Social media microinterventions: Testing information activism as a media and information literacy tool

Maia Klaassen
University of Tartu, Estonia

Maria Murumaa-Mengel
University of Tartu, Estonia

Marju Himma
University of Tartu, Estonia

ABSTRACT

Several studies have shown the effect of information activism and microinterventions, such as I Am Here International, the Elves and #NAFO to combat information disorder and hate online. Nevertheless, microinterventions have yet to be conceptualised in promoting media and information literacy (MIL) and informational resilience. This study positions microinterventions as information activism tools and empirically tests microinterventions in the context of higher education. Using an action research approach at a university MIL training course, we aim to understand what types of information activism are used and how the collective interventions affected the participants’ MIL-s. We construct a typology of information activism roles and corresponding study tasks: the Worker Bees (focused on microprotections), the Meerkats (microaffirmations) and Lions (microchallenges). These different types surfaced in each participant over a cycle of exhaustion, disappointment, group support, feedback and encouragement that we dubbed the phoenixing cycle.

Keywords: media and information literacy, microinterventions, information disorder, information activism, digital hostility.
INTRODUCTION

Our study is situated after the COVID-19 pandemic came with a vortex of information disorder and at the beginning of the Russian military invasion of Ukraine in 2022. In Estonia, we noticed our journalism and communication students at the University of Tartu were describing media avoidance strategies (Ytre-Arne & Moe, 2021) to cope with the situation. Similar signs of withdrawal have been noted in youth participation studies (Allaste & Saari, 2020; Briggs & Briggs, 2017). Individuals can feel disillusioned, discouraged, and fatalistic in the large-scale terminological confusion, platforms, media and technology. In the context of information overload (Kaur et al., 2021) and attention economy (Hendricks & Vestergaard, 2019), it seems incivility in online participation (Lutz & Hoffmann, 2017), violent e-bile, hostility, and vitriol have become lingua franca in most online spaces (Jane, 2016). The spiral of silence on social media (Sohn, 2022) and self-censorship (Gera et al., 2020) can make marginalised opinions seem mainstream but also make widely shared (silent) opinions seem less prevalent.

We view this as highly problematic in the context of media and information literacy (MIL) education and digital participation. To disclose our positioning – we are predominantly qualitative researchers (Sciarra, 1999), significantly engaged in societal outreach and science communication, and we are all Estonians, a neighbouring country of Russia. All this context forms “layers upon layers of personal and collective memories, emotions, and roles” (Murumaa-Mengel, 2023, p. 68) that cannot be omitted from research or teaching. It entails switching between roles, and we agree with the problematisation of “objectivity” as a constructed notion itself (Bennett, 2003), as science and teaching can only be performed through subjects, “fully objective knowledge in this view is impossible for us humans” (Lindhult, 2019, p. 24). Thus, we have partially embraced the roles of insider-researchers and activist teachers (Garton & Copland, 2010; Labaree, 2002; Taylor, 2011).

In this article, we describe the outcome of an optional experimental course, Media and Information Literacies’ Micro-Interventions, designed and carried out to test strategies that enable participation in often hostile online environments, which have been well described in the so-called “post-truth” world implications (Lewandowsky et al., 2017). One way to exercise the right for active participation and fight the spread of mis-, dis-, and malinformation (Wardle & Derakhshan, 2017) is information activism (Halupka, 2016) – a form of networked action that enables people to engage with democratic processes and social institutions via (small) political actions mainly taking place on digital platforms. Activism is often time-consuming and stress-inducing, a high-stakes activity from the perspective of personal safety, so we turned our attention to strategies requiring less effort, time and energy, submerged in the flow of everyday life (Bakardjieva, 2009). Microinterventions are conceptually established in health studies (e.g., Baumel et al., 2020) and studies focusing on race or gender (e.g., Sue et al., 2019).

Empirically, this study is inspired by specific cases of information activism and online grassroots movements such as #Iamhere (Buergar, 2021; Johansson & Scaramuzzino, 2022) and NAFO (Boichak & Hoskins, 2022). #Iamhere is now an international loosely tied movement where, locally, collectives of thousands of people have made a regular practice of responding en masse to what they regard as hateful comments online. NAFO (North Atlantic Fellas Organization) is a grassroots support movement for Ukraine, symbolised by the dog Shiba Inu. It aims to expose and ridicule Russian propaganda and raise funds to support the Ukrainian army.

However, a coordinated approach to microinterventions has not been tested as a study tool for MILs. Defining microinterventions as low-burden everyday actions (Baumel et al., 2020) taken by ordinary social media users to counteract, challenge, diminish, or neutralise incivility and information disorder, we set off to test microinterventions as a tool in the context of MIL training at a university course, approaching via action research method principles.

THEORETICAL BACKGROUND

Information disorder and incivility contagion in the context of networked publics

Information disorder (Wardle & Derakhshan, 2017) is an overarching term to describe mis-, dis- and malinformation, and also a term referring to complex information ecosystems (Allen et al., 2020) that people must navigate in their technologically saturated life. For terminological clarity – disinformation refers to fabricated, manipulated, imposter or misleading content created with the intent to harm. Misinformation can be intentional or unintentional, e.g., inaccurate misinterpretation (Rubin, 2019). Malinformation may
be true or false but is intended to be malicious and disseminated to harm the subject or the receiver (Zhou & Zafarani, 2020). As with microinterventions, people are engaged in connective actions (Wang & Zhou, 2021), they may target any category of information disorder.

The general hostile tone and the “recreational nastiness” (Jane, 2014) of the online communication culture are at the heart of this study. Hostility in online communication is a significant element in contemporary digital culture (Thompson & Cover, 2022); the jokes are borderline, and trolling (deliberately starting arguments or provoking an emotional response) is seen as entertainment (Laineste, 2013; Pasta, 2022). “Others” seem to be dehumanised. Furthermore, incivility online inflames polarisation (Suhay et al., 2018) on various topics. According to a large-scale study by Kim et al. (2021), platforms matter – Facebook (FB) comments are 77% more toxic than comments elsewhere, implying that the distorting prism of social media exaggerates people’s hostility. Kim et al. (2021) found empirical evidence for incivility contagion (Gervais, 2017) or incivility spiral (Andersson & Pearson, 1999) – exposure to toxic comments increases toxicity of the following comments. In other words, when targeted with incivility, people might return the rudeness, and the instigator may respond with further incivility. According to this model, incivility begets incivility, aggression, and possibly violence (Cortina et al., 2022), and digital hostility “has become a framing factor in the reduction of quality of public debate at a social level and, at an individual level, has been cited as responsible for withdrawal, disconnection and negative impact on health and mental health” (Cover, 2022, p. 79).

Despite structural characteristics and macro-level pressures, people’s agency ensures established ways of doing things can be changed when “people start to ignore them, replace them, or reproduce them differently” (Gauntlett, 2008, p. 102), as “human agency is characterised by an innate ability to imagine different outcomes” (Wiggins & Bowers, 2014, p. 1894). The strength of networked publics (Boyd, 2010) is not merely in the technology-related infrastructure but also in forming imagined collectives. Networked (counter)publics may be interpreted as imagined communities or groups bound together by shared experiences or worldviews (Livingstone, 2005) or connected through the creation and dissemination of media, all leading to participatory culture (Jenkins & Ito, 2015).

An example is the grassroots movement #Iamhere that builds on some of the participating audience members being invisible – many social media users prefer to stay in the background, lurking (Hargittai & Walejko, 2008), avoiding conflict and confrontation (Rashidi et al., 2020). This can be further proof of the spiral of silence on social media (Sohn, 2022) but also an illustration that not many people hold specific polarised views on different matters. They follow the discussions, slowly making up their mind about an issue. This (often invisible) movable middle should be the target of polite and respectful interventions “attempting to lessen the impact of the hateful speech by hiding it in the comment threads” (Buerger, 2021, p. 2).

Coping strategies and media and information literacies

The communicative culture of incivility and information disorder can be the source of stress and trigger different coping strategies, usually falling into the response framework of fight, flight, or freeze (Cortina et al., 2022; Elstad et al., 2021). Fighting or fleeing from a perceived threat or attack are two initial biobehavioural reactions to stressors (Cortina et al., 2022). Fight-reaction coping strategies are based on confronting the danger (Skinner et al., 2003). Flight reactions function essentially as running away from the stressor, an intended or realised escape from the present situation (Elstad et al., 2021).

Many researchers have expanded the model to include a freeze-reaction to fear – reduction in mobility, an absence of behaviour (Cortina et al., 2022) or paralysis of sorts (Elstad et al., 2021). When encountering online hostility, harassment, or information disorders, all three reactions are common: a person can operate from the basis of the fight (e.g., confronting the offender, arguing and engaging in the spiral of incivility), flight (e.g., moving away from the digital space and interactions, deleting and blocking stressful content, engaging in media avoidance practices) or freeze (e.g., staring the screen for long periods, unsure of actions to be taken or to be avoided).

Frameworks of digital literacies (e.g., DigiComp) and media and information literacies (e.g., UNESCO’s MIL-curriculum) often stress the value of an active response. According to Hobbs and Jensen (2009), media literacy is the “active inquiry and critical thinking about the messages we receive and create”. A model consisting of five digital and media literacy competencies by Hobbs (2017) lists the following – (i)
accessing information, (ii) analysing information, (iii) creating one’s message, (iv) reflecting on one’s consumption, and (v) being socially responsible.

Several scholars have indicated gaps in the scholarly conversations about media literacy (see Boyd, 2018; Bulger & Davidson, 2018). For example, there is no clear understanding of which of the many literacies MIL encompasses to help build individual information resilience. Przybylski et al. (2014, p. 4) understand resilience as an individual’s ability to adapt to changing and stressful environments and feel empowered to act instead of react in the face of novel and threatening challenges. Presenting information disorder and informational influencing as an existential threat can lead to fear and inactivity, as opposed to empowerment, which is what MILs should do (Klaassen, 2021). Bulger and Davidson (2018, p. 20) have – among others – urged the development of “curricula for addressing action in addition to interpretation”. Roozenbeek et al. (2023) propose democratising efficacy testing of misinformation interventions in real-world settings and call for identifying the causal impacts of misinformation interventions. Tamboer et al. (2024) used a learning-by-doing approach to writing and checking news articles. In short, similar to Hobbs, Kanižaj and Pereira (2019, p. 7), we see a need to move from transmission education to empowerment education.

**Microinterventions as a form of information activism**

Information activism has been used in interpersonal communication to demonstrate how marginalised groups create techniques for generating and sharing information inside and outside their group (e.g., Wang & Zhou, 2021), closing the gap between the personal action frames characteristic of new technology and the thick collective capacity of traditional action” (Halupka, 2016, p. 1501).

Microinterventions – established in the field of health studies (Baumel et al., 2020; Elefant et al., 2017) and research on race or gender (Baleria, 2019; Houshmard & Spanierman, 2021; Sue et al., 2019). “Microinterventions are individual actions that ordinary citizens can take to voice disapproval, educate others, and pressure those in authority to make changes” (Sue et al., 2020, p. 24-25). Sue, Calle, Mendez, Alsaidi and Glaeser (2019) view the actions and strategies in the context of individual and systemic racism and distinguish microinterventions, all of which are fight responses.

First, **microaffirmations** (microcompliments, microsupports, and microvalidations), from reaffirming comments (e.g., “She is right!”) to nonverbal shared eye-rolls and amplifications of utterances (e.g., “as they pointed out before, […]”), are about validating and making people feel seen, heard, valued and supported. In the context of our study, this could mean reacting positively to others’ comments, posting supportive comments, amplifying and reposting, etc.

Second, **microprotections** are a variety of everyday actions that try to overwrite social scripts. Usually, they revolve around cultural literacy or critical consciousness, promoting self-esteem and preparing for (harsh) realities of social life. Whilst rooted in race studies (Sue et al., 2020), microprotections are also resilience-building techniques, an “inoculation” against the realities of social life” (McGuire & Papageorgis, 1961; Roozenbeek et al., 2022) promises a possibility of preemptively building resilience against anticipated exposure to information disorder. Inoculation consists of forewarning (a perceived threat of an impending attack on one’s attitudes) and exposure to a (micro)dose of misinformation (Compton, 2013). Inoculation has proven to increase manipulation recognition, boost confidence, increase the ability to discern trustworthy from untrustworthy, and improve the quality of sharing decisions (Roozenbeek et al., 2022).

Third, **microchallenges** are aimed at confronting biased information, actively challenging biased institutional or societal policies and practices, engaging in social advocacy and even civil disobedience. “While microaffirmations and microprotections primarily focus on fortifying individual targets and/or their communities, the focus of microchallenges is directly aimed at the perpetrator’s behaviour” (Sue et al., 2020, p. 39).

This typology is useful in analysing responses to other social injustices (Awad & Connors, 2023; Nadal et al., 2021), and we aim to apply it in the context of empowering and resilience-building MILs needed to recognise strategically coordinated information disorders and digital hostility in online communication (Cover, 2022). Microinterventions can help “overcome one of the primary reasons for inaction: the sense of powerlessness and futility” (Sue et al., 2020, p. 23). Drawing from the above, we pose three research questions (RQ):

- RQ1: Which microintervention strategies were used by the participating students, and for what reasons?
• RQ2: How did participation in information activism affect the students’ self-reflected and described media and information literacies?
• RQ3: How can the concept of microinterventions be used in teaching information activism in higher education?

METHODOLOGY

Study process and data collection

In the spring of 2022, journalism and communication students (n = 28) of the University of Tartu, Estonia, participated in the optional course *Media and Information Literacies’ Micro-Interventions*, during which they were invited to microintervene in hostile online spaces over one month. Participants were asked to keep an eye on FB comment sections of Estonian online news portals and intervene where they perceived the spread of incivility or mis/disinformation. We followed the example of existing grassroots movements such as #Iamhere, which focus on scanning comment sections to engage in ongoing debates in public arenas and decrease hate via algorithmic influence (About Us - #Iamhere Movement, n.d.).

Methodologically, we use the action research framework, trying to both study and solve a problem. Action research is essentially a small-scale intervention, “a powerful tool for change and improvement at the local level” (Cohen et al., 2002, p. 226). We followed the typical cyclical nature of the four-stage action research study process (Zuber-Skerritt, 1996).

First, strategic planning: course design, discussions around ethical nuances, initial problem setting, preliminary research questions, aims, etc. The course was designed with the ethical principles of classroom research kept in mind (Hopkins, 2014). It was also pre-approved by experts from the Centre for Ethics of the University of Tartu, whose suggestions of 1) designing the assignment in a way that does not expose the student, 2) providing an opt-out choice to respect the autonomy of students and follow harm-minimization principles, 3) drawing attention to the risk of the students becoming propaganda mediators, were all included in the course prior to the intervention.

Second, implementing the plan (action): running the course, 12 weekly “mini-cycles” with students (consisting of planning microinterventions, implementation and reflection in the private FB group). Participants were given three options with detailed instructions, named in the task initially: Hero of the Comment Section (visible counterspeech), Secret Participant (small invisible acts of care and “upvoting” reasonable comments) and Theoretical Support (invisible research and analysis in the background). Out of 28 students, initially, 1 chose Theoretical Support, one chose Hero of the Comment Section, and 26 chose Secret Participant track. The above changed during the process.

The third stage of action research embraces observation, evaluation and self-evaluation (Zuber-Skerritt, 1996). In between weekly classes, we communicated in the course’s FB group. The support group for the students was on FB because all Estonian news outlets have a platform and can be equally monitored there. The FB group, with continuous descriptions of the process and participant observations, is our primary data source. We also used written and visual data from the course’s FB group, along with comments, screenshots, and interactions of any sort created by the instructors (n = 5) and students (n = 28). Private FB groups cannot be scraped; therefore, the data was collected by downloading each post via a direct link and then opening all comments (the “Read more...” button) before printing to PDF. The last dataset was self-reflections (n = 24) by the students, submitted directly to the instructors at the end of the course.

The final step of action research is a critical self-reflection to make decisions for the next cycle of research (Zuber-Skerritt, 1996). In addition to weekly reflections, we aimed to provide a structure for deeper critical reflection – becoming aware of perceptual biases, analysing situations students experienced, actively looking for taken-for-granted notions and creating plural structures (inspired by Winter, 1996). Writing this article also serves the purpose of engaging in deep reflection.

Reflections on researcher’s roles

In youth and subcultural studies, researchers have enthusiastically used opportunities that insiderness creates (Taylor, 2011). These advantages for the researchers could be categorized into four broad values: 1) shared experiences (Garton & Copland, 2010; Labaree, 2002); 2) greater access as there is closer and more regular contact with the participants and the field (Taylor, 2011); 3) cultural interpretation, which includes being able to understand the lingo (Taylor, 2011) and certain expressions and codes used; and 4) deeper understanding and clarity of thought (Labaree, 2002).
It is significant that we, the lecturers and instructors, also posted in the comment section and were familiar with the students from prior courses: a mutual relationship developed before the experiment may help build rapport (Garton & Copland, 2010). The problems arise when the participants do not see consent as voluntary, “particularly in the context of coercive relations” (Gallagher, 2009), as the teacher-student relationship is. To minimise this risk, the students were reminded that they were allowed to withdraw from the study at any time.

Data analysis

The analysis is based on, first, the students’ self-descriptions of their actions and feelings in the closed FB group (70 posts with 472 comments by 28 participants). Second, the diverse data built into FB, such as the number of reactions and comments. Third, self-reflections were submitted by the participants (68 A4 pages of data by 24 participants). We only included textual data in this research, i.e. screenshots are not included, as our focus was on the ones intervening, not the content where participants deemed microinterventions a necessity.

Analysis was conducted on thematic qualitative text analysis principles (Braun & Clarke, 2006), identifying “what is common to the way a topic is talked or written about and of making sense of those commonalities” (Braun & Clarke, 2006, p. 57). General research design combined inductive and deductive analytical logic, as the former was necessary to discover new themes and the latter to consider our theoretical reasoning and existing academic knowledge.

The datasets were coded using LiquidText and MaxQDA, starting with ten codes, such as: “microaffirmations”, “microprotections”, “micro-challenges”, “improved MILs”, and “Other Interesting Finds”, the latter serving as the starting point for secondary coding of new themes and ideas as the analysis progressed.

RESULTS AND DISCUSSION

We propose a readily usable educational frame engaging and accessible to the broader publics, following others who have made use of animal metaphors in academic work (e.g., Bame et al., 2013; Talebinejad & Dastjerdi, 2005). For example, “sealioning” is a term used for the practice of engaging in an online discussion with endless demands for answers and evidence, tiring the other (Green, 2022). Or, from a different discipline, Gherzi et al. (2014) described “the meerkat effect” in market research, hyper-vigilant behaviour and focused attention. “Ostriching” has been used for ignoring and hiding from problems (Karlsson et al., 2009), i.e. intended or realised escape from the present situation (Elstad et al., 2021). Johansson and Scaramuzzino (2022) used the phoenix metaphor to discuss netroots organisations’ strategies.

We will refer to three distinct microintervention-type actors: the Worker Bees (in the study task Theoretical Support), the Meerkats (Secret Participants and signallers) and Lions (originally in our task Heroes of the Commentaries), all paired with specific types of microinterventions from Sue et al. (2020). We also focus on the phoenixing cycle as an expected part of information activism.

Worker Bees and microprotections

The Worker Bees’ interventions are the most invisible, low barrier and effort, but essential for resilience-building to work (Sue et al., 2020). This type of information activism was for those who did not want to engage with uncivil content directly but still helped with knowledge-building and supporting others. Worker Bees sought out relevant academic and journalistic literature about various methods used in fighting disinformation (debunking, prebunking, shaming, raising awareness, capacity building, etc.) and summarised this knowledge for others. They constantly built the group’s resilience, preparing others for action on the “field” – strategies for developing literacies and fortifying both individuals and the group.

Theoretically, this is rooted in Sue and others’ (2020) work, recontextualising microprotections in information activism but also stemming from psychological inoculation. According to inoculation theory, it is possible to build psychological resistance against unwanted persuasion attempts, much like medical inoculations build physiological resistance against pathogens (Roozenbeek et al., 2022).

In our case, Worker Bees gave overviews of misleading arguments, persuasion and propaganda techniques, popular disinformation narratives, troll tactics and protective strategies that built resilience. For example, Participant T complained commentators only looked for things they already agreed with rather than being open to discussion. In response, Participant W researched echo chambers for the group and also presented a solution to avoid getting stuck in one:
I have practised joining different FB communities and groups that may not express my opinions entirely, and it has helped me to discover a variety of different perspectives. (Participant W)

Later, Participant C reflected on this newfound knowledge, demonstrating the inoculation has worked, as she has internalised the concept of echo chambers and is considering it in understanding others:

Commentators are in their echo chambers - they agree with those with whom they have the same opinion and who think otherwise must be foolish. (Participant C)

We also consider microprotections in the context of developing a critical consciousness of different information spaces and personal experiences. Here, weekly discussions and sharing evidence in the private FB group proved helpful. Participant I shared her process of developing better skills for dialogue with “The Other”. She posted a story to the FB group on how she answered questions from a friend, who did not end up changing their mind but their choice of sources, and was thankful for the dignified help. This indicates we should learn to navigate information disorder where troll tactics like sealioning – endless questions masked to be sincere but exhaustive by design (Green, 2022) – exist side by side with inquiries from people who need genuine help. It seems microprotections – if the specific case is inspiring and shared within the community – might bring on a range of encouragement and reflection as a response.

Meerkats and microaffirmations

The Meerkat was the prominent choice among the students, as this vigilant first responder is the one who alerts others on where to look or whom to follow. The main tools of the meerkats are behind-the-scenes notifications and microaffirmations – micro-compliments, microsupports, and microvalidations, that help validate people and make them feel supported (Sue et al., 2020). In the context of our study, microaffirmations include liking others’ comments, sometimes writing “agree!” or “they have a good point!” as a reply to a comment, but also reporting and flagging content. The latter could also be a part of the low-effort microchallenges (Lions in our study) – “direct actions that challenge and attempt to disarm, end, neutralise, or deflect the biased behaviour or policy” (Sue et al., 2020, p. 30). Because of the hidden and low-effort nature of these activities, we decided to include them in the Meerkats’ repertoire.

One of the Meerkats’ tasks was looking for comments to “like” to get them to the top of the “Most Relevant” comment sections. Done to nudge the general discourse away from the “recreational nastiness” (Jane, 2014), as several studies (Kim et al., 2021; Gervais, 2017; Suhay et al., 2018) have shown the framing effect of toxic comments – incivility spreading from comment to comment. Participants admitted to previously scrolling on without focusing, being disconnected from the text, and labelling comments “unnecessarily rude and demeaning”; something to be ignored, not actively challenged – a fatalistic approach to the algorithmic influences of social media platforms.

We aimed to decrease the disconnection levels of social media users (Cover, 2022), thus needing civil and constructive comments to be visible and algorithmically placed at the top of the comments sections. Participants engaged in conversation about the algorithmic logic of FB, trying to understand which microinterventions might be most effective and how the algorithm determines which comments to show first. The majority of participants expressed negative, divisive and polarising comments are on top but also highlighted when a comment section had been intervened in early and had created - as per previous studies (Gervais, 2017; Kim et al., 2021; Suhay et al., 2018) – an opposite framing effect, i.e. if a comment section started with positive comments on top, more would follow.

Microsupports were also presented to commentators with high MILs as positive examples, and participants grew more and more analytical of others, improving their MILs as a result:

I noticed a connection between the framing of the headlines. If the people in the headline were anonymous or an institution, the comment sections would have fewer personal attacks. There seems to be an effect between how the people are framed in the article and who is likely to be portrayed as the culprit or hero in the comment section. (Participant D)

The rapid evolvement of MIL skills was one of the most important results of the Meerkats’ work. The hyper-vigilance that made Meerkats excellent team players also helped train MIL skills further with informal peer-to-peer learning. In the end, participant J stated that noticing the content of the comment sections had become automatic, second nature. That is an improvement in MILs we hoped for when designing the course: a more active approach to social media use and participation.

Reporting uncivil comments and finding something to “like” put more pressure on the participants’
emotional well-being. However, in response to such reflections, participants in our closed FB group were encouraging and consoling each other.

I suppose everyone has realised that the comment sections of certain news stories are not suitable for the faint of heart. […] I kept my eye on the comment sections but did not want to waste my nerves by focusing in-depth and on the daily. (Participant B)

Participants with Russian mother tongue \( (n = 2) \) started ironically dubbing themselves the same derogatory terms that Russians were called in the Estonian comment sections. Other participants reacted swiftly, validating feelings and reiterating the individual injustice due to their ethnic roots.

**Lions and microchallenges**

Lions are visible in the comment sections, confronting biased information, challenging societal practices (Sue et al., 2020, p. 39), spreading counternarratives, fighting off trolls and voicing opinions. The most visible form of microinterventions, therefore, also inherently carries the most risk and needs the most civic courage.

I have a fear of reacting (by commenting) because I am afraid of getting that hailstorm of anger in return. At the same time, during this experiment, we were all "on the same battlefield," and there was really nothing to fear. As was constantly said during the course, everyone would have politely come to the rescue if there ever was such a situation. (Participant E)

This was also the case for Meerkats and Worker Bees, who expressed admiration for the patience and courage "to respond to comments like "everyone lies and is evil and tries to brainwash us" (Participant R). However, some topics discouraged microchallenges, with participants reporting an inability to be unbiased when faced with hateful comments about minorities or uncertainty in their competence when it comes to complicated and detailed discussions about the invasion of Ukraine. Challenging views did not come naturally to any of the participants, becoming a habitual calculation: how much time they had, how their mental state was, and how confident they felt in their objectivity and knowledge of the topic. Lions also showed signs of the overspill of civic courage to other areas, with one habitual Lion posting to the FB group:

I gave a lesson on information warfare to 10-12. Grade students today. I also discussed micro-interventions and provided many of the same examples I shared in this group. […] In addition to the guest lecture, I also wrote a news piece on if/how/why this topic is being taught to pupils in Southern Estonian schools. (Participant N)

Additionally, what else stood out in the reflections of the Lions was the amount of detail in the descriptions of the comments and commentators they chose to challenge.

I was amazed by the number of people who have no control over themselves and their use of words. However, it was no surprise that those who hurled negative and inappropriate comments knew no grammar rules. Commas, points, spaces - nonexistent. (Participant K)

The participants did not, however, focus on the content, political opinions, religious values or other aspects in the comments that are up to individual choice and preference. This is the key issue to keep in mind for similar projects, study activities and lesson plans in the future: the focus needs to remain on the form of the comment, the civility, the quality of sources and the factual basis of the arguments, not facilitate a unipolar understanding of the world, as researchers, lecturers or - why not - moderators in FB groups, our aim should be to highlight the realities of today’s online communication culture.

The real challenge was the high effort of this style of intervention, which can lead to burnout and hinder future activism. The data also helped us understand the cyclical nature of digital activist activities, as most students ended up switching up their initial intervention path according to what we consider an inevitable cycle in digital information activism. It must be highlighted that the changes between roles, were not designed into the course but happened as an organic decision of the students. Looking back, the role switch can be anticipated, especially within a longer time frame and thus would be advisable to incorporate it in the study design early on.

**The Phoenix cycle**

Initially, participants were shocked and vulnerable by the surprising elements observed in the comment sections. However, as they gained exposure and developed coping strategies, as the support in the form of microaffirmations or microprotections from their peers grew, they began to find joy in reasonable argumentations and highlight overwhelmingly positive comments. Looking for something to “like” and something to report shifted the participants’ focus on the
positive and negative extremes, but the exposure also built resilience, empathy and coping strategies.

Students often began the experiment with flight/freeze responses – i.e., ignoring and hiding from negative content and comments, as well as self-censoring their reactions. To draw from the immortal phoenix from Greek mythology: once the phoenix started regenerating, or our participants moved on from flight or freeze to fight stages, a pathway emerged – moving on from initial despair towards feelings of joy.

I found it joyful to see some people at least try to call out or doubt commentators sharing false info or blundering. I “liked” all comments like that. It could help tone down the spread of false information to some extent or at least let other readers know that it might not be all true. (Participant E)

This cycle also brought on different strategies to deal with the flight or freeze mode, like only looking for positive news headlines – of babies being born in the rubble of war, for example – with the expectation of finding more positive content. Avoiding specific topics, like the war and (foreign) politics, was still a better strategy than completely giving up on comment sections. On the other hand, Participant H chose exposure therapy, picking headings that most likely stir up a storm in the comment section. The search for an experience of success is not limited to finding comments to “like” but also relates to wanting a positive and empowering experience. The less successful their interventions were self-construed, the more the participants seemed to return to flight or freeze strategies.

In the case of more extended periods, “activist burnout” is to be expected - deteriorating the emotional and physical health of individual activists (Gorski & Chen, 2015). Similarly to Gorski and Chen (2015), we successfully anticipated how Lions might crave safe spaces to discuss their burnout and strategies to overcome them. The community support and the shared experiences positively affected students’ outlook, pushing them into another cycle of “phoenixing”.

CONCLUSIONS

This study used an action research approach to collect data from exercises used in a MIL training course. The aim was to understand what types of information activism were observed during the study process, what strategies were used and how the collective interventions affected the participants’ MILs. This enabled us to conceptualise three types of information activism roles: the Worker Bees (focused on microprotections), the Meerkats (microaffirmations) and Lions (microchallenges), which surfaced in the phoenixing cycle. Results indicate that participation in the experiment, whether as visible participants or not, increased the students’ understanding of the public discourse space in FB comment sections and individual MILs.

Many participants admitted to previously scrolling on without focusing on the comments, labelling them “unnecessarily rude and demeaning” and expressing worry of not finding anything to “like” as the general culture of discussion was overwhelmingly negative. This highlights that participants had given up on public discussion spaces, presenting a fatalistic approach towards their power to tip the balance of incivility online, signalling the platforms are too big and too powerful for individuals to be able to resist.

We highlight four significant results of the study. First, the Meerkats and microaffirming (liking, upvoting and empowering) created a more positive outlook for the participants on online public discourse spaces than looking for uncivil comments to report, i.e. what one looks for, one will notice (and be served by the algorithm). Second, rather than a specific strategy, the sense of community and support is the secret ingredient to feeling empowered enough for microinterventions, especially in “toxic” discussion spaces. Third, participants’ MILs improved as they reflected on something reminiscent of an inverted Broken Windows Theory: if the group is amongst the first in the comment section and leads by example, then a seemingly minor positive comment will contribute to a warm atmosphere and encourage more constructive discussions. Still, they also developed the habit of assessing the MILs of others in the comment sections, trying to reason why the Others spread mis-, dis- or malinformation and developing a self-preservation tactic in information activism. Fourth, the experiment proved helpful in improving individual “fight” strategies, carrying the participants through the inevitable cycle of information activism, “phoenixing”, from one regeneration to the next one.

We argue that microinterventions can be used as a tool in higher education and informal MIL trainings, as they support attentiveness towards information disorders, empower one to break the cycle of silence in online discussions, and teach the concept of algorithmic influencing. Further studies could test our model’s applicability in different contexts, e.g. community or workplace collaborative interventions. This could help
to answer the questions regarding motivations behind exchanging between different roles.

ACKNOWLEDGEMENTS

The design, implementation and publishing of this research has been supported by the IREX Media Literacy in the Baltics Program and the EU-funded project BECID (ID 118471).

REFERENCES


and young people: Research design, method and analysis. Sage Publications.


Lindhult, E. (2019). Scientific excellence in participatory and action research: Part II. Rethinking


Pasta, S. (2023). Social Network Conversations with Young Authors of Online Hate Speech Against Migrants. In A. Monnier, A. Boursier, & A. Seoane (Eds.), *Cyberhate in the Context of Migrations* (pp. 187-214). Springer International Publishing. https://doi.org/10.1007/978-3-030-92103-3_8


