaknesses, including the use of an opportunistic sample, the simplicity of the data collection process and the limited aim of this micro-exploratory investigation, and also the single item operationalization of the perceived online hate speech construct. The adolescents in this research are all enrolled in educational programs, coming from both urban and rural areas. Thus another limitation of this research is the non-representativeness of NEETs youth. Respondents consensually completed the online questionnaire at various places, such this non-restrictive control over environmental factor may have an influence over the results. As regarding the total number of respondents, the relationship between the studied constructs might have been more significant and the effect size greater if there was a larger pool of participants.

4. Conclusions and discussions

We have fitted a series of regression models, first predicting the internet content awareness (ICA) using the perceived online hate speech (step 2); then the helping attitudes disposal HAS using both perceived online hate speech and internet content awareness ICA (steps 3 and 4); and finally helping attitudes disposal HAS, using perceived online hate speech (step 1). While perceived online hate speech was a significant predictor for both helping attitudes disposal HAS and internet content awareness ICA, it is no longer significant when controlling for the mediator variable; confirming the full mediation effect. As depicted in Table 2, online time spent moderates the relationship between perceived online hate speech and helping attitudes disposal. The moderating effect is manifested by diminishing the relationship between perceived online hate speech and helping attitudes disposal as the time spent online increases, becoming no longer significant at the increased values of the online time spent moderator.

The main purpose of our quantitative research was to balance the literature's core observations with a more detailed account of the youth's viewpoints and perceptions about online hate speech. Our research paints an

image of sophisticated teens very familiarized with the subject of hate speech and can relate it to their personal experiences with digital media. And if there is certainly a fundamental knowledge of principles such as freedom of expression and hate speech, teenagers typically struggle to express consistently what digital hate speech looks like, and the mechanism that inflicts it in the modern digital environment. Most frequently than not, there is often a convincing picture of victims, perpetrators and witnesses that appears.

Youth tend to understand some of the possible implications that hate speech online can have, both at the human and social levels. Nonetheless, as the youth claim, apart from controlling and appropriate supervision and reporting practices, increasing consciousness and education seem to be necessary, and hopefully that should start from an early age.

Education can help learners systematize their learning about the meaning, causes and effects of online violence. They will also tackle the normalization of hatred, which can induce ignorance and division. A radically new approach is required for this, one that lets youth understand and control their own feelings, attitudes and behaviors, while empathizing with peers, particularly others considered to be different.

Emerging phenomena are properties of the systems that arise from the collective interaction of their components, properties that cannot be predicted by analyzing only the structure or behavior of a small number of constituents subject to fundamental laws. Although emergent phenomena are ubiquitous in contemporary science and a significant part of today's research is focused on their understanding, their acceptance and incorporation into the structure of the thinking pattern seems to face invisible resistance in psychology exploratory research.

Reductionism, the possibility to deduce fundamental laws of human behavior in the real and digital environment through simplification, does not imply the possibility to describe the initial complexity by reconstructing it from the fundamental laws thus discovered. And maybe that interferes with our desire that the effort to understand has not been in vain. The mere fact that we can formulate these laws does not give us the right to reduce the complexity of the phenomenon of digital human aggression, in the form of hate speech. The fascination of these emergent phenomena makes them an exceptional candidate for the psychology of the 21st century, and the inclusion of their exploration within the fundamental knowledge in cognitive psychology can only broaden the horizons.

Emergent behavior in social systems is one of the most fascinating research topics of modern times, as, in my opinion, it is in this type of

phenomena that one of the solutions to the most complex problems of the 21st century lies. According to the paradigm shift, the digital sustainable of the 21st century will be those that, perceiving and integrating strategic management, their complex nature, will create conditions for positive emerging behavior.

The role of 21st century digital influencers becomes to create the necessary elements of the ecosystem, fostering and conserving networks of interdependent relationships between them "stewardship". More than imposing behaviors, it is about creating conditions for them to appear collectively and spontaneously, supporting and leveraging the small initiatives that contribute to the purpose of the digital environment, creating feedback mechanisms and independent instances that promote dialogue and participation instead of criticism and hate speech.

The times of centralized management are over and the complex systems paradigm offers innovative perspectives on how to manage complex social systems in a distributed way, thus making them more resilient. This is a process that requires time, patience but above all transparency, trust and turning the focus back to human relationships, as these are the ones that really have a sustainable impact over the digital behaviors of peers.

References

Aguirre, B. E., Wenger, D. E., & Vigo, G. (1998). A test of the emergent norm theory of collective behavior. *Sociological Forum, 13*(2), 301–320. https://doi.org/10.1023/a:1022145900928

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.

Anthony, R. J. (2004). Natural inspiration for self-adaptive systems. Database and expert systems applications. *Proceedings of 15th International Workshop*: 732 – 736. https://ieeexplore.ieee.org/document/1333561

Barchia, K., & Bussey, K. (2011). Predictors of student defenders or peer aggression victims: Empathy and social cognitive factors. *International Journal of Behavioral Development, 35*(4), 289–297. https://doi.org/10.1177/0165025410396746

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Bekkers, R. (2007). Measuring altruistic behavior in surveys: The all-or-nothing dictator game. *Survey Research Methods*, *1*, 139-144. https://ojs.ub.uni-konstanz.de/srm/article/view/54

Bellmore, A., Ma, T. L., You, J. I., & Huhges, M. (2012). A two-method investigation of early adolescents' responses upon witnessing peer victimization in school. *Journal of Adolescence, 35*(5), 1265-1276. https://doi.org/10.1016/j.adolescence.2012.04.012

Buckingham, D. (2003). *Media education. Literacy, learning and contemporary culture*. Polity Press.

Buckingham, D. (2008). Defining digital literacy – What do young people need to know about digital media? In C. Lankshear & M. Knobel (Eds.), *Digital literacies* (pp. 73–91). Peter-Lang.

Citron, D. K., & Norton, H. (2011). Intermediaries and hate speech: Fostering digital citizenship for our information age. *Boston University Law Review, 91*, 1435-1483. https://digitalcommons.law.umaryland.edu/fac_pubs/1056/

Couch, C. J. (1968) Collective behavior: An examination of some stereotypes. *Social Problems*, *15*(3), 310–322. https://doi.org/10.1525/sp.1968.15.3.03a00040

De Cremer, D., & Van Lange, P. A. M. (2001). Why prosocials exhibit greater cooperation than proselfs: The roles of social responsibility and reciprocity. *European Journal of Personality*, *15*(S1), 5-18. https://doi.org/10.1002/per.418

Duffy, M. (2003). Web of hate: A fantasy theme analysis of the rhetorical vision of hate groups online. *Journal of Communication Inquiry*, *27*(3), 291-312. https://doi.org/10.1177/0196859903252850

Goldin, C., & Katz, L. (2008). *The race between education and technology*. Harvard University Press.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

Johnson, G. (1997). Changes in earnings inequality: The role of demand shifts. *The Journal of Economic Perspectives*, *11*(2), 41-54. https://doi.org/10.1257/jep.11.2.41

Johnson, N. R. (1987). Panic and the breakdown of social order: Popular myth, social theory, empirical evidence. *Sociological Focus, 20*(3), 171–183. https://doi.org/10.1080/00380237.1987.10570950

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, *39*(10), 1123-1134. https://doi.org/10.1037/0003-066x.39.10.1123

Leets, L., & Giles, H. (1999). Harmful speech in intergroup encounters: An organizational framework for communication research. In M. Roloff (Ed.), *Communication yearbook*, *22* (pp. 91–37). Sage Publications.

Livingstone, S. (2004). Media literacy and the challenge of new information and communication technologies. *The Communication Review*, *7*(1), 3–14. https://doi.org/10.1080/10714420490280152

Macaulay, T. (2017). Threats and impacts to the IoT. *RIoT Control Understanding and Managing Risks and the Internet of Things*, 221-278. https://doi.org/10.1016/B978-0-12-419971-2.00012-1

Neuman, S. B., & Celano, D. (2006). The knowledge gap. Implications of leveling the playing field for low-income and middle-income children. *Reading Research Quaterly*, *41*(2), 176–201. https://doi.org/10.1598/rrq.41.2.2

Nickell, G. (1998). *The helping attitudes scale*. Paper presented at 106th Annual Convention of the American Psychological Association at San Francisco, August, 1998. http://web.mnstate.edu/nickell/APA%201998%20paper.pdf

Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M., & Räsänen, P. (2014). Exposure to online hate among young social media users. *Sociological Studies of Children & Youth*, *18*(1), 253-273. https://doi.org/10.1108/s1537-466120140000018021

Potter, W. J. (2004). *Theory of media literacy: A cognitive approach*. Sage.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, 42(1), 185-227. https://doi.org/10.1080/00273170701341316

Rad, D., Dixon, D., & Rad, G. (2020). Digital outing confidence as a mediator in the digital behaviour regulation and internet content awareness relationship. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, *11*(1), 84-95. https://doi.org/10.18662/brain/11.1/16

Rad, D., Dughi, T., Roman, A., & Ignat, S. (2019). Perspectives of consent silence in cyberbullying. *Postmodern Openings*, *10*(2), 76-84. https://doi.org/10.18662/po/71

Reicher, S. D. (1987). Crowd behaviour as social action. In J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher & M. S. Wetherell (Eds.), *Rediscovering the social group: A self-categorization theory*. Blackwell.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and non-experimental studies: New procedures and recommendations. *Psychological Methods*, *7*(4), 422. https://doi.org/10.1037/1082-989x.7.4.422

Tilly, C. (1993). Contentious repertoires in Great Britain, 1758–1834. *Social Science History, 17*(1), 253–280. https://doi.org/10.2307/1171282

Turner, R. H., & Killian L. M. (1987). *Collective behavior* (3rd ed.). Prentice-Hall.

Turner, R. H., & Killian, L. M. (1972). *Collective behavior* (2nd ed.). Prentice-Hall.