
ANNALES
UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA
LUBLIN – POLONIA

VOL. LVI, 5

SECTIO H

2022

ZUZANNA KRAUS

zuzannakraus@gmail.com

SYLWIA KUCZAMER-KŁOPOTOWSKA

sylwia.klopowska@ug.edu.pl

University of Gdańsk. Faculty of Management

101 Armii Krajowej St., 81-824 Sopot, Poland

ORCID ID: <https://orcid.org/0000-0003-4781-0118>

MARIUSZ CHMIELEWSKI

mariusz.chmielewski@ug.edu.pl

University of Gdańsk. Faculty of Management

101 Armii Krajowej St., 81-824 Sopot, Poland

ORCID ID: <https://orcid.org/0000-0002-0775-621X>

*Impact of Hate Speech on Building the Value of Pro-Social
Activities. Perspective of Author and Social Media User*

Keywords: social marketing; social campaigns; social media; social media influencers; hate speech; sources of competitive advantage; value of pro-social activity; CSR

JEL: M14; M37; M21

How to quote this paper: Kraus, Z., Kuczamer-Kłopotowska, S., & Chmielewski, M. (2022). Impact of Hate Speech on Building the Value of Pro-Social Activities. Perspective of Author and Social Media User. *Annales Universitatis Mariae Curie-Skłodowska, sectio H – Oeconomia*, Vol. 56, No. 5.

Abstract

Theoretical background: The value of a given project (including those of a pro-social nature) is determined by a number of factors, not only those of a financial nature. Within the corporate social responsibility concept, ethics was indicated as one of the factors influencing the value of the project. Currently, one of the problems

in the field of ethics is the widespread online hate speech, affecting also the effectiveness of the processes of building the value of pro-social activities that fit into the concept of social marketing. Hate speech has come to form a significant portion of online communication. Only in the first quarter of 2022, Facebook removed 15.1 million pieces of content with hate speech, constituting only a minor part of the problem.

Purpose of the article: The authors have noted a research gap in the area of analysing the impact of hate speech on building the value of pro-social activities. The intention of the authors was to evaluate whether the analysis of the phenomenon in question will allow to verify whether there is a relationship between online hate speech and building the value of online pro-social activities. The following research question was formulated: Can online hate speech have an impact on building the value of pro-social activities?

Research methods: The authors decided that the quality dimension of the conducted studies should have the form of an individual in-depth interview (the author's perspective) and a focus group interview supplemented with projection techniques – connotation test, evaluation of emotions and level of controversy (the perspective of the user of social media).

Main findings: Hate actions significantly impede the functioning of profile with pro-social content which is meant to function as the so-called safe space, which may have an impact on building the value of the influencer's pro-social activities. During the interview with the recipients of controversial social campaigns, the occurrence of contradictory emotions was confirmed. The paper may offer a basis for further deepened scientific studies in the area of building the value of pro-social activities and a guideline for authors of social activities. Further research in the discussed direction may indicate new modes of using the social media with the participation of influencers who are active in the pro-social area and who have a community of engaged followers.

Introduction

According to the concept of building enterprise value (Rappaport, 1999), a company's competitive advantage is one of the six main value generators. Value creation in a company is realized through all activities undertaken in the field of business, which applies not only to the operational activities of a company but also to various types of projects, such as shaping an image of a company. Competitive advantage can be developed by business entities with a variety of tools. Typically, significant attention is focused on financial tools, however, in recent years attention has also been paid to the use of ethics as an area that can provide a competitive advantage for an organization over other entities (Kazojć, 2014). It seems that in the case of a company's competitive advantage, the set of tools influencing it is very broad.

It should be noted that like other factors from different areas, ethics can act not only as a factor that increases the value of the project but also as a factor that decreases the value of the project, including pro-social activities. The issue of ethics is very much a part of the concept of social responsibility, the use of which affects the building of the value of any projects (Adamczyk, 2009). What is important, in recent years, hate speech has become one of the main ethical issues of modern media (Białek-Szwed, 2018). Hate speech can be used by competitors acting unethically as a destroyer of corporate value. From the point of view of building enterprise value, it is crucial to identify hate speech activities and to determine the impact of this factor on perceptions of the company and its operations.

Online hate speech is a widely noticeable phenomenon without one universally accepted definition. Nevertheless, the concept of hate speech as such is not a new phenomenon, yet the development of the Internet and changes in user behaviour have extended its definition and now it comprises multiple speech acts and other ill and socially unaccepted behaviours online and offline (Assimakopoulos et al., 2017; Laaksonen et al., 2020). The Internet and social media have come to be an ideal place for the growth of online hate speech, which is considered a serious problem and forms the subject matter of an international public debate and initiatives aimed at working out solutions that aim to limit the distribution of online hate speech, which is a problematic task (Kovács et al., 2021; Tontodimamma et al., 2021). On the other hand, one of the biggest social media platforms, Facebook, removed 15.1 million pieces of content containing hate speech, only in the first quarter of 2022 (Statista, 2022). In 2021, the result for the entire year was 96.4 million pieces of deleted hate speech content, which highlights the scale of the discussed problem but only partially since Facebook is only one of many places online where hate speech is noticeable (Statista, 2022). Facebook, as a platform that deals with online hate speech every day, defines it as a direct attack against people according to their: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease (Facebook, 2021).

Hate speech is analysed by the researchers in various areas, such as: exposure to on-line hate among children/young social media users (Alshamrani et al., 2021; Ananthakrishnan & Tucker, 2021; Oksanen et al., 2014), or in the problematic context of hate speech detection (Arango et al., 2019; Kovács et al., 2021; Watanabe et al., 2018). However, the authors have noticed a research gap in the context of defining, identifying and analysing the potential effects of on-line hate speech in the area of actions comprising social marketing. Social campaigns often become a part of a public discussion (including in social media) and may affect the recipients in diverse ways (Kohls et al., 2017; Kunst et al., 2019). Such discussion may result from specific features of the campaign, especially when it is considered controversial by the recipients and the media. Given the multidimensional nature of hate speech and the current discussion about the blurry line between humour or freedom of expression and hate speech (MacAvaney et al., 2019; Tontodimamma et al., 2021), the authors decided to examine the attitude of an influencer, i.e. the author of pro-social content on the Instagram with respect to hate speech and the social campaigns which were deemed controversial or even offensive in the media and the public discourse.

The subject of the research is the phenomenon of online hate speech, accompanying pro-social activities conducted on the Internet, and more specifically in social media. The intention of the authors was to evaluate whether the simultaneous analysis of the phenomenon in question, both from the perspective of the creator of online pro-social content and the recipients of pro-social messages, will allow to verify whether there is a relationship between online hate speech and building the value of online pro-social activities.

This article attempts to obtain an answer to the research question: Can online hate speech have an impact on building the value of pro-social activities? Given the specific nature of the studied phenomena (the sensitivity of the topic), the authors decided on the qualitative nature of the conducted research, namely an individual in-depth interview (survey from the author's perspective) and a focus group interview supplemented with projection techniques – connotation test, evaluation of emotions and level of controversy (study from the perspective of the user of the social media).

Literature review

The existence and the role of social marketing was highlighted in Kotler's publication a number of years ago; it was stated there that it is an element of sustainable marketing (Kotler et al., 2012). Sustainable marketing is one of the main elements of corporate social responsibility (Adamczyk, 2009). The sustainable marketing defined in the publication assumes that an enterprise should operate in a manner that its marketing decisions support long-term beneficial operation of the marketing system and its significant part is social marketing that takes place when the marketing decisions are made in observance of the broader context and the interest of the entire community (Kotler et al., 2012). In response to the needs of the modern consumer, the concept of corporate social responsibility has developed, which means voluntary incorporation of social issues into the enterprise's economic activities (Kotler et al., 2012). As part of this concept, companies carry out social initiatives known as the socially engaged marketing, which has three main goals: social, image-related and economic, as opposed to social marketing, the main goal of which is only the aforementioned social element (Adamczyk, 2009; Latapí Agudelo et al., 2019).

Growing interest in the subject of social marketing resulted in increased attention of researchers and a growing number of published articles (Bhat et al., 2019). When elaborating on this concept, it is necessary to emphasize that social marketing is defined in a number of ways (Bhat et al., 2019; Dann, 2010; Kumar Nanda, 2013) and that digital technologies have a tangible impact on social marketing in many areas (Flaherty et al., 2021). One of the classic definitions is the one drawn up by Kotler and Lee; according to these authors, social marketing is a process that applies marketing principles and techniques to create, communicate and deliver value in order to influence target audience behaviour that benefits the society (public health, safety, the environment and communities), as well as the target audience (Kotler & Lee, 2008). Thus, social marketing is implemented with the use of numerous tools, principles, techniques and theories underlying marketing (Dann, 2010). Social marketing campaigns, created as part of social marketing, are analysed in academic articles, for example, with respect to the social problems that they intend to address, e.g. driving under the influence of alcohol, impact of advertisements in the social media on children or social stigmatisation focused on psychological diseases (Cismaru et al., 2009; Evans, 2008; Sampogna

et al., 2017). Extensive academic studies also focus on the issue of efficiency of social campaigns, as on account of their character and expected changes within the scope of broadly understood social behaviour, their impact and effects are often hard to measure (Helmig & Thaler, 2010; Lahtinen et al., 2020; Stead et al., 2007; De Vries & Duque, 2018). An important element from the perspective of this paper are studies pertaining to the so-called controversial social campaigns, not only on account of their financing by large concerns, e.g. tobacco (Hastings & Angus, 2011), but primarily in the context of the controversial stunts applied in them, which may affect the recipients' emotions (Hastings et al., 2004). Brennan and Binney analysed the feelings of recipients of social campaigns where negative appeals to their sense of fear, guilt and shame were used (Brennan & Binney, 2010). Feelings evoked in the recipients with the use of a social campaign devoted to *anorexia nervosa* were studied by Gomes and Casais with the use of text and emoji analysis (Gomes & Casais, 2018).

Hate speech and hating are phenomena that are relatively easy to notice in online communication, while the international debate about them relies on a number of examples of how they affect our reality, which is shown by such studies as *Histories of Hating* (Shepherd et al., 2015). Feeling of anonymity in the context of the content produced on-line and the idea of freedom and universal nature which guided the growth of the Internet affected the character of online communication which, in spite of encompassing a number of socially harmful behaviour, remains predominantly beyond any legal regulations (Banks, 2010). As mentioned in the introduction, hate speech does not have one uniform definition. Generally speaking, it is a particular form of offensive language that makes use of stereotypes to express an ideology of hate (Warner & Hirschberg, 2012). On the other hand, attention should be paid to the definition of "hate speech" according to which,

hate speech is defined as bias-motivated, hostile, malicious speech aimed at a person or a group of people because of some of their actual or perceived innate characteristics. It expresses discriminatory, intimidating, disapproving, antagonistic, and/or prejudicial attitudes toward those characteristics, which include gender, race, religion, ethnicity, colour, national origin, disability, or sexual orientation. (Cohen-Almagor, 2011)

When discussing hate speech, reference must be made to the persons who distribute hate speech. These are the so-called haters, who are the object of numerous academic publications. Haters are simply persons or users who post hate speech (Mondal et al., 2017). Following the DeTACT report (Detect Then ACT), the actions of haters may be divided into individual and group (DeTACT, 2020):

- individual haters often engage spontaneously, mostly in conversational rhetoric, individual goal is to seek to validate themselves,
- haters groups' actions are much more organized, often in private environments where they work on different ways of expressing their views, developing hateful imagery, creating new narratives, groups' aim is to gain power in society and enforce their normative systems onto others.

Furthermore, a survey carried out on Twitter has shown that the so-called hate instigators target more popular and high profile Twitter users, and that participating in hate speech can result in greater online visibility (Elsherief et al., 2018). Looking at it from a broader perspective, it may be concluded that hate speech is a multidimensional phenomenon and a certain type of a side effect of technological development, which was discussed in a UNESCO publication, which also offers the following remark:

Hate speech online is situated at the intersection of multiple tensions: it is the expression of conflicts between different groups within and across societies; it is a vivid example of how technologies as the Internet bring with them both opportunities and challenges; and it implies complex balancing between fundamental rights and principles, including freedom of expression and the defence of human dignity. (Gagliardone et al., 2015)

In the scientific discourse, there are relatively few studies presenting research combining the above-mentioned areas: social marketing and online hate speech. Among the available publications, the analysis of the international activist movement known as Sleeping Giants and a social-media “campaign to make bigotry and sexism less profitable” (Braun et al, 2019) deserves attention. The authors consider how the social campaign’s interventions speak to the larger debate around the normative relationship between advertising and the performance of the news ecosystem.

Another researcher presents the conclusions of the study of the reception of a communication campaign against hate speech online and cyber-bullying (Sundquist, 2017). Furthermore, Hodalska presents interesting recommendations and practical implications for prevention and intervention programs designed by marketing practitioners, foundations and organizations that aim to reduce online harassment, through various campaigns worldwide (Hodalska, 2016).

Research related to online hate speech and the social underlying forces active in the online environment in relation to risk and ill-treatment has also been presented in several other publications (Balica, 2017; Apriyani, 2021). However, the authors have not identified any publications addressing the issue of online hate speech from the perspective of a pro-social influencer.

Research methods

The authors decided to examine the aspect of receipt of controversial social campaigns content and the accompanying online hate speech in a broad research perspective, accounting for the analysis of the phenomenon from the point of view of an author of pro-social content on the Internet, as well as a user of social media. The procedure of quality studies of the phenomenon in question encompassed an in-depth interview (author’s perspective) and a focus group interview supplemented with projection techniques – connotation test, evaluation of emotions and level of controversy (study from the perspective of the user of the social media).

After a literature review and data analysis from a focus group interview, it seemed necessary to conduct an individual in-depth interview with the creator of pro-social content. Why was this step so important? During the FGI analysis, new issues and questions appeared that needed to be answered. Finally, in order to verify the research material, methodological triangulation was used, consisting of a comparison of conclusions from both applied research methods.

Individual in-depth interview

The outstanding actions of influencers focused on pro-social issues are observable in the social media. The choice of a specific influencer to conduct the IDI study was intentional and in line with the assumptions of the research. Kaya Szulczewska is a pro-social influencer who promotes the body positive movement and popularizes tolerance and acceptance among her audience on her Instagram @cialopozytyw_polska. This influencer owns one of the most popular profiles of the international body positive movement on the Polish Instagram, @cialopozytyw_polska with 54.4 thousand followers (January 2022). The content presented by the influencer is consistent with the premises of the movement – thus, it is a space that promotes acceptance of the human body along with its imperfections and overall social tolerance. The profile is devoted to pro-social actions and not commercial ones, as in the case of the majority of influencers; as a result, it often focuses on controversial subjects. Due to this, Kaya Szulczewska encounters a great amount of hate speech.

Given the time, financial and geographic restrictions, the individual in-depth interview with the influencer was carried out between February and March 2020 in a remote mode. This eliminated the adverse time pressure and guaranteed that the respondent had the possibility of thinking the answers through (Batorski & Olcoń-Kubicka, 2006). The influencer received a list of thematically diversified questions. The received answers provided a base for further interpretation of the research material.

Focus group interview with connotation test

Advertisement campaigns, along with social campaigns which construct their message on negative appeals or threat appeals in order to strongly influence the recipients' emotions, may evoke diverse reactions of the recipients, from positive, through constructive criticism up to hate (Moraes et al., 2016; Waller, 2006). The object of the studies carried out by the authors covered two Polish social campaigns which, in press releases and comments of online recipients, evoked extreme emotions and online hate speech:

(1) "Eat Carefully" ("Jedz ostrożnie"), with the form of a competition organised by outdoor advertisement agency, Art Marketing Syndicate SA (AMS SA), as part of the Galeria Plakatu AMS project (AMS, 2018), which was deemed a controversial

campaign, evoking negative emotions (Bosomtwe, 2019; Godziński, 2019; Nowy Marketing, 2019),

(2) “If you love, dare to change” (“KOCHASZ? OdWAŻ się na zmiany”) a campaign organised by Polskie Centra Dietetyczne, while the concept and performance of actions were the task entrusted to Agencja Reklamowa Brandbay.pl (Polskie Centra Dietetyczne, 2018), which did not evoke such extreme emotions as the former one (Redakcja PR, 2018).

The posters from the analysed campaigns were selected at random, but the choice of the campaigns and the subject matter that they referenced was not accidental. One of the reasons for such non-random choice was the fact that both campaigns referred to the same social problem, namely obesity; furthermore, in both campaigns posters were used as the form of communication. However, in spite of these common features, each campaign presented a different approach and tackling of the subject matter which resulted in evoking different emotions in the recipients.

The selection of people surveyed using the FGI method was related to the subject of the study and the research question posed. Six people (three women and three men), between 18 and 26 years of age, who use the Internet on a daily basis, with an emphasis on daily activity in social media, were surveyed. Such characteristics of the study group allowed us to assume that these people are relatively often exposed to vulgar, controversial content online. The choice of the survey participants was purposeful and took place through the recruitment form, which allowed to eliminate those who do not meet the conditions and, at the same time, to select the actual recipients of the social campaigns in question. All participants of the group interview were informed about the subject before the study, and participation was voluntary. The study took the form of a stationary meeting of participants with a moderator and began with an explanation of the purpose of the study. In the initial phase of the study, the moderator asked general questions not directly related to the subject of the study, in order to familiarize the participants and break the barriers of public speaking. After these introductory elements, the main part of the study was conducted.

The study was carried out with the use of a focus group interview supplemented with projection techniques: connotation test, evaluation of emotions and level of controversy according to the Likert scale. The participants of the interview were shown six posters (two sheets, each containing three posters), taken from both of the analysed social campaigns (Figures 1, 2). Their task was to:

- indicate connotations that are evoked by the presented images (open-end question),
- evaluate the emotions that the poster evokes in them (a 5-point Likert scale from 1 – *very negative* to 5 – *very positive*),
- respond to the question whether the poster was controversial (a 5-point Likert scale from 1 – *definitely no* to 5 – *definitely yes*).

The analysis of the information and the formulation of conclusions from the conducted FGI study were carried out on the basis of transcripts of the participants' statements.

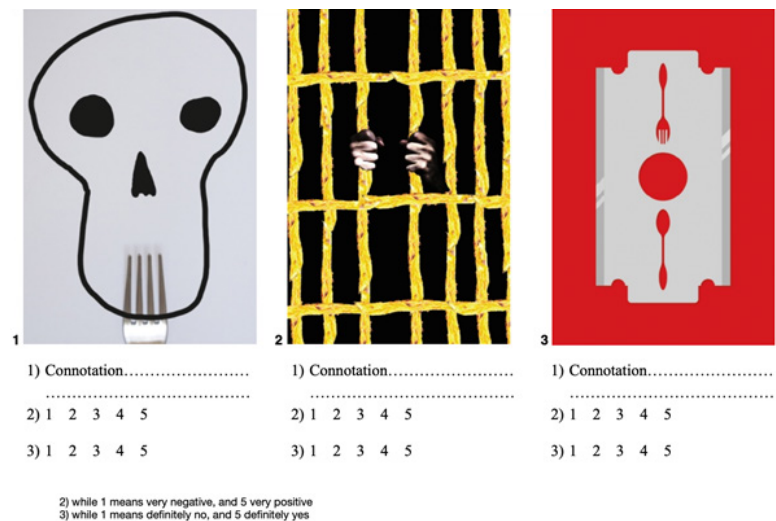


Figure 1. Focus group interview, connotation test form for the participants (p. 1)

Source: Authors' own study.



Figure 2. Focus group interview, connotation test form for the participants (p. 2)

Note: Translation of text from poster No. 5: “If you love, dare to change”. Make an appointment for a free dietary consultation for your child.¹ Translation of text from poster No. 6: “Live / Devour”.²

Source: Authors' own study.

¹ The caption in poster 5 makes use of a play on Polish words, in this case referencing weight and courage to change (translator's note).
² The caption in poster 6 makes use of a play on Polish words, rendering direct translation impossible. The word merges two words: “to live” (żyć) and “to devour” / “eat too much” (żreć).

Results

Authors' perspective

As could have been expected, the respondent influencer who addresses social problems did not have any major difficulties with defining and identifying manifestations of hate speech online. One of the first questions that Kaya had to answer was the frequency of hate content on her profile. The respondent noted that everything depends on the frequency of content added by her; in general, the frequency is on a daily level. Kaya often stressed that she wished for her profile *@cialopozytyw_polska* "to be a »safe space«, so I do not allow any critical comments about the body or advice pertaining to the appearance or other comments which could make the person publishing the photo on my profile feel offended. Haters are quickly pacified". On this basis, a conclusion may be drawn about a high level of engagement in fighting the hate content and a great amount of time that has to be devoted to such actions. Kaya also emphasised that hate also refers to her private account, *@kayaszu*, where, instead of sharing stories of others, she posts her own thoughts and where she is attacked not only with hate in comments but also "[she] receive[s] erotic type messages, which is not comfortable for [her]".

The aforementioned conclusion about the significant amount of time devoted to the removal of hate content was emphasised by the respondent in the next question. When answering the question about the impact of hate on her person, Kaya indicated several major issues. "Not only the mere content of hate is tiring, but its intensity, repetitiveness and thoughtlessness. It is also difficult to bear the thought that hate often reaches persons who may have problems with self-esteem and such thoughtless comments may destroy somebody's life. I am greatly saddened by it and I try to make sure that at *@cialopozytyw* nobody has to read it". It should be noted that several features were indicated here with which the respondent identifies hate, i.e. repetitiveness and thoughtlessness; Kaya also emphasised the harmfulness of hate, primarily for the readers and the followers of her profile and not herself. Furthermore, she discussed further actions that follow hate and which are much more aggravating for her, i.e. "It is not hate as such that affects my psyche, but personal libel, smear campaigns and public lynching". Kaya also talked about the impact of hate on her actions: "As far as my actions are concerned, removing hate and spam from posts and mailboxes does take a lot of time and sometimes slows down my actions; in some cases, I need a week to rest after the »shift during a shit-storm«". The last part of her response allows us to conclude that even a strong person, immune to hate content, may be overpowered by it and removal of hate resembles emptying the ocean with a teaspoon, in particular in reference to the aforementioned frequency of hate. Kaya was also asked whether she noticed any common characteristics in the hate statements appearing at the *@cialopozytyw_polska* profile.

The author indicated the consistency of content which triggers hate speech posts. These are, among others: "What appals people the most is obesity, bodies of fat

people, menstruation blood, hair on female bodies, abortion and feminism". The author also divides hate into more personal attacks: "people attack me personally for many different reasons, but they often get mad when I ask for support on Patronite or express my opinion on a subject". Thus, we have a situation where the author divides her actions into subjects that refer to her and her community, and the ones that she undertakes personally to maintain her operation. Eventually, the respondent added some additional features of haters and emphasised that based on her own observations, they usually derive from the so-called "fake accounts", set up exclusively to offend. Kaya also noted that according to her, haters are usually young persons, not intent on a discussion or talk, but only on venting their emotions. In addition, she claimed that haters appear in groups or pairs; she even talked about "hate raids" coordinated by other profiles or external groups. Such actions significantly hinder the functioning of the profile and contradict its main idea, because – as the author reiterated – "I would like to create a safe space, where one can anonymously share one's photos, stories and not be afraid of negative evaluation".

In the last question pertaining to the issues of hate, the author was asked if she agreed with the opinion that pro-social profiles on the Instagram attract more haters than commercial profiles. The author had a problem with an unequivocal answer to this question, which highlights the difficulties in determining the scale of the phenomenon; however, she noted that according to her, the greatest role is played by the subject matter, indicating pro-social profiles that do not address greatly controversial subjects which do not experience such a great volume of hate. In her last statement, she claimed that it was a problem pertaining to the entire industry: "As far as I know, the majority of bloggers and influencers struggle with some forms of hate, even if they do it professionally and there is a lot of commercial content".

Perspective of social media user

The first batch of questions referred to social problems. The participants did not have any difficulties with explaining a social problem, referring to its major features such as: it refers to a large group of people, entails divergence from certain standards and the fact that "a social problem does not unite the entire society, but divides it". The respondents stressed that it is necessary to react to social problems, emphasising that not all of them can be solved via legal regulations.

Hence, they noted the growing role of non-governmental institutions that address such phenomena, stressing the "available media, television, radio, even instead of advertisement posters, there should be more of these related to some social campaign instead of, for example, Media Markt advertisements". Thus, the participants manifested quite good knowledge and willingness to express their opinions. In case of questions pertaining to hate speech, the respondents in the study indicated such features as "emotionally negative statement (...) under the influence of some prej-

udice”; furthermore, in the course of the discussion, a problem with differentiating between hate speech and criticism emerged: “some people believe that hate speech is every type of negative opinion about anybody (...) sometimes I am not able to define what is permissible criticism and what is not.” In their statements, the respondents focused on online communication, thus, they were asked whether they saw hate speech outside of the Internet and they noted without any greater problems: public space referring, for example, to the writings on buildings, graffiti, and then “in the Catholic Church, there is a lot of hate speech which is disguised (...) it is often camouflaged, so that there are no vulgar words, sometimes it is nicely packaged” and “public speeches (...) some completely imaginary data were provided about paedophilia and homosexuality, and it was so... Well, it was clearly hate speech”. It may be concluded from the respondents’ statements that hate speech does not have to be directly communicated and full of vulgarisms; often, it is deeply embedded in the content of ostensibly innocent communications.

When asked if they have ever used hate speech, the participants seemed ill-at-ease and made an impression as if nobody knew what to say, until one participant jokingly claimed “I am wondering if anybody would admit to using it in general”. This statement and the respondents’ strange behaviour shows that the use of hate speech, in spite of its commonness, is not something socially accepted and is something bashful; additionally, it led to another remark of one of the respondents that a problem appears when: “I mean, there are ideological [people] and they do not really think that this is hate speech (...) but a grounded belief about the superiority of this race, etc. And for a person from the outside, this is sheer hate and this is hate speech, but that first person would not have called it like this. (...) nobody wants to think that they are using hate speech because this is... This is a grave accusation”. Here, another aspect worth emphasising was noticed, i.e. individuals dangerously cherishing their beliefs and ideals, who do not fully understand that they are using hate speech because they are deeply convinced that their views are justified.

The next batch of questions focused on social campaigns and when asked about their objectives, the respondents noted: popularisation of a problem, drawing attention to a problem, an attempt at starting changes; they also emphasised the essence of creating social campaigns which, according to them, not only are important, but – if carried out with a mission and vision – may offer a number of changes and show truth, which indicates a high level of trust to social campaigns.

The participants unanimously stated that in their immediate environment, they saw a lot of social campaigns in the public space; they listed predominantly posters and more extended and creative campaigns on the Internet. The respondents did not believe that there were too many campaigns and the ones that were created referred to various subjects, thanks to which the recipients did not grow indifferent to them.

The moderator also asked if it was easy to exhaust a problem and address it *ad nauseam*: “It does not really get boring, because (...) the problem is still present and

will be present; it seems that as long as it is present, various versions of the campaign will be visible and available". In further responses, the respondents emphasised how careful one had to be when using controversy or toying with emotions in social campaigns, so as not to turn an important issue into a joke; on the other hand, they stressed that "(...) when a campaign plays with emotions, it is remembered and does not matter whether it makes us laugh or makes us angry and we remember it anyway, so this is rather necessary. The campaign has to be remembered".

The last part was intended as an analysis and confrontation with the respondents' answers in the connotation test carried out before the start of the survey. Earlier, the respondents, without in-depth familiarity with the research problem and issues of selected social campaigns, were asked to perform three tasks. Projection techniques were used for it: connotation test, evaluation of emotions and level of controversy according to the Likert scale. The connotation test comprised six posters, completely different, which offered diverse conclusions:

(1) The first poster (skull) evoked connotations related to death and danger, in spite of the fact that in majority, the emotions defined on the Likert scale were at 3 (neutral), similarly to the level of controversy.

(2) The second poster (French fries) was associated with obesity and slavery, depending on whether the respondents understood the message or not. Emotions were determined primarily as definitely negative, while the level of controversy was defined as low, even though two persons marked it as controversial.

(3) The third poster (razor blade) was associated with eating supper; its reception was determined as neutral, while the level of controversy was diverse, with dominant determination as controversial.

(4) The connotations related to the fourth poster (sausage) mainly refer to an explosion and evoke rather negative emotions and was definitely controversial.

(5) The fifth poster evoked completely diverse connotations among the participants, from pregnancy to sadness; the emotions were determined on the neutral level and the level of controversy was rather high.

(6) The last poster (devour) evoked inconsistent connotations from politics to hate, emotions were primarily determined as negative, while the level of controversy was diversely assessed, from high to very low.

During the analysis of the respondents' answers, certain dependences may be noted. "I also chose neutral, because I did not really know what to answer", which shows lack of overall familiarity with the subject matter; unclear signals referring to the interpretation of the posters caused the respondents' disorientation; this was also visible during the completion of the form, as with respect to some of the posters, the response was entered automatically, while next to others, it required several focused facial grimaces and several minutes to decide. "I decided that this was about (...)" is an example of another phenomenon, i.e. thinking, guessing and the respondents' satisfaction if their connotations matched the main subject of the poster, as if they solved a riddle, and it seems that a poster should not be a riddle for the recipients.

The respondents unanimously conceded that they did not guess the problem to which the first of the presented posters referred to, which blocked determination of their emotions. Responding to the question why they entered a specific number, they were confused; they shrugged their shoulders and answered that they did not know. It is possible that they were afraid of positively assessing posters that referred to a negative phenomenon, especially that as mentioned earlier by the respondents, the majority of these posters addressed the issue of obesity in a non-standard manner. During the study, a number of non-verbal signals were also observed, sent by the participants, such as: frowning, scratching one's back, nervously tapping fingers on the chair, raising one's eyebrows, which indicate awkwardness and uncertainty during the completion of the survey.

At the end of the interview, the respondents confirmed that the familiarity and clearness of the subject matter influences the reception of a poster and facilitates the possibility of determining emotions. It is worth drawing attention to the fact that at the beginning, when answering questions about the "incomprehensible" posters, there was uneasy silence, until one of the participants confessed that he did not really understand what the poster was about. Right after this confession, the participants jokingly admitted that he was right and exchanged inaccurate connotations. This situation stresses the fact that the subjects discussed in such a survey were not the easiest ones. During the interview, the respondents stated that the poster presenting an obese child titled "If you love, dare to change" is so neutral that it would not have attracted their attention and they would remain indifferent towards it; they noted that the last poster was much more controversial for them; it was definitely much more offensive.

This is an interesting conclusion; in the media, the presentation of an obese child on the poster was widely discussed; such a child could have been exposed to annoying and unacceptable remarks, similarly to the last poster where the content was deemed provocative and potentially too painful for persons who are struggling with a specific problem. This leads to the conclusion that possibly nowadays, in the era of social media, a message focused more on words than on images may be much more controversial and offensive for the recipients, while the human body in various dimensions appears in the social media so often that it is no longer controversial.

Discussions and conclusions

Use of the aforementioned research methods allowed for capturing the full perspective of the described phenomena: the perspective of the author of pro-social content in the social media and the recipients of social campaigns, which provided an opportunity for drawing thorough conclusions.

The performed survey confirmed the diversity of emotions evoked by pro-social campaigns created with the use of emotional appeals, also encompassing emotions

opposite from the ones that were expected. This is consistent with the publications of Brennan and Biney, who thoroughly examined the effects of using the following emotions in pro-social campaigns: fear, guilt and shame (Brennan & Binney, 2010). The survey was carried out in reference to a campaign created to increase compliance with income reporting requirements and shown that in this case, the use of negative appeals more often invoked self-protection and inaction (Brennan & Binney, 2010). A similar stance could be noted among the respondents during the focus group interview, where in case of objects presented on posters in a negative context, even when the main message was not understood, they evoked the feelings of fear, aversion or indifference. Gomes and Casais analysed a social campaign where treat appeals were used, and the survey relied on the analysis of reactions of its recipients on Facebook and YouTube (Gomes & Casais, 2018). Similarly, the survey results indicate extremely positive and negative emotions; the positive ones refer to the attempt at showing support or compassion, while the negative ones expressed primarily fear and sadness (Gomes & Casais, 2018).

Conclusions derived from the analysis of the online hate speech phenomenon from the perspective of the author of online pro-social content confirm the supposition that hate affects not only the author, but also the entire community, which may have an impact on building the value of the influencer's pro-social activities. The influencer's role is to remove such harmful content, which is a never-ending activity, burdening not only the psyche, but also time-consuming. Hate adopts more extended forms, which are meant to dominate the influencer and prevent removal of specific content. An interesting remark is the fact that in spite of earlier conclusions about the influencer's deep identification with her work, there is a distinction between the portions of hate that may affect her community and the ones that are more painful for her. This shows the cause-and-effect relationship between the daily removal of such content, simultaneous resistance to it and a feeling of responsibility for own community which should not have contact with hate speech in such a place.

From the perspective of social media users, it may be noted that in spite of not fully accurate connotations before being introduced to the subject matter, the participants of the in-depth group interview after understanding the leitmotif of the social campaign, decided that the most interesting, but also simultaneously the most controversial, was the minimalist poster relying on the word *żryj* ("devour") noting that for them, it was more suggestive, interesting and convincing than an image of an obese child on another poster. This may lead to a conclusion that in the world full of images and visual communication, a campaign based on words may offer a distinguishing trait. The respondents did not conceal their difficulties with a clear-cut determination and designation of borders between hate and offending someone, as well as expressing own opinions, because anonymity and the scale of the phenomenon makes them flooded with such content on a daily basis. It is interesting to note how, at moments, they were throwing up their hands helplessly during the discussion about fighting hate on-line; they said that in their personal experience,

they ignore hate content. Faith in the essence and the potential of social campaigns that the respondents see are also worth emphasising; the respondents stressed that they are much needed and may actually influence the formation of reality; in case of the message and the creation of content that stimulates to action, the respondents were aware of the fact that the message has to stand out in order to be noticed in the flood of messages and advertisements, but simultaneously cannot be offensive.

The authors are aware that the qualitative nature of the conducted research may determine the extent to which the conclusions drawn from it will be used. However, taking into account the limited research achievements so far, analyzing the relationship between the activities of pro-social influencers and hate speech on the Internet – the analyses carried out can be treated as a preliminary study with interesting findings and a starting point for further, deepened analyses. The results of this research could be used during the preparation and implementation of various, not only pro-social projects, in particular using the concept of corporate social responsibility. Generally, this concept could be used as an effective tool for creating a competitive advantage, which is the basis for the growth of the value of any activity.

Limitations and future research

The survey was carried out in line with the premises of a quality survey which was affected by a complex and multi-dimensional character of the research questions set at the very beginning of the research process. In this case, it was not about the quantitative description of the reality; the focus was not on the question of how many influencers are engaged in social activities. To the contrary, the research questions that were asked used such words as: “What?”, “How?” and “Why?”. For example: What is the cause of hate?, Why do the influencers focus on social problems?, How to talk about a social problem in the modern times? Thus, the results of the survey constructed and carried out in this manner and conclusions derived from it cannot be fully satisfactory on account of the limited range of quality data that were received. However, such knowledge may offer an important assumption for further scientific research in the area of ethics of communication in social campaigns, as well as provide a guideline for the authors of social campaigns in the social media. Performance of further research in the discussed direction may contribute to the popularisation of this relatively new subject matter in the scientific discourse and indicate new modes of using the social media channels and reaching the selected target groups not only via paid advertisements but with the participation of influencers who are active in the pro-social area (not commercially) and who have a community of engaged followers.

References

- Adamczyk, J. (2009). *Spółeczna odpowiedzialność przedsiębiorstwa*. Warszawa: PWE.
- Alshamrani, S., Abusnaina, A., Abuhamad, M., Nyang, D., & Mohaisen, D. (2021). Hate, obscenity, and insults: Measuring the exposure of children to inappropriate comments in YouTube. *Companion Proceedings of the Web Conference 2021*. ACM, New York.
- AMS. (2018). Galeria Plakatu. Retrieved from <https://galeriaplakatu.ams.com.pl/poster/19-edycja-konkursu-jedz-ostroznie-2018/posters>
- Ananthakrishnan, U.M., & Tucker, C.E. (2021). The drivers and virality of hate speech online. *SSRN Electronic Journal*. doi:10.2139/ssrn.3793801
- Apriyani, M.N. (2021). Optimalisasi Internet Sehat Sebagai Upaya Pencegahan Tindak Pidana Hate Speech Di Pondok Pesantren Darul Falah Ponorogo. *Veteran Society Journal*, 2(1). doi:10.33005/vsj.v2i1.16
- Arango, A., Pérez, J., & Poblete, B. (2019). Hate speech detection is not as easy as you may think: A closer look at model validation. *SIGIR 2019 – Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, Inc.
- Assimakopoulos, S., Baider, F.H., & Millar, S. (2017). *Online Hate Speech in the European Union*. Cham: Springer. doi:10.1007/978-3-319-72604-5
- Balica, R. (2017). The criminalization of online hate speech: It's complicated. *Contemporary Readings in Law and Social Justice*, 9(2), 184–190. doi:10.22381/CRLSJ92201710
- Banks, J. (2010). Regulating hate speech online. *International Review of Law, Computers and Technology*, 24(3), 233–239.
- Batorski, D., & Olcoń-Kubicka, M. (2006). Prowadzenie badań przez Internet, podstawowe zagadnienia metodologiczne. *Studia Socjologiczne*, 3(128). Retrieved from <https://depot.ceon.pl/handle/123456789/2202>
- Bhat, S.A., Darzi, M.A., & Hakim, I.A. (2019). Understanding social marketing and well-being: A review of selective databases. *Vikalpa*, 1 June.
- Białek-Szwed, O. (2018). Mowa nienawiści problemem etycznym współczesnych mediów. *Zeszyty Naukowe KUL*, 61, 4(244), 331.
- Bosomtwe, O. (2019). 'Jedz ostrożnie' kontrowersyjna kampania AMS 'Żryj' nic nie wnosi. Retrieved from <https://noizz.pl/opinie/jedz-ostroznie-kontrowersyjna-kampania-ams-zryj-nic-nie-wnosi/8spnx56>
- Braun, J.A., Coakley, J.D., & West, E. (2019). Activism, advertising, and far-right media: The case of Sleeping Giants. *Media and Communication*, 7(4): Peripheral Actors in Journalism: Agents of Change in Journalism Culture and Practice. doi:10.17645/mac.v7i4.2280
- Brennan, L., & Binney, W. (2010). Fear, guilt, and shame appeals in social marketing. *Journal of Business Research*, 63(2), 140–146.
- Cismaru, M., Lavack, A.M., & Markewich, E. (2009). Social marketing campaigns aimed at preventing drunk driving: A review and recommendations. *International Marketing Review*.
- Cohen-Almagor, R. (2011). Fighting hate and bigotry on the Internet. *Policy & Internet*, 3(3), 89–114.
- Dann, S. (2010). Redefining social marketing with contemporary commercial marketing definitions. *Journal of Business Research*, 63(2), 147–153.
- DeTACT. (2020). Online hate speech: Introduction into motivational causes, effects and regulatory context. Media Diversity Institute. Retrieved from <https://www.media-diversity.org/resources/online-hate-speech-introduction-into-motivational-causes-effects-and-regulatory-context/>
- Elsherief, M., Nilizadeh, S., Nguyen, D., Vigna, G., & Belding, E. (2018). Peer to peer hate: Hate speech instigators and their targets. *Proceedings of the International AAAI Conference on Web and Social Media*, 12. Retrieved from <https://dev.twitter.com/streaming/>
- Evans, W.D. (2008). Social marketing campaigns and children's media use. *Future of Children, Future Child*, March.
- Facebook. (2021). Community Standards. Retrieved from https://en-gb.facebook.com/communitystandards/hate_speech

- Flaherty, T., Domegan, C., & Anand, M. (2021). The use of digital technologies in social marketing: A systematic review. *Journal of Social Marketing*. doi:10.1108/JSOCM-01-2021-0022
- Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). Countering online hate speech – UNESCO Digital Library. UNESCO. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000233231>
- Godziński, B. (2019). Kampania AMS „Jedź Ostrożnie!”. O co chodzi z plakatem „Żryj – żyj”? Wywiad. *Na Temat*. Retrieved from <https://natemat.pl/260885,kampania-jedz-ostroznie-wyjasnienie-plakatu-zryj-od-ams-wywiad>
- Gomes, R.F., & Casais, B. (2018). Feelings generated by threat appeals in social marketing: text and emoji analysis of user reactions to *anorexia nervosa* campaigns in social media. *International Review on Public and Nonprofit Marketing*, 15(4), 591–607.
- Hastings, G., & Angus, K. (2011). When is social marketing not social marketing? *Journal of Social Marketing*, 1(1), 45–53.
- Hastings, G., Stead, M., & Webb, J. (2004). Fear appeals in social marketing: Strategic and ethical reasons for concern. *Psychology and Marketing*, 21(11), 961–986.
- Helmig, B., & Thaler, J. (2010). On the effectiveness of social marketing – what do we really know? *Journal of Nonprofit and Public Sector Marketing*, 22(4), 264–287.
- Hodalska, M. (2016). Cyberbullying, fear and silence: from bystanders to cyber-samaritans. In M. Hodalska, C. Ghita, & J. Bapat (Eds.), *Perils of the Web : Cyber Security and Internet Safety* (pp. 23–36). Inter-Disciplinary Press.
- Kazojć, K. (2014). Koncepcja społecznej odpowiedzialności i jej obszary w organizacjach. *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania*, 38(1), 58.
- Kohls, E., Coppens, E., Hug, J., Wittevrongel, E., Van Audenhove, C., Koburger, N., ... & Arensman, E. (2017). Public attitudes toward depression and help-seeking: Impact of the OSPI-Europe depression awareness campaign in four European regions. *Journal of Affective Disorders*, 217, 252–259.
- Kotler, P., Keller, K.L., Ancarani, F. G., & Costabile, M. (2012). *Marketing Management*. Italy: Pearson, Italia.
- Kotler, P., Lee, R.N. (2008). *Social Marketing: Influencing Behaviors for Good*. Thousand Oaks: Sage Publications. Retrieved from <https://www.amazon.com/Social-Marketing-Influencing-Behaviors-Good/dp/1412956471>
- Kovács, G., Alonso, P., & Saini, R. (2021). Challenges of hate speech detection in social media. *SN Computer Science, Springer Science and Business Media LLC*, 2(2), 1–15.
- Kumar Nanda, A. (2013). Social marketing: A literature review. *International Journal of Science and Research (IJSR)*, 4(9). Retrieved from www.ijsr.net
- Kunst, J.R., Bailey, A., Prendergast, C., & Gundersen, A. (2019). Sexism, rape myths and feminist identification explain gender differences in attitudes toward the #metoo social media campaign in two countries. *Media Psychology*, 22(5), 818–843.
- Laaksonen, S.-M., Haapoja, J., Kinnunen, T., Nelimarkka, M., & Pöyhtäri, R. (2020). The datafication of hate: Expectations and challenges in automated hate speech monitoring. *Frontiers in Big Data*, 3(3), 3.
- Lahtinen, V., Dietrich, T., & Rundle-Thiele, S. (2020). Long live the marketing mix. Testing the effectiveness of the commercial marketing mix in a social marketing context. *Journal of Social Marketing*, 10(3), 357–375.
- Latapi Agudelo, M.A., Jóhannsdóttir, L., & Davídsdóttir, B. (2019). A literature review of the history and evolution of corporate social responsibility. *International Journal of Corporate Social Responsibility*, 4(1), 1–23.
- MacAvaney S., Yao H.-R., Yang E., Russell K., Goharian N., Frieder O. (2019). Hate speech detection: Challenges and solutions. *PLoS ONE*, 14(8), e0221152. doi:10.1371/journal.pone.0221152
- Mondal, M., Silva, L.A., & Benevenuto, F. (2017). A measurement study of hate speech in social media. *HT 2017 – Proceedings of the 28th ACM Conference on Hypertext and Social Media*, Association for Computing Machinery, Inc.

- Moraes, C., Ferreira, C., Michaelidou, N., & McGrath, M. (2016). Consumer ethical judgement and controversial advertising avoidance on social media. In L. Petruzzellis & R. Winer (Eds.), *Rediscovering the Essentiality of Marketing* (pp. 189–193). Cham: Springer.
- Nowy Marketing. (2019). „Jedź ostrożnie” – kontrowersyjna kampania AMS dzieli odbiorców [opinie]. *Nowy Marketing*. Retrieved from <https://nowymarketing.pl/a/20880,jedz-ostroznie-kontrowersyjna-kampania-ams-dzieli-odbiorcow-opinie>
- Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M., & Räsänen, P. (2014). Exposure to online hate among young social media users. *Sociological Studies of Children and Youth*, 18, 253–273.
- Polskie Centra Dietetyczne. (2018). *OdWAŻ SIĘ na zmiany – akcja społeczno-edukacyjna*. Retrieved from <https://odwazsienazmiany.pl/>
- Rappaport, A. (1999). *Wartość dla akcjonariuszy*. Warszawa: WIG_Press.
- Redakcja PR. (2018). *Rusza kampania społeczno-edukacyjna „Kochasz? Odważ się na zmiany”*. Retrieved from <https://publicrelations.pl/rusza-kampania-spoleczno-edukacyjna-kochasz-odwaz-sie-na-zmiany/>
- Sampogna, G., Bakolis, I., Evans-Lacko, S., Robinson, E., Thornicroft, G., & Henderson, C. (2017). The impact of social marketing campaigns on reducing mental health stigma: Results from the 2009–2014 time to change programme. *European Psychiatry*, 40, 116–122.
- Shepherd, T., Harvey, A., Jordan, T., Srauy, S., & Miltner, K. (2015). Histories of hating. *Social Media + Society*, 1(2). doi:10.1177/2056305115603997
- Statista. (2022). Global number of hate speech-containing content removed by Facebook from 4th quarter 2017 to 1st quarter 2022. Retrieved from <https://www.statista.com/statistics/1013804/facebook-hate-speech-content-deletion-quarter/>
- Stead, M., Gordon, R., Angus, K., & McDermott, L. (2007). A systematic review of social marketing effectiveness. *Health Education*, 107(2), 126–191. doi:10.1108/09654280710731548
- Sundqvist, J. (2017). *N0 H8 M8 (No hate mate) : En enkätstudie om lärarstudenters mottagande av en kampanj mot nâthat [N0 H8 M8 (No hate mate) : A survey and reception study of teacher students' attitudes towards a campaign against online hate speech]*. Independent thesis Basic level (degree of Bachelor), Jönköping Univ.: Publications.
- 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.
- De Vries, E.L.E., & Duque, L.C. (2018). Small but sincere: How firm size and gratitude determine the effectiveness of cause marketing campaigns. *Journal of Retailing*, 94(4), 352–363.
- Waller, D.S. (2006). A proposed response model for controversial advertising. *Journal of Promotion Management*, 11(2–3), 3–15.
- Warner, W., & Hirschberg, J. (2012). *Detecting Hate Speech on the World Wide Web*. Retrieved from <http://info.yahoo.com/legal/us/yahoo/utos/utos-173.html>
- Watanabe, H., Bouazizi, M., & Ohtsuki, T. (2018). Hate speech on Twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection. *IEEE Access*, 6, 13825–13835.