

Spence, R., Bifulco, A., Bradbury, P., Martellozzo, E., & DeMarco, J. (2023). The psychological impacts of content moderation on content moderators: A qualitative study. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 17(4), Article 8. <https://doi.org/10.5817/CP2023-4-8>

The Psychological Impacts of Content Moderation on Content Moderators: A Qualitative Study

Ruth Spence, Antonia Bifulco, Paula Bradbury, Elena Martellozzo, & Jeffrey DeMarco

Centre for Abuse and Trauma Studies, Middlesex University, London, UK

Abstract

Content moderation generally refers to the practice of monitoring user generated content and applying a set of rules defining what is acceptable. As a result, Content Moderators (CMs) are responsible for reviewing and removing harmful online material, which has the potential to cause psychological harm. However, little is known about the perceived impacts of the job. Semi-structured interviews were conducted with 11 commercial CMs who were exposed to child sexual abuse material (CSAM) as part of their job. Interviews were analysed using framework analysis to investigate any psychological impacts. Results demonstrated that CMs manifested with a range of symptoms consistent with experiencing repeated trauma. This included experiencing intrusive thoughts of CSAM, triggered by situations with similar contexts to those encountered at work, avoidance of children and negative cognitive and emotional effects such as cynicism, anxiety and detachment. These symptoms fit into a framework of post traumatic and secondary traumatic stress and suggest CMs wellbeing and exposure to stressful stimuli may be comparable to professionals working in the emergency services or caring professions, such as social workers. These personnel are at heightened risk of various problems including compassion fatigue, vicarious trauma and burnout. Companies which employ moderators should learn from these comparable professions and provide psychoeducation and trauma-informed care to CMs which have been demonstrated to help mitigate these challenges.

Keywords: child sexual abuse material; content moderation; trauma; post-traumatic stress disorder; work stress

Editorial Record

First submission received:
November 27, 2022

Revisions received:
March 29, 2023
July 14, 2023

Accepted for publication:
August 7, 2023

Editor in charge:
Lenka Dedkova

Introduction

Social media and the internet have become ubiquitous. It is estimated that nearly three billion people use Facebook each day (Facebook, 2022b), while Instagram has 2 billion active users (Statista, 2022). However, with the freedom to share information comes the ability to upload and distribute more troubling content. In the second quarter of 2022 alone, Facebook actioned 20.4 million pieces of child sexual exploitation (Facebook, 2022a). The increasing amount of user generated uploaded content means that companies need to hire more and more content moderators (CMs), these are the individuals who evaluate and remove material that is flagged by users as harmful or in violation of platform policy.

There is a wealth of evidence that exposure to other people's trauma at work affects wellbeing and mental health (e.g., Léonard et al., 2020; Ralph, 2020) and research focusing on CMs suggest they suffer from impacts similar to those found in other professionals (Steiger et al., 2021). However, despite a growing number of businesses

employing moderators (Fact.MR, 2022), research attempting to evaluate the impact on CMs has largely focused on volunteer moderators (Cook et al., 2022; Schöpke-Gonzalez et al., 2022). It is likely that volunteer and commercial CMs have distinct working conditions, likely leading to experiencing the role differently, in that it may be associated with different expectations, training, and levels of support (Burns et al., 2008).

The present study was designed to qualitatively explore any psychological impacts in commercial CMs. This was to illustrate the potential difficulties experienced within their roles, while also providing information for more effective work practice in this field. The use of in-depth semi-structured interviews allows participants to express themselves freely and articulate a diversity of experiences. The results can be used to inform organisations about how the work may affect employees and therefore what types of support should be put in place. Indeed, without a sound understanding of any specific psychological effects on CMs, attempts to ameliorate or counter any impacts could be misguided or even harmful (Wessely et al., 2000). This would impede efficiencies within the role and could adversely affect staff retention and recruitment. In the remainder of this introduction, a review of the different work environments populated by CMs and how the work CMs undertake might give rise to psychological difficulties is considered.

The Work of a Content Moderator

CMs enforce the online rules which tell users how to behave and what content is acceptable on a particular internet site (Gerrard, 2022). Some sites use a “community reliant approach” where communities define their own standards that are then put into effect by volunteer moderators, whilst other sites use an industrial approach, where workers are employed to enforce a set of standardised rules (Caplan, 2018). As such, the ways in which CMs are employed vary. Whilst some are volunteers, paid moderators can range from in-house workers employed directly by the company needing moderation; to boutique firms specialising in content moderation for other companies; to outsourced third party vendors and microlabour platforms (Roberts, 2019). This means CMs can experience a range of working conditions where some have less workplace protection as technically, they are not employees - thus the company requiring moderation work has a level of “plausible deniability” to the harm faced by these CMs (Barrett, 2020; Roberts, 2019).

Nevertheless, CMs play a critical role in protecting firms’ reputations, ensuring online platforms do not become unusable, hostile spaces that are unprofitable for these business (Drootin, 2021). Furthermore, they shape what is seen online as they make hundreds of daily decisions, and for some this can be with limited oversight (Parks, 2019). In 2014 Chen estimated there were over 100,000 paid CMs worldwide. It is likely that in the intervening years this number has increased. In 2021 Messenger and Simmons reported Facebook and Instagram had expanded their moderation workforce to 15,000 people, and companies continue to hire more with a proliferation of businesses that specialise in brand protection and moderation (Fact.MR, 2022). These figures do not include the numerous unpaid CMs used by sites like Reddit (Witynski, 2022).

As part of their role, material deemed offensive or harmful is sorted into queues and sent to CMs for review where they must make decisions about suitability in line with the set of policies they are currently enforcing (Ahmad & Krzywdzinski, 2022). CMs can be required to view depictions of child sexual abuse and violence, as well as less graphic but no less noxious content that displays cruelty, humiliation and discrimination. Thus, moderators engage in emotional labour as they are expected to regulate their own emotions when exposed to this material, as well as during interactions with users as they work to represent the platform (Steiger et al., 2021). The content they view can range from text, to still photos to explicit video and the amount of time spent viewing material varies; it can however be substantial in terms of hours every working day (Dwoskin, 2019). In addition, CMs have to maintain high accuracy levels and throughput quotas to meet acceptable job performance standards, which contribute to stress above that which might be experienced through viewing the content alone (Dwoskin et al., 2019; Wohn, 2019). Dwoskin et al. (2019) described how moderators had their performance regularly audited and were expected to maintain an accuracy score of over 95 percent when identifying and correctly removing harmful content in accordance with company standards.

There is a body of evidence highlighting the consequences of chronic job-related stress or exposure to difficult or traumatic content at work being serious and long-lasting (Vasconcelos et al., 2021). Research with front-line professionals such as first responders has revealed that exposure to other people’s traumatic experiences can lead to post-traumatic stress disorder (PTSD), a disorder that can develop after exposure to actual or threatened death, serious injury, or sexual violence either directly, through experiencing or witnessing the event, or experiencing repeated or extreme exposure to aversive details of the traumatic event(s). It is marked by intrusion,

avoidance, negative symptoms and hyperarousal (American Psychiatric Association, 2013). Professionals in caring professions such as social workers and therapists who are exposed to second-hand accounts of trauma can develop syndromes such as secondary traumatic stress (STS), vicarious trauma (VT) and burnout (Craig & Sprang, 2010; Greinacher et al., 2019; Lee et al., 2018). These have been referred to as the “cost of caring” and relate to the stress reaction experienced by someone in response to another’s distress (Figley, 1995). STS has similar symptoms to PTSD such as intrusive thoughts, hyperarousal and hypervigilance, as does VT, although it is also associated with a transformation of self-image and worldview (Foley & Massey, 2021; Krause, 2009). Burnout is characterised by emotional exhaustion, cynicism and reduced feelings of accomplishment (Brady, 2017; O’Malley et al., 2019).

This is also true for professions not traditionally considered front-line staff. For example, journalists exposed to traumatic events while working with difficult content such as child abuse, child cruelty, war and aviation accidents, report higher levels of PTSD and the internalization of guilty cognitions (Browne et al., 2012). Drone pilots involved in military operations have reported higher levels of burnout, and an increased likelihood of developing anxiety, depression and PTSD (Asaro, 2013). Judges and lawyers report sleep disturbances, feeling isolated, concentration problems and interpersonal issues (Jaffe et al., 2003; Léonard et al., 2020). Lastly, police officers who specialise in internet child exploitation and are repeatedly exposed to images and videos of child abuse as part of their investigations show raised rates of STS, mental health difficulties and absenteeism (Bourke & Craun, 2014; Ralph, 2020). Furthermore, disorders such as PTSD, STS and burnout are often comorbid with other mental health problems such as depression and anxiety and in some cases their onsets are linked (e.g., Hakanen et al., 2008; Schindel-Allon et al., 2010).

Anecdotal accounts from investigative journalists suggest the work of CMs places them at risk for a range of psychopathological difficulties. CMs report anxiety, depression, nightmares, fatigue and panic attacks, with their relationships and physical health also affected (Newton, 2019, 2020b). Both Microsoft and Facebook have been sued for substantial sums by CMs who developed PTSD on the job and held the companies responsible (Ghoshal, 2017; Newton, 2020a). Similarly, academic research highlights high stress levels and fatigue (Cook et al., 2022) and symptoms of insomnia, anxiety, low mood and experiencing intrusive images (Benjelloun & Otheman, 2020) and work by Roberts (2019) found that although CMs would report not being negatively impacted by the work they would demonstrate via anecdotes that the work was affecting them and their relationships.

More broadly, moderators have reported feeling burnout and emotionally exhausted (Dosono & Semaan, 2019; Riedl et al., 2020; Schöpke-Gonzalez et al., 2022), apathy (Lo, 2018) and underappreciated (Wohn, 2019), as well as being harassed by users or feeling guilty for enforcing rules or taking actions they do not necessarily agree with (Dosono & Semaan, 2019; Wohn, 2019). Moral injury occurs when someone acts, fails to prevent or bears witness to acts that transgress their moral beliefs (Litz et al., 2009) and therefore for some, the work of CMs could manifest as moral injury. Indeed, a mismatch between the community and moderator’s values is associated with quitting in volunteer moderators (Schöpke-Gonzalez et al., 2022).

Nevertheless, some individuals inevitably are more susceptible to trauma than others and there is individual variability to the degree of negative impact as a result of the role (Powell et al., 2015). People exposed to trauma can experience post-traumatic growth, whereby their coping leads to a greater sense of purpose and life meaning (Steiger, 2020). A survey of internet child exploitation police officers found they had higher wellbeing than a normative adult sample (Tomyn et al., 2015), and similarly, a study of volunteer moderators found low levels of STS (Schöpke-Gonzalez et al., 2022). Organisational support can also raise feelings of compassion satisfaction where the focus rests on the benefits rather than costs of the job, and positive feelings are experienced when helping someone who has been traumatised (Brady, 2017; Hunsaker et al., 2015). This can act as a buffer against compassion fatigue, STS and burnout (Brady, 2017; Conrad & Kellar-Guenther, 2006). Thus, the role of CM may produce significant stress for some, but not all.

The evidence suggests that content moderation is associated with a variety of negative impacts, although it is possible for some individuals the work will be associated with positive effects. To date, there have been few academic studies that explore what impacts the demands of the job might engender and many that have been conducted have focused on the experience of volunteer CMs (e.g., Dosono & Semaan, 2019; Lo, 2018; Wohn, 2019). However, volunteer moderators may experience different types of psychological distress due to their distinct motivational factors and the range of activities they engage in which commercial CMs do not (Matias, 2019). The current study aims to qualitatively explore the impact of the job on commercial CMs, with the objective of using these results to highlight areas where problems may arise. Specifically, the study set out to investigate a) what, if

any, behavioural changes moderators may experience due to their job and b) how, if at all, the work affects their relationships with others. The findings can act as a first step to guide larger scale studies on the impacts.

Methods

Participants

Eleven CMs with experience of moderating CSAM were recruited. Initial recruitment was through online advertising on social media platforms and utilising existing networks, with further recruitment relying on a snowball technique where interviewees provided the research team's details to potential participants. Participation was voluntary and CMs were not remunerated for their time. Moderators who wanted to participate were given an information sheet which explained the study and outlined that they had to have experience of moderating CSAM. All participants were given the opportunity to ask questions before, during and after involvement. There was no minimum requirement with regards to how much exposure a participant had as the study was interested in exploring various CM experiences. All participants spoke English but for the majority it was not their native language. Specific employment-related details such as the participant's employer or specific sites they may have moderated were not requested, however at the beginning of each interview participants were asked to outline what their job entailed and confirm they were exposed to CSAM as part of their work. All participants worked full-time as CMs either for commercial companies or CSAM specific hotlines. The sample included six females and five males, from eight countries: the Netherlands, UK, Ireland, Germany, Brazil, Colombia, Malaysia and India.

Procedure

All participants took part in an online interview, and could participate with their camera on or off. Interviews were administered by two of the research team and were conducted from April to June 2022. They ranged in duration from 47 to 78 minutes, with average interview time of an hour. The interviews all followed a similar structure with participants asked about their role, how the material affected them, how they coped, other occupational stressors, how the job was perceived by others and the impact of COVID on their working conditions. This paper focuses on data relating to impacts of the work. Interviewers asked open-ended questions and follow up probes were used as needed to elicit further information and clarify answers. Informed consent was recorded verbally before the interview commenced. Ethical approval was granted for the study by the university's psychology research ethics committee (ref: 20888). The interviewers did not specifically ask about non-disclosure agreements (NDAs; contracts CMs may be required to sign that prohibit them from talking about their work) but anonymity and confidentiality were guaranteed through the use of participant codes, and participants personal details were not stored. All data was secured on password protected drives in line with Data Protection and UK GDPR regulations and the interview recordings were deleted after transcription.

Data Analysis

All interviews were audio-recorded, transcribed verbatim and anonymised through removal of identifiable information. A Framework approach was used to analyse the data (Gale et al., 2013). This enables in-depth exploration of the data whilst making the process of data analysis transparent and is particularly suited to the analysis of cross-sectional descriptive data (Ritchie & Lewis, 2003). It involved six stages, *Familiarisation*: the transcripts are read repeatedly to enable familiarisation with the data and identify key themes. *Coding*: descriptive or conceptual labels are systematically applied to interesting or notable excerpts of raw data. The coding framework is then discussed and agreed by the research team. *Developing an analytical framework*: codes are grouped into categories representing interrelated ideas or concepts. *Applying the analytical framework*: the codes and categories are applied across all the transcripts. *Entering the data into the framework matrix*: a spreadsheet is used to generate a matrix. In the matrix, each column represents a key category and each row represents an individual participant. The summarised data from each transcript are entered into the matrix so each cell represents a category by participant. *Interpreting the data*: the final analytic stage involves working through the data in the matrix, drawing out the range of experiences and views expressed during the interviews, identifying similarities and differences across participants, and interrogating the data to seek to explain emergent patterns and findings.

Results

The CMs reported a number of different impacts across a variety of domains, particularly emotional and cognitive. They also discussed factors that exacerbated work stress aside from the content. The identified themes and sub-themes are presented in Table 1.

Table 1. *Themes Identified Through Framework Analysis.*

Superordinate Themes	Sub-Themes
Impact Awareness	Barriers to talking about impacts, Harmful Content
Detachment	Desensitisation, Keeping a distance from victims, Suitability, View of Self, Sensitivity to Some Material
Negative Cognitive Changes	Thoughts of CSAM, Hypervigilance Regarding Children, Incompatible with Family Life, World is Darker Place, Cynical Views of People
Emotional Toll	Anger, Anxiety, Mental Health, Fatigue, Sleep Disturbance, Self-Care, Effects Relationships
Stressful Work Environment	Respect, Ill-Equipped, Quotas

Impact Awareness

Regardless of if CMs felt they were personally affected by the work, there was an awareness amongst all the moderators that the content they were being exposed to was harmful and had the potential to affect them long-term.

“I think the turnover’s always higher in the more problematic content. We didn’t see that much turnover in terms of bullying or election misinformation or adult nudity because those policies are, that kind of content doesn’t really traumatise you as much as the other ones.” (CM8)

Nevertheless, some CMs did assert they were unaffected by the job. However, further probing always elicited either current or past examples of difficulties. The reluctance to talk about impacts was notable and possibly reflects attitudes towards the work, such as not wanting to talk about the difficulties, because they agreed to take on the job and accepted that exposure to content was part of their profession. Additionally, many effects may come on gradually and may be hard for CMs to recognise, quantify and put into words. CMs are aware that something has changed and can talk about being affected in general, but find it difficult to identify specifics. “It has an effect on you detrimentally when you’re on your own. It affects your imagination in ways that you can’t quantify, but it still affects you.” (CM3)

Therefore, although all CMs gave examples of how they had been negatively affected by their role, the level of distress caused and insight about effects varied. A couple reported they could not identify any impacts within themselves and viewing traumatic content was “just their job” but noticed impacts in others, or gave examples of impacts that occurred in the past, possibly reflecting that any effects or symptoms associated with the work change over time for some CMs.

Detachment

Associated with change in effects over time was the ability of CMs to detach from the material. It seemed after the initial shock of viewing content, a common way to cope with the feelings evoked was to shut down emotionally. This ability seemed to develop over the first few months in the job and enabled CMs to look at content without experiencing any adverse reactions. “They kind of get used to see those tickets and to see that kind of content and it’s like they transform into robots and suppressing their emotions and stuff like that.” (CM7)

In particular, CMs tried to keep an emotional distance from the victims. This both involved remaining analytical, but also a conscious effort not to think about the children as people or think about what had happened to those victims both before and after that particular piece of content was created.

“I’m just trying to keep my mind to analysing the material, and it’s been working so far... It helps me because then I don’t remind myself that it’s actual people and actual stuff that happened. It helps me to just think about it as something like a TV series that they’re analysing or whatever”. (CM1)

However, others struggled with maintaining that boundary and could not remain detached from the victims. This resulted in them subjecting themselves to the content more than they had to as a result.

“That child, when on his own, showed his trauma and as a reviewer I kind of think this is my responsibility to make sure that, even if I do it years after the event happened, it’s my responsibility to make sure that the child in some way wasn’t on his own going through this trauma.” (CM9)

At the extreme end the emotional blunting could turn into desensitisation, where CMs became anaesthetised to what they were witnessing. Although perhaps an adaptive response overall, it made a couple of CMs question themselves and if their ability to do the work meant there was something wrong with them. However, it may be that those who were unable to turn off their emotional response in this way did not last long in the job.

“If you see a child being sexually abused how do you look at this in a numb way or can you ever look at it? It leads to a lot of burnout. A lot of people in the child sexual exploitation or terrorism spheres particularly have a quick turnaround in terms of the work that they do because it’s traumatic for them. It’s not something that people want to view.” (CM8)

Desensitisation did not occur across all content. Some CMs had personal sensitivities and reported certain types of content continued to be upsetting, even after they had adjusted to what they were being exposed to overall. This could be both in terms what was in the content but also across different types of content, e.g., text rather than images was more disturbing for some. “My trigger is bestiality. I cannot work on the bestiality content, but I had to do it at some point, and it was really tough for me because I have my dog.” (CM5)

Negative Cognitive Changes

CMs experienced intrusive thoughts and discussed how images of CSAM remained in their minds after initial exposure. The intrusive thoughts were generally associated with feelings of psychological distress or unease.

“So that was quite disturbing for that moment. Yes, I did take about two or three days off from work after that... I was not physically ill. It just was immediate that I couldn’t take off my head from that.” (CM6)

CMs reported that intrusive thoughts were particularly prevalent during, and indeed triggered by, sexual relations with their partners. This often resulted in them having to stop and/or deliberately distract themselves from thoughts of CSAM. This could leave them questioning their sexuality and reduced feelings of intimacy within the relationship.

“Same with my own sexuality, sometimes I say to my boyfriend, just let me have a minute for myself and then I’ll try to bring my thoughts to nice thoughts and then it’s okay, but that wasn’t there before. That’s new since I started the job.” (CM2)

The presence of children also appeared to trigger thoughts of CSAM and worry about child welfare. In a couple of cases this even led to CMs trying to avoid children, especially situations where a child might become upset as this was particularly distressing and triggered or intensified intrusive thoughts of CSAM. CMs who had children particularly struggled with the content and it was reported that staff often left the job when they were either planning on or having children of their own, demonstrating a recognition that having children would impact their ability to do the work. CMs in general demonstrated heightened concerns about child safety and feeling more protective of children.

“When my aunt had her child... I tried to explain to her what’s happening, what’s actually happening... and that you need to be more attentive as a parent when you’re leaving your child even to a simple game. It’s not just a game anymore.” (CM7)

More general changes in cognition were also evident, with the exposure to disturbing content leading to the development of a more negative world view. CMs talked about having lived in a bubble or having been sheltered but that their job had revealed another darker side of the world. This was also associated with the development of a cynical view of people and their behaviour. This general distrust of others experienced by CMs could result in difficulties with interpersonal relationships. In particular, a couple of female CMs expressed concerns about men in general as they developed suspicions of their possible involvement in perpetrating or watching CSAM. As a result, some described experiencing a more general loss of hope and fewer positive emotions.

"But sometimes I feel like my joy has been taken away from life. It's not in a very serious way but, I don't know, I just see the world a little bit less colourful. Yes, it's a bit difficult to explain, but I guess I know now how dark the world can be." (CM1)

Emotional Toll

The work also took an emotional toll on CMs. Many talked about feelings of anger or frustration and this was discussed generally in the context of the content, feeling anger when they thought about what they had seen or the people that were involved. However, a couple were more explicit in saying their anger was more generalised than that. CMs also spoke about physical affects experienced at work, with one even describing having panic attacks. The descriptions offered varied but all were indicative of experiencing a "fight or flight" response to the content.

"Sometimes I feel not exactly nervous but like my body is very awake but not in the sense of not being tired, more like being a little, maybe shaking a little but not shaking on the inside. More like nervous but without a reason when I see very much content or content that is very disturbing." (CM2)

In general though, CMs did not articulate the exact emotional affects. They tended to talk in generalities about mental health issues, getting "sick" due to the content and experiencing emotional stress. This suggested the job might be associated with a general malaise rather than specific problems.

"He just said, 'I knew that there would be some really gross content, but I really didn't expect this.' Honestly, I didn't have really words to tell him. I mean, literally, I just go and tell him that that's the part of work that you're feeling." (CM6)

They also discussed feeling fatigue at work and being drained or exhausted by the work. This might be a slightly different way of expressing the malady experienced by workers in this role. This weariness could be linked to the emotional effects of being exposed to content, or a business need in meeting quotas. "When I started here, I actually was off sick for 60 per cent, because I had no energy and I wasn't feeling good. I couldn't even stay awake when I was here." (CM1)

CMs reported sleep disturbances, such as nightmares, which often were connected to the violence they had witnessed during work. There was also difficulty falling or staying asleep, which may have contributed to day time tiredness. This was due to replaying images as they were trying to get to sleep or waking up for fear of being asleep.

"I started to have sleeping problems, I don't know it was really irrational, everything comes to me and I had problems sleeping, I was afraid to sleep for a week. I wanted to sleep, but I didn't sleep well, that week I would sleep 2 or 3 hours. I would wake up and be like no no, I wasn't sleeping, no I have to be awake." (CM4)

CMs also mentioned that self-care suffered as a result of the work. This was attributed to stress, whether caused by the content itself or having to hit specific targets and quotas, individuals developed ways to comfort themselves at work.

"I know of many people who say that they do eat a lot of chocolate while working. Chocolate is one; coffee is another... Thirdly, smokers who smoke, they tend to take a lot more smoking breaks... One is to help them cope with what they see. Secondly, the stress of meeting their daily numbers." (CM6)

CMs felt unable to talk to friends and family about their work. They felt that others did not want to know or that they would be traumatised if they did know. However, for some this was associated with a feeling of disconnection because not only were they shielding others from what they had seen, but also they were hiding the affects the job could have on them. For some this contributed to a sense of isolation.

"For my grandparents, I'm a researcher. That's all that they know about me, they don't know more. My parents know that I'm working on CSAM but I don't tell them anything else. Same thing for my partner, but because of this, when you're having a bad day, when you have your emotions all over the place, you still have to pretend that you're okay, that everything is fine." (CM9)

Stressful Work Environment

The emotional toll involved in the work could be exacerbated or alleviated by the response of management, and this seemed to hinge on whether managers respected the role that CMs played within the company. This was often more likely if leaders had direct experience of the work and could empathise with what CMs experienced, both in terms of content but also the targets that CMs had to hit.

"I got affected by the management when they started pressuring you. Lots of things, and you didn't even realise that how much we are affected by the content as well. It's really disturbing content what we're working in, specifically in a minor content, but still, the managers didn't care about that. They were like, 'Now, oh, your performance this way. We need to extend something, or...'" (CM5)

Quotas were also mentioned by many of the CMs as something which added to a feeling of pressure or stress on top of what they were witnessing. This was both the levels of accuracy that had to be maintained and throughput. This too, was often linked to the approach of management with a feeling that their focus was on reaching the quotas with no appreciation of the wider context of the job.

"'You're not getting your quota.' Bring you into the office, '[CM's name] you need to improve on this, this, and this,' and it's like, 'Woah, hold on a second, you should watch this stuff.' There's a bit of a divorce going on there, a disconnect." (CM3)

Lastly, CMs felt that the tools they had to work with often were not designed with content moderation and their role in mind. This added to the frustration and stress associated with their job because they felt it prevented them from being able to do it effectively.

Discussion

Exposure to other people's trauma at work can be detrimental to employees (Greinacher et al., 2019; Lee et al., 2018) and to the efficient fulfilment of their work. Existing literature suggests that this is also the case for CMs (Cook et al., 2022; Schöpke-Gonzalez et al., 2022). Research to date examining the psychological effects of this work is sparse and tends to focus on the experience of volunteer moderators who may have a different set of motivations and job-related tasks (Matias, 2019). The present study adds to this literature by qualitatively exploring the psychological effects of content moderation in commercial CMs. The results show both negative cognitive and emotional changes take place and that these are associated with behavioural changes such as avoidance and hypervigilance. Overall, the effects are consistent with commercial CMs experiencing a trauma response in relation to their work, and provides evidence demonstrating why organisations employing CMs need to provide trauma-informed care for their employees.

Similar to previous research, although CMs knew the content was difficult and that people struggled, there was a reluctance to talk openly about the personal impacts of the job (Roberts, 2019). There are several possible reasons for this, such as: a lack of awareness around how they are being impacted or the inability to put the impacts into words. Syndromes such as VT and burnout are cumulative and gradual, this can make it difficult for individuals to recognise any changes within themselves or others (Krause, 2009; Ledingham et al., 2019). Alternatively, CMs may not want to admit to impacts either because they agreed to the job or because as Roberts (2019) argued, adverse effects may suggest they are unable to perform the job adequately. Equally, the participants all worked full-time as content moderators and therefore some may have signed NDAs, which can be associated with a fear of speaking out about the job (Schiffer, 2021).

Nevertheless, some CMs did appear to be coping well. Again, there are several possible reasons for this: some CMs reported they were more sensitive to particular types of content. It could be that certain people, for example those who do not interact with children in their daily lives, are more resilient to working with CSAM material. Similarly, people who stay in in this profession for a long time, may be the ones who are affected less, cope relatively well and experience little in the way of impairment, or even undergo a period of post-traumatic growth. There appeared to be an initial period at the beginning of the job whereby people either tended to adjust to the role or leave. Therefore, CM resilience may be dependent on the content they are exposed to, or may change over time. However, the adjustment seemed to be associated with the ability to suppress emotions, which may have longer-term consequences for social and mental wellbeing (Chervonsky & Hunt, 2017; Hu et al., 2014). Future

research should investigate what predicts resiliency in CMs, but also the potential long-term effects both in terms of ongoing resiliency and any possible deleterious consequences of this.

Many of the impacts described by CMs are in line with the symptoms described by volunteer moderators (Dosono & Semaan, 2019; Schöpke-Gonzalez et al., 2022). The symptoms are also reminiscent of those occurring after repeated or extreme exposure to aversive details of trauma as described by PTSD or STS, and the majority of themes could be grouped in this way (see Table 2). When CMs gave examples of how they were affected, they talked about intrusive thoughts and images of CSAM including that these were triggered by children and their own sexual activity. Additionally, some described avoiding children and there were a number of negative cognitive and emotional affects, including hypervigilance around children, anger and increased distrust of others. There were also symptoms of hyperarousal, presenting as sleep disturbance and bodily sensations of anxiety. However, the interviewers asked open ended questions regarding possible impacts. Specific questions that could be used to assess or diagnose different disorders were not utilised. More research is needed to investigate whether CMs reach the threshold for any potential diagnoses, to what extent symptoms reflect one or many different potential syndromes such as burnout, PTSD or VT and to estimate their potential prevalence in this population.

Nevertheless, this study suggests the psychological impacts experienced by commercial CMs are comparable to those experienced by professionals such as first responders and police tasked with analysing child exploitation material (Greinacher et al., 2019; Bourke & Craun, 2014). In these professions, these risks are more widely recognised, for the most part, contributing to support being implemented and facilitating more understanding from managers and the community (Burns et al., 2008; Krause, 2009). In contrast, despite their centrality to the business model of many online platforms (Barrett, 2020) moderators remain an undervalued, often hidden profession (Gillespie, 2018), frequently located in the global south and commonly not afforded the same benefits and provisions as offered to other professionals (Roberts, 2019; Jereza, 2021). However, the increasing amount of literature demonstrating the potential negative consequences of this work (e.g., Benjelloun & Otheman, 2020; Roberts, 2019) illustrates that companies need to take more responsibility for the welfare of these employees and volunteers and provide better protection against the possible deleterious psychological effects.

Understanding the effects this work has on CMs from a psychological perspective allows the symptoms to be positioned in a framework that can draw on the large psychological literature around trauma and its sequelae. It can link the experiences and emotional responses direct from the voices of CMs to clinical practice and empirically ground potential recommendations. This includes creating interventions based on “what works” and for whom, as well as better designing systems to protect CMs. Many organisations provide technological solutions such as blurring to minimise the impact (Das et al., 2020). However, these are not necessarily considered useful by moderators (Spence et al., 2023) and a lack of support has been associated with quitting in volunteer moderators (Schöpke-Gonzalez et al., 2022). Nevertheless, there are potentially processes which could be put in place to alleviate emotional distress, for instance playing Tetris after trauma exposure has shown some early promise in reducing the occurrence of intrusive memories (Iyadurai et al., 2018). Regardless, these results suggest that companies should provide trauma-informed care at work, which is guided by the principles of safety, trustworthiness, peer support collaboration and mutuality, empowerment, and cultural, historical and gender issues (Substance Abuse and Mental Health Services Administration, 2014) and is shown to reduce STS and burnout amongst staff (Handran, 2015; Schepers & Young, 2022).

Table 2. Themes Rearranged Into PTSD Symptom Domains.

PTSD Domain	Sub-Themes
Intrusion	Thoughts of CSAM
Avoidance	Incompatible with Family Life
Negative Symptoms	Cynical Views of People, World is Darker Place, Mental Health, Desensitisation, Anger, Fatigue, Self-Care, Effects Relationships, View of Self
Hyperarousal	Hypervigilance Regarding Children, Sleep Disturbance, Anxiety

Although trauma responses such as intrusive thoughts might be disturbing, what may be more insidious is the slow change in perceptions of the world and those in our communities, which could be more pervasive and affect all spheres of the CMs life. Research with volunteer moderators found they experience cognitive changes like increased apathy and guilt (Dosono & Semaan, 2019; Lo, 2018). In this study, commercial CMs described having a darker, more cynical view of the world and other people, and these views may not change once they leave their job (O'Malley et al., 2019). Although the specific ages of the CMs were unknown, they tended to be young and,

generally, most CMs are younger than thirty (Mukhopadhyay, 2020). Although the evidence is inconsistent, younger age has been associated with greater distress after experiencing trauma and more symptoms (Acierno et al., 2006; Adams & Boscarino, 2006), as well as higher levels of burnout (Brady, 2017). This may be because when compared to older workers, they lack a wide array of life experiences to draw on, making them less equipped to deal with stressful situations (Lerias & Byrne, 2003). Older workers are also more likely to use positive emotion regulation strategies and possess higher emotional competencies to stave off negative consequences (Steiger, 2020). Therefore because of their age, CMs may be a group that are particularly vulnerable to repeatedly working with traumatic material.

There is currently limited research that investigates the psychological impacts of the role. However, there is some evidence that suggests the harmful impacts of moderation are not necessarily restricted to those viewing overtly disturbing material such as CSAM. In work by Jereza (2022, p. 12) a moderator commented “working with hate speech so much is it would get into my head. Like *these images and words would flash across my brain like with no warning, and I didn't want them there*. It was—it was awful. I was like, I would kind of torturing myself, um, and so I do think, you know, it can get into your head easier than you would think perhaps? Or when I was out and about like if *I ever saw a white guy with a tattoo, I would immediately start scanning his body for hate symbols*. And every white guy has a tattoo in [City] . . . I was like hating myself for it. They're not supposed to be there. I'm not this person” (emphasis added). This suggests that intrusive thoughts and hypervigilance may also be associated with working with hate speech. This content is the most prevalent type that CMs have to analyse on social media platforms (Facebook, 2022a; Twitter, 2021) therefore, research into the effects of working with this content specifically is needed. In particular, if it is associated with VT, where the world views of moderators may be changed, possibly even in agreement with what they being exposed to. Indeed, it is possible that many of the same impacts might be present across different content types but the substance of the impacts could vary in line with the material the CM is exposed to.

Limitations

The current study was a cross-sectional qualitative design therefore was not able to capture changes over time. Additionally, time in post was not reported, therefore it cannot be ascertained if certain impacts were more likely to be described by those earlier or later in their career as a CM. Similarly, to encourage participation CMs were not asked about details of their role such as if they were employed directly by the company or were outsourced labour, although all participants were in paid positions and were able to conduct an interview in English. There are numerous volunteer CMs used by sites like Reddit (Caplan, 2018) and moderation is a global phenomenon. Thus, this study was unable to fully capture the full range of different work environments CMs will be exposed to, not only in terms of support provided but also working conditions, which are likely impact CM resilience (Burns et al., 2008; Substance Abuse and Mental Health Services Administration, 2014). However, this study opens up future avenues for research, including larger scale studies that can quantify potential impacts and explore potential factors that increase CM resilience.

Conclusions

Many of the CMs described a range of symptoms that were associated with experiencing trauma at work, including intrusive thoughts, avoidance and hypervigilance around children. This suggests it is crucial for companies to provide psychoeducation, intervention and trauma informed care. Of particular concern were the cognitive biases that highlighted a cynical and suspicious framework as these may be harder for CMs to recognise and be resistant to change, even after leaving the role. Larger scale quantitative and longitudinal research are both needed to quantify the possible effects and their long-term consequences.

Conflict of Interest

The authors have no conflict of interest to declare.

Authors' Contribution

Ruth Spence: project administration, formal analysis, writing—original draft. **Antonia Bifulco:** supervision, methodology, writing—review & editing. **Paula Bradbury:** data curation, methodology. **Elena Martellozzo:** funding acquisition, investigation, writing—review & editing. **Jeffrey DeMarco:** conceptualization, funding acquisition, writing—review & editing.

Acknowledgement

This work was conducted with the help of Amy Harrison.

References

- Acierno, R., Ruggiero, K. J., Kilpatrick, D. G., Resnick, H. S., & Galea, S. (2006). Risk and protective factors for psychopathology among older versus younger adults after the 2004 Florida hurricanes. *The American Journal of Geriatric Psychiatry, 14*(12), 1051–1059. <https://doi.org/10.1097/01.JGP.0000221327.97904.b0>
- Adams, R. E., & Boscarino, J. A. (2006). Predictors of PTSD and delayed PTSD after disaster: The impact of exposure and psychosocial resources. *The Journal of Nervous and Mental Disease, 194*(7), 485–493. <https://doi.org/10.1097/01.nmd.0000228503.95503.e9>
- Ahmad, S., & Krzywdzinski, M. (2022). Moderating in obscurity: How Indian content moderators work in global content moderation value chains. In M. Graham & F. Ferrari (Eds.), *Digital work in the planetary market* (pp. 77–95). The MIT Press.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>
- Asaro, P. M. (2013). The labor of surveillance and bureaucratized killing: New subjectivities of military drone operators. *Social Semiotics, 23*(2), 196–224. <https://doi.org/10.1080/10350330.2013.777591>
- Barrett, P. M. (2020). *Who moderates the social media giants? A call to end outsourcing*. New York University Stern Center for Business and Human Rights. https://bhr.stern.nyu.edu/tech-content-moderation-june-2020?_ga=2.220235672.1205101538.1694289365-761940979.1694289364
- Benjelloun, R., & Otheman, Y. (2020). Psychological distress in a social media content moderator: A case report. *Archives of Psychiatric Mental Health, 4*, Article 7375. <https://doi.org/10.29328/journal.apmh.1001024>
- Bourke, M. L., & Craun, S. W. (2014). Coping with secondary traumatic stress: Differences between U.K. and U.S. child exploitation personnel. *Traumatology: An International Journal, 20*(1), 57–64. <https://doi.org/10.1037/h0099381>
- Brady, P. Q. (2017). Crimes against caring: Exploring the risk of secondary traumatic stress, burnout, and compassion satisfaction among child exploitation investigators. *Journal of Police and Criminal Psychology, 32*(4), 305–318. <https://doi.org/10.1007/s11896-016-9223-8>
- Browne, T., Evangeli, M., & Greenberg, N. (2012). Trauma-related guilt and posttraumatic stress among journalists. *Journal of Traumatic Stress, 25*(2), 207–210. <https://doi.org/10.1002/jts.21678>
- Burns, C. M., Morley, J., Bradshaw, R., & Domene, J. (2008). The emotional impact on and coping strategies employed by police teams investigating internet child exploitation. *Traumatology, 14*(2), 20–31. <https://doi.org/10.1177/1534765608319082>
- Caplan, R. (2018). *Content or context moderation? Artisanal, community-reliant, and industrial approaches*. Data & Society Research Institute. <https://datasociety.net/library/content-or-context-moderation/>
- Chen, A. (2014, October 23). *The laborers who keep dick pics and beheadings out of your Facebook feed*. Wired. <https://www.wired.com/2014/10/content-moderation/>
- Chervonsky, E., & Hunt, C. (2017). Suppression and expression of emotion in social and interpersonal outcomes: A meta-analysis. *Emotion, 17*(4), 669–683. <https://doi.org/10.1037/emo0000270>

- Conrad, D., & Kellar-Guenther, Y. (2006). Compassion fatigue, burnout, and compassion satisfaction among Colorado child protection workers. *Child Abuse & Neglect*, 30(10), 1071–1080.
<https://doi.org/10.1016/j.chiabu.2006.03.009>
- Cook, C. L., Cai, J., & Wohn, D. Y. (2022). Awe versus aww: The effectiveness of two kinds of positive emotional stimulation on stress reduction for online content moderators. In *Proceedings of the ACM on human-computer interaction* (pp. 1–19). ACM. <https://doi.org/10.1145/3555168>
- Craig, C. D., & Sprang, G. (2010). Compassion satisfaction, compassion fatigue, and burnout in a national sample of trauma treatment therapists. *Anxiety, Stress, & Coping*, (23)3, 319–339.
<https://doi.org/10.1080/10615800903085818>
- Das, A., Dang, B., & Lease, M. (2020). Fast, accurate, and healthier: Interactive blurring helps moderators reduce exposure to harmful content. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* (pp. 33–42). <https://doi.org/10.1609/hcomp.v8i1.7461>
- Dosono, B., & Semaan, B. (2019). Moderation practices as emotional labor in sustaining online communities: The case of AAPI identity work on Reddit. In CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–13). ACM. <https://doi.org/10.1145/3290605.3300372>
- Drootin, A. (2021). “Community guidelines”: The legal implications of workplace conditions for internet content moderators. *Fordham Law Review*, 90(3), 1197–1244. <https://ir.lawnet.fordham.edu/flr/vol90/iss3/4>
- Dwoskin, E. (2019, May 8). *Inside Facebook, the second-class workers who do the hardest job are waging a quiet battle*. The Washington Post. <https://www.washingtonpost.com/technology/2019/05/08/inside-facebook-second-class-workers-who-do-hardest-job-are-waging-quiet-battle/>
- Dwoskin, E., Whalen, J., & Cabato, R. (2019, July 25). *Content moderators at YouTube, Facebook and Twitter see the worst of the web – and suffer silently*. The Washington Post. <https://www.washingtonpost.com/technology/2019/07/25/social-media-companies-are-outsourcing-their-dirty-work-philippines-generation-workers-is-paying-price/>
- Facebook. (2022a). *Community standards enforcement report*. <https://transparency.fb.com/data/community-standards-enforcement/child-nudity-and-sexual-exploitation/facebook/>
- Facebook. (2022b). *Meta reports first quarter 2022 results*. <https://investor.fb.com/investor-news/press-release-details/2022/Meta-Reports-First-Quarter-2022-Results/default.aspx>
- Fact.MR. (2022). *Content moderation solutions market overview (2022–2032)*. <https://www.factmr.com/report/4522/content-moderation-solutions-market>
- Figley, C. R. (Ed.). (1995). *Compassion fatigue: Coping with secondary traumatic stress disorder in those who treat the traumatized*. Brunner/Mazel.
- Foley, J., & Massey, K. L. D. (2021). The ‘cost’ of caring in policing: From burnout to PTSD in police officers in England and Wales. *The Police Journal: Theory, Practice and Principles*, 94(3), 298–315.
<https://doi.org/10.1177/0032258X20917442>
- Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology*, 13, Article 117. <https://doi.org/10.1186/1471-2288-13-117>
- Gerrard, Y. (2022). Social media moderation: The best-kept secret in Tech. In D. Rosen (Ed.), *The social media debate* (pp. 77–95). Routledge.
- Ghoshal, A. (2017, January 12). *Microsoft sued by employees who developed PTSD after reviewing disturbing content*. TNW. <https://thenextweb.com/news/microsoft-sued-by-employees-who-developed-ptsd-after-reviewing-disturbing-content>
- Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press. <https://doi.org/10.12987/9780300235029>
- Greinacher, A., Derezza-Greeven, C., Herzog, W., & Nikendei, C. (2019). Secondary traumatization in first responders: A systematic review. *European Journal of Psychotraumatology*, 10(1), Article 1562840.
<https://doi.org/10.1080/20008198.2018.1562840>

- Hakanen, J. J., Schaufeli, W. B., & Ahola, K. (2008). The job demands-resources model: A three-year cross-lagged study of burnout, depression, commitment, and work engagement. *Work & Stress*, 22(3), 224–241. <https://doi.org/10.1080/02678370802379432>
- Handran, J. (2015). Trauma-informed systems of care: The role of organizational culture in the development of burnout, secondary traumatic stress and compassion satisfaction. *Journal of Social Welfare and Human Rights*, 3(2), 1–22. <https://doi.org/10.15640/jswahr.v3n2a1>
- Hu, T., Zhang, D., Wang, J., Mistry, R., Ran, G., & Wang, X. (2014). Relation between emotion regulation and mental health: A meta-analysis review. *Psychological Reports*, 114(2), 341–362. <https://doi.org/10.2466/03.20.PR0.114k22w4>
- Hunsaker, S., Chen, H.-C., Maughan, D., & Heaston, S. (2015). Factors that influence the development of compassion fatigue, burnout, and compassion satisfaction in emergency department nurses. *Journal of Nursing Scholarship*, 47(2), 186–194. <https://doi.org/10.1111/jnu.12122>
- Iyadurai, L., Blackwell, S. E., Meiser-Stedman, R., Watson, P. C., Bonsall, M. B., Geddes, J. R., Nobre, A. C., & Holmes, E. A. (2018). Preventing intrusive memories after trauma via a brief intervention involving Tetris computer game play in the emergency department: A proof-of-concept randomized controlled trial. *Molecular Psychiatry*, 23, 674–682. <https://doi.org/10.1038/mp.2017.23>
- Jaffe, P. G., Crooks, C. V., Dunford-Jackson, B. L., & Town, J. M. (2003). Vicarious trauma in judges: The personal challenge of dispensing justice. *Juvenile and Family Court Journal*, 54(4), 1–9. <https://doi.org/10.1111/j.1755-6988.2003.tb00083.x>
- Jereza, R. (2021). Corporeal moderation: Digital labour as affective good. *Social Anthropology*, 29(4), 928–943. <https://doi.org/10.1111/1469-8676.13106>
- Jereza, R. (2022). “I’m not this person”: Racism, content moderators, and protecting and denying voice online. *New Media & Society*, 1–17. <https://doi.org/10.1177/1461444822112222>
- Krause, M. (2009). Identifying and managing stress in child pornography and child exploitation investigators. *Journal of Police & Criminal Psychology*, 24(1), 22–29. <https://doi.org/10.1007/s11896-008-9033-8>
- Ledingham, M. D., Standen, P., Skinner, C., & Busch, R. (2019). “I should have known”. The perceptual barriers faced by mental health practitioners in recognising and responding to their own burnout symptoms. *Asia Pacific Journal of Counselling and Psychotherapy*, 10(2), 125–145. <https://doi.org/10.1080/21507686.2019.16346>
- Lee, J. J., Gottfried, R., & Bride, B. E. (2018). Exposure to client trauma, secondary traumatic stress, and the health of clinical social workers: A mediation analysis. *Clinical Social Work Journal*, 46(3), 228–235. <https://doi.org/10.1007/s10615-017-0638-1>
- Léonard, M.-J., Saumier, D., & Brunet, A. (2020). When the lawyer becomes traumatized: A scoping review. *SAGE Open*, 10(3), 1–10. <https://doi.org/10.1177/2158244020957032>
- Lerias, D., & Byrne, M. K. (2003). Vicarious traumatization: Symptoms and predictors. *Stress and Health*, 19(3), 129–138. <https://doi.org/10.1002/smi.969>
- Litz, B. T., Stein, N., Delaney, E., Lebowitz, L., Nash, W. P., Silva, C., & Maguen, S. (2009). Moral injury and moral repair in war veterans: A preliminary model and intervention strategy. *Clinical Psychology Review*, 29(8), 695–706. <https://doi.org/10.1016/j.cpr.2009.07.003>
- Lo, C. (2018). *When all you have is a Banhammer: The social and communicative work of volunteer moderators* [Doctoral thesis, Massachusetts Institute of Technology, Cambridge, MA]. <https://cmsw.mit.edu/wp/wp-content/uploads/2018/05/Claudia-Lo-When-All-You-Have-Is-a-Banhammer.pdf>
- Matias, J. N. (2019). The civic labor of volunteer moderators online. *Social Media + Society*, 5(2), 1–12. <https://doi.org/10.1177/2056305119836778>
- Messenger, H., & Simmons, K. (2021, May 10). *Facebook content moderators say they receive little support, despite company promises*. NBC News. <https://www.nbcnews.com/business/business-news/facebook-content-moderators-say-they-receive-little-support-despite-company-n1266891>
- Mukhopadhyay, B. R. (2020, August 26). *WARNING: The (open) secret lives of content moderators*. Cassandra Voices. <https://cassandravoices.com/society-culture/society/warning-the-open-secret-lives-of-content-moderators>

- Newton, C. (2019, December 16). *The terror queue: These moderators help keep Google and YouTube free of violent extremism – and now some of them have PTSD*. The Verge. <https://www.theverge.com/2019/12/16/21021005/google-youtube-moderators-ptsd-accenture-violent-disturbing-content-interviews-video>
- Newton, C. (2020a, May 12). *Facebook will pay \$52 million in settlement with moderators who developed PTSD on the job*. The Verge. <https://www.theverge.com/2020/5/12/21255870/facebook-content-moderator-settlement-scola-ptsd-mental-health>
- Newton, C. (2020b, January 24). *YouTube moderators are being forced to sign a statement acknowledging the job can give them PTSD*. The Verge. <https://www.theverge.com/2020/1/24/21075830/youtube-moderators-ptsd-accenture-statement-lawsuits-mental-health>
- O'Malley, M., Robinson, Y. A., Hydon, S., Caringi, J., & Hu, M. (2019). *Organizational resilience: Reducing the impact of secondary trauma on front line human services staff*. SAMHSA ReCAST Issue Brief. https://eastsideforall.org/wp-content/uploads/2020/01/ReCAST-Issue-Brief_Secondary-Trauma.pdf
- Parks, L. (2019). Dirty data: Content moderation, regulatory outsourcing, and the cleaners. *Film Quarterly*, 73(1), 11–18. <https://doi.org/10.1525/fq.2019.73.1.11>
- Powell, M., Cassematis, P., Benson, M., Smallbone, S., & Wortley, R. (2015). Police officers' perceptions of their reactions to viewing internet child exploitation material. *Journal of Police and Criminal Psychology*, 30(4), 103–111. <https://doi.org/10.1007/s11896-014-9148-z>
- Ralph, N. (2020, August 21). *Dealing with the personal impact of crimes against children*. College of Policing. <https://www.college.police.uk/article/dealing-personal-impact-crimes-against-children>
- Riedl, M. J., Masullo, G. M., & Whipple, K. N. (2020). The downsides of digital labor: Exploring the toll incivility takes on online comment moderators. *Computers in Human Behavior*, 107, Article 106262. <https://doi.org/10.1016/j.chb.2020.106262>
- Ritchie, J., & Lewis, J. (2003). *Qualitative research practice*. Sage Publications.
- Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press. <https://doi.org/10.2307/j.ctvhrcz0v>
- Schepers, O. C., Young, K. S. (2022). Mitigating secondary traumatic stress in preservice educators: A pilot study on the role of trauma-informed practice seminars. *Psychology in the Schools*, 59(2), 316–333. <https://doi.org/10.1002/pits.22610>
- Schiffer, Z. (2021, July 22). *Facebook content moderators call for company to put an end to overly restrictive NDAs*. The Verge. <https://www.theverge.com/2021/7/22/22587757/facebook-content-moderators-ireland-end-restrictive-ndas>
- Schindel-Allon, I., Aderka, I. M., Shahar, G., Stein, M., & Gilboa-Schechtman, E. (2010). Longitudinal associations between post-traumatic distress and depressive symptoms following a traumatic event: A test of three models. *Psychological Medicine*, 40(10), 1669–1678. <https://doi.org/10.1017/S0033291709992248>
- Schöpke-Gonzalez, A. M., Atreja, S., Shin, H. N., Ahmed, N., & Hemphill, L. (2022). Why do volunteer content moderators quit? Burnout, conflict and harmful behaviors. *New Media & Society*. <https://doi.org/10.1177/14614448221138529>
- Spence, R., Harrison, A., Bradbury, P., Bleakley, P., Martellozzo, E., & DeMarco, J. (2023). Content moderators' strategies for coping with the stress of moderating content online. *Journal of Online Trust & Safety*, 1(5), 1–18. <https://doi.org/10.54501/jots.v1i5.91>
- Statista (2022, July 27). *Number of monthly active Instagram users from January 2013 to December 2021*. <https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/>
- Steiger, M. (2020). *Building a resilient workforce: Programming for commercial content moderation staff* [Doctoral dissertation, St. Mary's University]. <https://commons.stmarytx.edu/dissertations/46>
- Steiger, M., Bharucha, T. J., Venkatagiri, S., Riedl, M. J., & Lease, M. (2021). The psychological well-being of content moderators: The emotional labor of commercial moderation and avenues for improving support. In *CHI '21*:

Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1–14). ACM. <https://doi.org/10.1145/3411764.3445092>

Substance Abuse and Mental Health Services Administration (2014). *SAMHSA's Concept of Trauma and Guidance for a Trauma-Informed Approach* (HHS Publication No. (SMA) 14-4884). Abuse and Mental Health Services Administration. https://ncsacw.acf.hhs.gov/userfiles/files/SAMHSA_Trauma.pdf

Tomyn, A. J., Powell, M. B., Cassematis, P., Smallbone, S., & Wortley, R. (2015). Examination of the subjective well-being of Australian internet child exploitation investigators. *Australian Psychologist*, 50(3), 203–211. <https://doi.org/10.1111/ap.12119>

Twitter. (2021). *Rules enforcement report*. <https://transparency.twitter.com/en/reports/rules-enforcement.html#2021-jul-dec>

Vasconcelos, A. G., Lima, E. P., Teoh, K., do Nascimento, E., MacLennan, S., & Cox, T. (2021). Work-related factors in the etiology of symptoms of post-traumatic stress among first responders: the Brazilian Firefighters Longitudinal Health Study (FLoHS). *Cadernos de Saúde Pública*, 37(9), Article e00135920. <https://doi.org/10.1590/0102-311X00135920>

Wessely, S., Rose, S., & Bisson, J. (2000). Brief psychological interventions (“debriefing”) for trauma-related symptoms and the prevention of post traumatic stress disorder. *The Cochrane Database of Systematic Reviews*, 2, Article CD000560. <https://doi.org/10.1002/14651858.CD000560>

Witynski, M. (2022, 31 May). *Unpaid social media moderators perform labor worth at least \$3.4 million a year on Reddit alone*. Northwestern Now. <https://news.northwestern.edu/stories/2022/05/unpaid-social-media-moderators/>

Wohn, D. Y. (2019). Volunteer moderators in Twitch micro communities: How they get involved, the roles they play, and the emotional labor they experience. In *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). ACM. <https://doi.org/10.1145/3290605.3300390>

About Authors

Ruth Spence is a senior researcher at the Centre for Abuse and Trauma Studies at Middlesex University working across various projects that focus on trauma experiences and their psychological sequela, as well as developing online assessment and intervention approaches. She has worked in partnership with charities, industry and third sector organisations using a range of online methodologies and analytic techniques to create insights from data. She has published extensively on the determinants of mental health, including life events and the psychological effects of online content.

<https://orcid.org/0000-0002-6197-9975>

Professor **Antonia Bifulco** is a Lifespan psychologist in the Department of Psychology at Middlesex University, and director of the Centre for Abuse and Trauma Studies. Her research over 30 years has investigated the circumstances around a range of childhood abuses, including sexual abuse, together with the consequences for later life adjustment, relationships, adult trauma and psychological disorder. This has been published in numerous academic journal articles, and several books.

<https://orcid.org/0000-0001-8316-9706>

Paula Bradbury is doctoral researcher at Middlesex University exploring the appropriateness of policy, policing practice and legislation in response to adolescents sharing nudes. She is also an active senior researcher at the Centre of Abuse and Trauma Studies. She is a published, peer reviewed author, specialising in the field of online harms, child sexual abuse and exploitation, stalking, sexual violence, offender pathways and adolescent digital behaviours. She has a keen interest in understanding sexual behaviours in cyberspace and how our safeguarding policies and policing practices need to evolve to meet the challenges of the Metaverse. She provides private online harms consultation to public, private, and civil service organisations, both in the UK and abroad.

<https://orcid.org/0000-0002-7837-7233>

Dr **Elena Martellozzo** is an Associate Professor in Criminology and Associate Director at the Centre for Child Abuse and Trauma Studies (CATS) at Middlesex University. Elena is a world leading expert in cybercrime, focusing primarily on online harms, online safety and is involved in policy debates at the intersection of technology and human behaviour. One of Elena's key focuses is to develop innovative research and sharing her findings in international fora. She continually attracts research funding in highly sensitive areas with government departments including the police, the Internet Watch Foundation, the NSPCC, the Children Commissioner for England, the Home Office.

<https://orcid.org/0000-0002-1249-7611>

Jeffrey DeMarco is a forensic psychologist and Senior Lecturer at Middlesex University. The majority of his operational, policy and insight work explores the intersection between psychology and social institutions such as the Criminal Justice System. This has included work for the European Commission, enhancing the policing of sexual abuse; investigating youth justice systems for UNICEF across the MENA region and eastern Africa, and establishing educational programmes for parents and young people; improving partnership between local communities and military in conflict zones, including Iraq and Afghanistan while developing well-being "hubs" for families to seek health, education, immigration and criminal justice advice; and assessing the psychopathology of adolescent victims/offenders of violence presenting to the police and statutory services.

<https://orcid.org/0000-0002-7160-2100>

✉ Correspondence to

Ruth Spence, Centre for Abuse and Trauma Studies, Middlesex University, London, NW4 4BT.
r.spence@mdx.ac.uk

© Author(s). The articles in *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* are open access articles licensed under the terms of the [Creative Commons BY-SA 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/) which permits unrestricted use, distribution and reproduction in any medium, provided the work is properly cited and that any derivatives are shared under the same license.