THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Troubling content

Citation for published version:
Brownlie, J, Ho, JC, Dunne, N, Fernández, N & Squirrell, T 2021, 'Troubling content: Guiding discussion of death by suicide on social media', *Sociology of Health & Illness*. https://doi.org/10.1111/1467-9566.13245

Digital Object Identifier (DOI):
10.1111/1467-9566.13245

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Sociology of Health & Illness

OPEN ACCESS

ORIGINAL ARTICLE

Troubling content: Guiding discussion of death by suicide on social media

Julie Brownlie[1] ⓘ | Justin Chun-ting Ho[2] | Nikki Dunne[3] |
Nichole Fernández[3] | Tim Squirrell[3]

[1]College of Humanities and Social Science, University of Edinburgh, Edinburgh, UK

[2]Centre for European Studies and Comparative Politics, Paris, France

[3]Previously University of Edinburgh, Edinburgh, UK

Correspondence
Julie Brownlie, School of Social and Political Science, Edinburgh University, Rm 1.4, 22 George Square, Edinburgh, UK.
Email: julie.brownlie@ed.ac.uk

Abstract

Growing concerns about "online harm" and "duty of care" fuel debate about how best to regulate and moderate "troubling content" on social media. This has become a pressing issue in relation to the potential application of media guidelines to online discussion of death by suicide—discussion which is troubling because it is often transgressive and contested. Drawing on an innovative mixed-method analysis of a large-scale Twitter dataset, this article explores in depth, for the first time, the complexities of applying existing media guidelines on reporting death by suicide to online contexts. By focusing on five highly publicised deaths, it illustrates the limits of this translation but also the significance of empathy (its presence and absence) in online accounts of these deaths. The multi-relational and politicised nature of empathy, and the polarised nature of Twitter debate, suggests that we need to step back from calls for the automatic application of guidelines produced in a pre-digital time to understand more about the sociocultural context of how suicide is discussed on social media. This stepping back matters because social media is now a key part of how lives and deaths are deemed grievable and deserving of our attention.

KEYWORDS
bereavement, empathy, guidelines, social media, suicide

INTRODUCTION

> Violation of guidelines on reporting of suicide all over this app. Across all social media these days. Its vile and intrusive
>
>   (Suzanne Moore, Twitter, 6 June 2018, https://twitter.com/suzanne_moore)

In the UK, talk about "online harm" and "duty of care" is everywhere. In 2019, the government published a White Paper (Javid & Wright, 2019), announcing its intention to establish a statutory duty of care by social media companies, to be overseen by an independent regulator. The disputed nature of harmful online content, however, meant that the White Paper was inevitably dealing with "grey areas", not least because harm often results not "from deliberate criminality, but from ordinary people released from the constraints of civility" (Guardian, 2019). As such, the White Paper is part of a wider debate on digital citizenship and ethical behaviour (McCosker, 2014) including whether "aberrant" online participation is best understood as trolling, provocation or as some other form of public engagement (Phillips, 2011; Jane, 2017).

This article is concerned with one specific grey area: discussion online of death by suicide. Recognition in death, as in life, is uneven (Butler, 2004). Before social media, deaths by suicide tended to go unreported in mainstream media unless they involved high profile people (Jamieson et al., 2003). In those cases, "hierarchies of information release" (Gibson, 2015) shaped who got to tell, and to know, about a death but these hierarchies have now been usurped by a digitalised "culture of participatory mourning", where even the dead can report on their own demise. "Celebrity deaths", including those by suicide, have become a prominent feature of this mediatised landscape of mourning (Burgess et al., 2018; Döveling et al., 2018).

These social changes have significant potential health implications because of the "Werther Effect" (Phillips, 1974)—where the reporting of suicide of public figures leads to an increase in deaths by suicide in the general population. While the representation of suicide in mainstream journalism has been extensively studied (Gould et al., 2014), investigation of how others' deaths by suicide are discussed on social media, including Twitter, is still in its infancy, particularly compared to discussion of suicidal *intent* in digital spaces (Horne & Wiggins, 2009; Robinson et al., 2016). This work is urgently needed because of the relatively young user base of Twitter (Pew Research Center, 2019) and a growing concern about the suicide rate of young people globally (World Health Organization, 2019).

Unlike the White Paper (Javid & Wright, 2019) which primarily addresses issues of regulation and moderation, our concern is with bottom-up approaches to change, that is, with if, and how, users might be guided to discuss suicide more sensitively. The distinction between regulation, moderation and guidance, however, is not clear-cut: moderation is a means of responding to breaches of community norms, and guidelines themselves, while advisory, can feel regulatory in practice. Nevertheless, we focus primarily on the role of guidelines in proactively shaping online speech, as well as being the basis for moderation and intervention *after* individuals have breached community norms or regulations. We aim to explore the challenges of operationalising guidelines through an empirical investigation of Twitter data, presenting evidence about the extent to which existing guidelines on reporting of deaths by suicide are being breached "all over this app" and considering how they might be extended to social media.

We focus on everyday Twitter "talk", primarily by individual users rather than organisations or journalists, about the death by suicide of five well-known people. In particular, our aim is to explore cases of death by suicide that have been publicly discussed, though we remain sensitive to the risks of ventriloquising unempathetic accounts. We begin by bringing social science and health research into

closer dialogue with media and communication literature on the reporting of suicide and by engaging with research on how to conceptualise the sharing of emotion on social media.

TOWARDS A MEDIATISED UNDERSTANDING OF DISCUSSION OF SUICIDE ONLINE

Media narratives about suicide are part of "cultural sense-making" but at the same time the media itself is culturally shaped (Mueller, 2017). A *mediatised*, rather than simply mediated, understanding of Twitter accounts of suicide recognises this mutual shaping of media and social life (Döveling et al., 2018). We explore this interplay below through research on reporting of suicide and on the sharing of emotion online.

As noted, media reporting of death by suicide has become a key public health concern because of its perceived relationship to subsequent suicidal behaviour (Marchant et al., 2020). In response, both reporting guidelines (endnotes 5 & 6), and tools for assessing media content, have emerged in recent years (John et al., 2014). The difficulty, however, of isolating the specific impact of reporting (Mueller, 2017), and establishing causal mechanisms, means that "contagion" is not easy to evidence. There is, moreover, no consensus about what the focus for regulation should be, and a concern that too much regulation might increase the stigma, and hence the risk, of suicide.

These complexities are writ large on social media. The scale, immediacy and networked nature of digital spaces such as Twitter lead to discursive intensification and amplification (Gillespie, 2018). This makes moderation, or regulation, of the accuracy and/or sensitivity of digital content extremely difficult. Twitter posts about suicide have been studied (O'Dea et al., 2017), but there is still relatively little work on the guidance, regulation or moderation of discussions of suicide on this platform. The relationship between social media and the Werther effect is, however, now being recognised with calls for further research on how the tenor and emotional content of social media discourse impacts "at-risk" groups and individuals (Fahey et al., 2018). The recent UK White Paper has reignited debate about social media companies as "publishers" and hence responsible, and liable, for their content. Others, however, use the analogy of social media as "public space": where the onus is on those who own these spaces to ensure *prevention of harm* (Perrin & Woods, 2018).

The UK White Paper suggests a step change towards statutory regulation, but it relies on two existing techniques for finding and/or classifying "troubling" content: reporting and algorithms (Boyd et al., 2011).[1] "Flagging" allows users to report material within the rubric of a platform's "community guidelines". A flag's meaning, however, is not straightforward; as interactions between users, flags, content moderators and platforms are mediated and modulated by the algorithmic procedures underpinning the "flagging" interface (Crawford & Gillespie, 2014). Algorithmic regulation is also subject to subversion, with users finding multiple ways of accessing and disseminating "inappropriate" content, in spite of measures such as keyword blocks (Gerrard, 2018). Beyond such technical challenges, moderation of troubling content also depends "on contested theories of psychological impact and competing politics of culture" (Gillespie, 2018, p. 10). These are visible in approaches to the rise of pro-self-harm sites, with some seeing such sites as encouraging harm (Boero & Pascoe, 2012) and others emphasising their role in reducing stigma (Boyd et al., 2011).[2]

Thinking in terms of mediatisation, however, raises questions about how tweets, any more than offline speech, can be meaningfully subject to notions of "best practice". In grappling with these questions, we need to conceptualise how and why emotions are shared online. We know from existing research that emotion is shared in different ways across and also *within* digital platforms (Zappavigna, 2012). For some, this variation reflects differences in social networks: Facebook, for instance, involves

known networks and their interconnectedness can lead to users choosing to share some of their views on less socially dense platforms. Even within Twitter, however, those who are part of a higher density social network are less likely to share some emotions. This reflects the performative nature of online environments—users' *beliefs* about the platform and the social networks they imagine they are, or wish to be, part of, are key to sharing (Litt & Hargittai, 2016) but so too, as much as architecture and algorithms, are the *practices* actually employed in situated contexts (Costa, 2018). While this study is not designed to address users' perceptions, in previous work we found that Twitter users who felt suicidal believed that the social diffuseness of the platform allowed them to express the "unsayable" (Brownlie, 2018). As we will go on to see, however, those who are suicidal or who die by suicide are also judged on Twitter.

The scale, immediacy and networked nature of social media is increasingly being made sense of through the lens of collective or shared emotion with strong emotion seen as triggering and sustaining discussion on Twitter (Thelwall & Kappas, 2014). While the "contagious" power of emotions—particularly in the context of gatherings—has been a feature of sociological analysis from Durkheim (1915) onwards, in online contexts this power is explored increasingly through affect a wide ranging concept that is often used to emphasise the less than conscious and embodied aspects of emotions (Sampson et al., 2018). Ahmed's (2014) concern, however, that the concept of contagion assumes emotions pass smoothly between people in uncomplicated or uncontested ways and indeed that what we pass on is shared in the sense of *being the same thing* is pertinent. Her focus instead is on the circulation of objects, defined as anything emotion is directed towards, and the influence of values on how we direct our emotions (2014:208, 214). In what follows, our analysis is based neither on contagion nor affect but we find it useful to focus on how value (and emotions) come to be attached to those who choose to die by suicide and the challenges of managing these attachments in the context of Twitter.

RESEARCHING DISCUSSION OF SUICIDE ON TWITTER

Drawing on a mixed-methods approach to researching Twitter (Brownlie & Shaw, 2018), and moving iteratively between qualitative and quantitative data, our analysis aims to add to studies which use machine learning to explore reactions to deaths by suicide on social media (Fahey et al., 2018; Niederkrotenthaler et al., 2019). We draw on a large-scale Twitter dataset comprised of tweets about five highly publicised deaths and produced as part of an earlier research project on online trust and empathy (Karamshuk et al., 2017). While not representative of all highly publicised deaths by suicide, these cases were chosen because, at the time of the study, they were relatively recent and data could be easily retrieved. They were also selected for their diversity—including people of different ages and genders and those who were well known before their death as well as those made famous by their death. The cases also spoke to a range of social issues including cyber bullying, digital activism and mental health. Twitter posts were retrieved for 20 days following each of the five deaths. A full dataset of tweets was retrieved for three of these cases—Amanda Todd, Leelah Alcorn and Charlotte Dawson—and sampled datasets for Robin Williams and Aaron Swartz. 1.88 million tweets were collected in total (see Table 1).[3] We provide fuller details about these cases elsewhere (Karamshuk et al., 2017) but offer here a brief overview.

Aaron Swartz, at the time of his death in 2013, was being indicted for data theft after downloading journal articles from the online database at Massachusetts Institute of Technology (MIT). Swartz's death was framed by many through the lens of digital activism with both his legal prosecutors and MIT criticised on Twitter. Canadian teenager, Amanda Todd, died by suicide in 2012. Her death was publicised following the widespread sharing of an online video in which she detailed her experience

TABLE 1   Number of tweets

Case	Count
Aaron	91,564
Amanda	519,244
Charlotte	40,563
Leelah	389,111
Robin	836,009

of cyberbullying. Charlotte Dawson, a New Zealand-Australian television personality and campaigner against cyberbullying, died by suicide in 2014. Leelah Alcorn died by suicide in the United States also in 2014; her suicide note, posted on Tumblr, attracted global attention and her parents were widely criticised on social media for allegedly not accepting her transgender identity. Robin Williams, an internationally known actor, who at the time of his death had reportedly been diagnosed with depression and Parkinson's disease, also died by suicide in 2014.

In the qualitative stage of this project, we worked with a random subsample of tweets (6680) taken from the larger Twitter sample and created as part of the earlier study noted above (see Karamshuk et al., 2017). For the purposes of this project, we qualitatively coded the 6680 tweets across the five cases. A subset of these tweets had already been coded by the original research team to explore empathy, and we draw on this original coding[4] to inform discussion in the latter half of this article on the role of empathy in Twitter talk about suicide.

Some of the tweets analysed were recognised through likes and retweets and as such could be identified as conversations. Our focus, however, similar to Burgess et al. (2018, p. 233), was to treat a tweet "as a kind of a calling out, or a calling forth, that does not necessarily request or require a direct response". Platforms shape discussion about suicide in specific ways; some are highly fluid, others are stabilised through "traces" (Papacharissi, 2015, p. 4). Both these temporal aspects feature in the Twitter dataset: there were day-by-day fluctuations across the five cases depending on what was happening off and online, but patterns could also be identified across the 20 day retrieval period.

We began by creating a coding frame through a close reading of (predominantly offline) media guidelines about death by suicide. Using Google and a clean research browser (Rogers, 2013, p. 111), the term "suicide reporting guidelines" generated a list of 12 of the most highly ranked sets of media guidelines on the reporting of suicide.[5] Five of these (Samaritans, Reporting/Blogging on Suicide, One in Four and New Zealand Mental Health Foundation) contained guidelines specific to digital media, or other references to digital material. The remaining seven were not medium-specific. Additional online-specific guidelines were also searched, using combinations of the words "suicide, social media, guidelines". Reflecting the current research interest in identifying suicidal intent online, the majority of results for this latter search related to what social media users ought to do if they saw content that indicated another user might be suicidal. These results, however, were not incorporated into the coding framework.[6]

The coding scheme produced through a close reading of the above guidelines identified eight practices to be deterred and seven to be encouraged. Those to be deterred included the use of details about the method of suicide; location or individuals involved; over-simplification of the causes; stereotyping; insensitive or incorrect language such as references to people "committing" suicide; photographs or videos of the deceased or links to their social media accounts; reporting the contents of a suicide note; use of fake news articles, or information that contain factually incorrect information; and, finally, insensitivity to the bereaved. Practices that were encouraged included the use of trigger warnings;

reflection on the tragic loss of life; acknowledgement of the complex and multifaceted nature of suicide causality; provision of details of places where support can be found; commentary on the impact of suicide on those left behind; information on the warning signs of suicide; and the use of correct phrasing to refer to suicide.

To help with reliability, the coding scheme was then compared with the work of leading suicide prevention researchers. Using NVIVO 12, the scheme was then applied to the dataset of 6680 tweets. As noted, this dataset had also previously been coded for "lack of empathy" [7] so we also analysed these (N 400) for their characteristics.

Initially content analysis, "directed" by specific guidelines, was carried out, exploring, for example, use of particular language across the dataset (Hsieh & Shannon, 2005). This was followed by a more reflexive approach towards developing themes from the dataset (Brownlie, 2011), including drawing on researchers' experiences of attempting to apply the guidelines. These included, the uncertainty of coding a tweet as simultaneously meeting one guideline while contradicting another, or of working out whether apparent over-simplification was intended by a user or was simply a result of the brevity imposed by the medium. Analysing text "for" emotion is clearly complicated by the fact that emotion does not simply reside in any one document but is constructed, and reconstructed, through reading (Brownlie & Shaw, 2018). Analysing for "lack of empathy", therefore, involved the researcher sharing with the research team their decision-making about why a tweet had been coded as unempathetic on the basis of content—such as a direct personal attack—but also its tone. The latter is notoriously difficult to identify. Even where lack of empathy is detected, for instance, it might not be clear who it is directed at: calls to pay the deceased less attention, for instance, while apparently lacking in empathy could be read as critical of media that helps commodify tragedy.

After the Twitter subsample had been analysed qualitatively, we turned to the full Twitter corpus of 1.88 million tweets (see Table 1).

It is worth noting that there are established methods for detecting sentiment expressed in text, though these would not have been well suited to the exploratory aims of this article. Moreover, such approaches have their own limitations as they often rely on predefined lexical-based dictionaries and it is difficult to create a universal dictionary that works for all contexts (Gonçalves et al., 2013).

In order to examine the prevalence of guideline breaches, we adopted a two-step procedure: first, using keyword analysis, word lists were generated from the coded subsample for each guideline-breaching practice, and then used to count occurrence of guideline-breaching practice within the full Twitter corpus.

Keyword analysis is a technique to identify keywords, "words which occur unusually frequently in comparison with some kind of reference corpus" (Scott, 2016, para. 2), with statistical tools, such as chi-squared tests. These keywords can reflect themes or characteristics of the documents of interest. Our analysis involved several stages. For each of the eight negative practices identified, we split the coded subsample into two sets: the tweets coded as guideline-breaching, and those that were not. The former was used as the target corpus while the latter was used as the reference corpus. Chi-squared tests were used to compare the relative frequency of each word occurrence in the two sets. Second, all words occurring with statistically significant ($p < 0.05$) high frequency in the guideline-breaching set were extracted for further analysis. Third, the contexts of each extracted word were examined using Key Word in Context (KWIC) to determine whether the word was a meaningful indicator of guideline-breaching practice. We were able to apply seven of the eight guideline components relating to deterrence to the full Twitter corpus: insensitivity to the bereaved, details about the suicide, use of fake information, reference to a suicide note, over-simplification, use of stereotypes and incorrect phrasing. We were not able to operationalise guidelines around the use of photographs as they were too technically complex. Fourth, the occurrence of keywords within the full corpus was estimated and

the prevalence of keywords per day analysed so as to understand the temporal dimension of guideline-breaching practices.

The final stage of analysis involved moving between qualitative and quantitative data, asking questions about emerging patterns and whether or not these were consistent across the two datasets. Overall, the methodology was qualitatively driven and iterative and, as we have suggested elsewhere, aimed to "bridge the gap between close readings and large scale patterning" (Karamshuk et al., 2017).

Ethically, we recognise that tweets, while publicly available, may be understood by their authors as remaining within Twitter and/or imagined as relatively anonymous given the diffuseness of the Twittersphere (Franzke et al., 2020). In order to avoid transferring potentially searchable data, from one context to the other, we agree with Markham (2012) that there is a need to anonymise creatively online data. Building on our previous published work in this area (Brownlie & Shaw, 2018), we have accordingly paraphrased rather than reproduced individual tweets.

In addition to the well-documented challenges of researching Twitter (Tromble et al., 2017), our analysis was restricted by the specificity of the cases, the sampling of the guidelines and the extent to which we were able to operationalise themes from the guidelines quantitatively—with only seven practices applied to the full Twitter dataset, there may be other themes that occurred more often but which were difficult to operationalise in this kind of analysis. Even allowing for these limitations and the impossibility of knowing exactly how widespread breaching is, its extent is worth investigating; not least because research suggests that the volume of reporting on Twitter is more strongly related to changes in suicide rates following a celebrity's death than the volume of media reporting in newspapers or on Television (Ueda et al., 2017). Moreover, the dataset allows us to ask useful exploratory questions about *how* this relationship differed across cases and across *time*, and these questions guide the following analysis.

THE EXTENT OF BREACHING OF REPORTING GUIDELINES ONLINE

For illustrative purposes, we focus on those guidelines which the qualitative analysis suggests raise the most and least challenge in the context of social media. Even in relation to only seven guidelines, it is clear that practices in breach of these are commonplace.[8] Figure 1 below shows the percentage of tweets that contain one or more keywords linked to such practices. Insensitivity to the bereaved was the most prevalent, especially in relation to Leelah Alcorn and Amanda Todd: more than 60 per cent of the tweets relating to the former and almost 50 per cent relating to the latter contain keywords for this practice. Over-simplification and stereotyping are also pervasive; approximately 40 per cent of tweets relating to Amanda Todd and 25 per cent of tweets relating to Charlotte Dawson showed evidence of both as did around 5 to 15 per cent of the tweets in the other 3 cases. Details about the suicide, use of fake information, reference to a suicide note and incorrect phrasing were comparatively less prevalent but this could be a methodological limitation, as some practices are harder to identify through relevant keywords.

Our analysis gives some indication of the patterning of different types of breaches by case, but looking at the patterning over time offers further insights. Figure 2 shows that the trend in relation to insensitivity towards the bereaved was the most complex, with high prevalence and some fluctuations over the period, but with no matching trend across the five cases. Over-simplification, on the other hand, for the most part declined over time or, put differently, the complexity of all the cases is increasingly recognised as time passes. A similar trend was observed for stereotyping: higher prevalence of stereotyping was observed just after the suicide but decreased over time. In the immediate
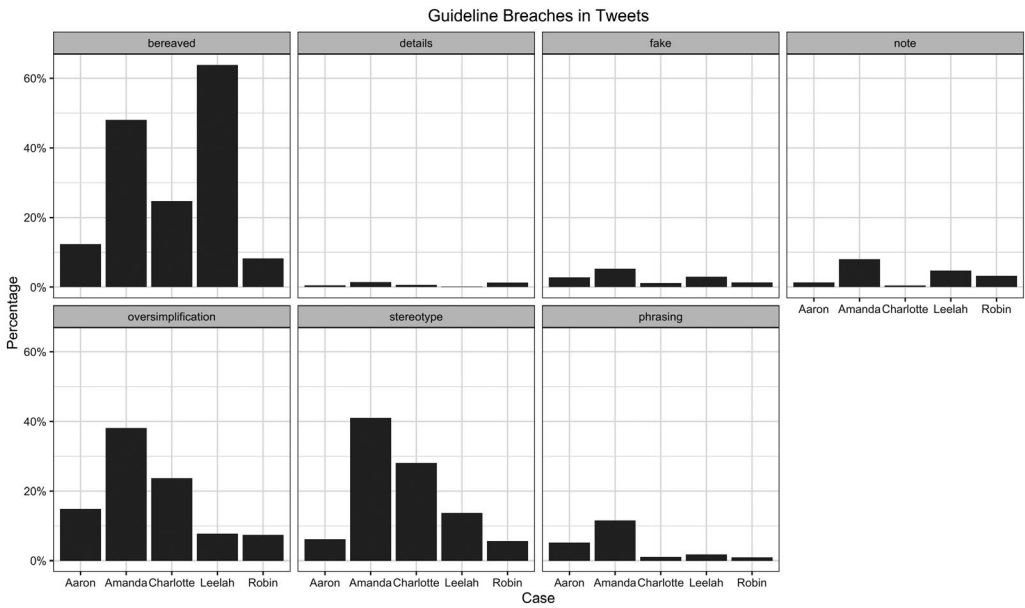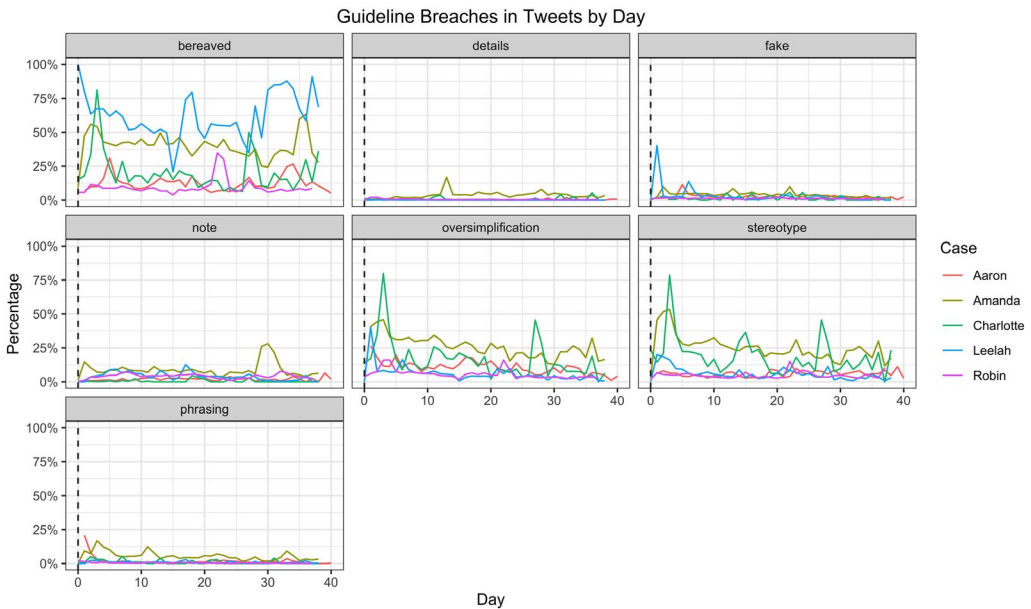
FIGURE 1 Guideline breaches in tweets

FIGURE 2 Guideline breaches in tweet by day

aftermath of a death by suicide, such deaths tend to be portrayed as monocausal but this practice decreases as more information which contextualises the death is released. Immediately aftermath of Leelah Alcorn's death, and the publication of a suicide note, for instance, tweets about her parents were particularly vitriolic. Over time tweets conveying the complexity of her family circumstances increased though new information (for instance that Leelah would be buried as Joshua) and caused

renewed waves of anger towards Leelah's parents. In the case of Amanda Todd, however, this story of increasing awareness of complexity is itself complicated. "Jokes" that circulated immediately after Todd's death, and which were initially drowned out by the volume of "stop bullying"-type tweets, became more prevalent over time. One reading of this is that in circumstances where empathy is lacking from the outset—and we will see below this was the case for Amanda Todd—those who initially refrained from making such comments on the basis that it was "too soon", may come to feel less inhibited over time.

This quantitative analysis points to the extent of breaching across the dataset, and across time, but also offers a preliminary sense of what kind of breaching emerged in relation to different cases.

THE NATURE OF BREACHING

Methods and language use

Qualitative analysis allows us to make greater sense of these patterns. One practice almost universal to the guidelines we studied, and comparatively well avoided in the dataset, is the mention of method. The small number of breaches of this guideline across the dataset mostly occurred in relation to Amanda Todd and took the form of "jokes" about a previous suicide attempt. This demonstrates the challenge of applying guidelines in a top-down fashion to specific acts/words, particularly in a context where, unlike traditional journalism, content is often deliberately transgressive. Keyword searches, however, were effective in identifying language that guidance suggests is insensitive or incorrect including (i) suicide as akin to a crime or a sin, for example "committed" suicide; (ii) suicide as success and failure, for example "successful" suicide or "failed attempt" and (iii) language which sensationalises the rate of suicide, for example "epidemic", "skyrocketing" or (iv) mystifies such deaths, for example "inexplicable", "without warning". These descriptions are prevalent across the dataset, and their detection is fairly straightforward, even if what to *do* about such content is far less so. However, the fact that such language appeared in otherwise sensitive and nuanced accounts about the individuals who had died, however, including in tweets and articles that tried to "call out" insensitive or irresponsible coverage, suggests that lack of knowledge rather than of empathy per se may help explain such breaches. This, in turn, points to the role guidelines can potentially play in proactively shaping online speech rather than reactively identifying breaches.

Over-simplification

"Over-simplification" or "failing to acknowledge the complexities of suicide", a key practice mentioned in the guidelines, was, as we have seen, also prevalent across the dataset. When coding for this we looked to posts in which a single cause was posited as the main or only reason for suicide, such as suggestions that Robin Williams' death was as a result of finding out that he had Parkinson's Disease, or that "stop bullying" headlines led to Amanda Todd's. This guideline is difficult to apply to tweets, however, because their brevity can obscure (even if unintentionally) the multifaceted nature of the death described. The nuance necessary to avoid over-simplification is also difficult to achieve on gamified platforms like Twitter where the simplest and most emotionally resonant messages are often those that are rewarded with likes and retweets (Brady et al., 2017). The nature of the platform—specifically retweeting—also makes other media guidelines, such as avoiding repetition of stories, redundant.

Stereotypes

Avoiding images, phrasing or language that perpetuates unhelpful stereotypes or myths around suicide was also advised across media guidelines. In practice, such content was prevalent in the Twitter dataset including the notions that suicide is a "cry for help", that it constitutes a solution to one's problems (normalisation), or that suicide can be romantic (e.g. "they wanted to be together for eternity"). In Aaron Swartz's case, stereotyping took the form of "martyrdom" whereas in Amanda Todd's, tweets included vindication of suicide as a response to bullying (#71 "Think about what you write in case you hurt somebody, she doesn't deserve what happened"); the romanticisation of death (#8 "the world is more ugly but heaven has become more pretty #stopbullying"); the idea that individuals are only taken seriously after suicide (#154 "it's horrible that we only bother about amanda todd now she has died") and the argument that Todd was responsible for her own death (#864 "[…] not disrespecting but she brought this on herself"). Again because of the specific nature of the content of individual cases, it is difficult to identify such content ex ante, pointing again to guidelines being potentially more effective as an educational rather than regulatory tool.

Photographs, videos and notes

Half of the guidelines we reviewed referenced avoiding photographs or videos of the deceased, as well as links to their social media accounts and sites where they were being mourned. The intent behind this guideline is to avoid over-identification with the deceased and hence reduce the risk of "contagion". Given the hyperlinked nature of the Internet, however, this guideline is again difficult to translate online and, across the dataset, it was breached in different ways: for instance, through memorials to Aaron Swartz or fan art about Leelah Alcorn's death. This particular guideline also illustrates how there can be tensions across guidelines: guidelines that encourage comment on the impact on the bereaved, for example, may be undermined by those advising against links to social media sites.

Just under half the guidelines suggested the content of suicide notes should not be shared. This is increasingly relevant on social media given that some who die by suicide, (including Leelah Alcorn), choose to leave notes (or note-like texts) on their social media or blog, rather than (or in addition to) leaving an offline note. Most of the Twitter data related to this guideline involved Amanda Todd and Leelah Alcorn. Todd's video posted a month before her death can be read as akin to a suicide note and fulfils the criteria for why suicide notes' contents can be harmful: namely that they can glorify or romanticise suicide, present it as an acceptable option, or make it seem like a reasonable solution to one's problems. The number of people accessing this "document"—17 million views as of December 2012 (Baur and Ninan, 2019)—suggests the population potentially exposed to such risk is huge. Amanda Todd's parents, however, specifically advocated for her video to be kept on YouTube as a message *against* cyberbullying, and there have been debates over whether the video should be shown in Canadian classrooms (CBC News, 2012).

Conversely, a note left on Tumblr by Leelah Alcorn, which was auto-published the day of her death, became a source of contestation when her parents decided to take it down. Multiple mirrors[9] were created which contained not just the tweet but the rest of Leelah's Tumblr account. Those advocating for its removal were accused of transphobia, while those promoting its dissemination were accused of potentially placing others at risk. Again the potential for conflict between guideline objectives is clear—in this case between avoiding "contagion" in the immediate aftermath of a death by suicide and encouraging longer-term suicide prevention strategies through identification of identity oppression.

Insensitivity to the bereaved

The relational complexity involved in such deaths is relevant to another guideline commonly breached across the dataset—avoiding insensitivity to those who have been bereaved. Guidelines reviewed suggest that those bereaved by suicide are also at increased risk of suicide themselves. As we have already intimated however, sensitivity to one bereaved group (the online transcommunity for instance) can lead to insensitivity to another (parents). Balancing the needs of different groups in the aftermath of a death is particularly challenging on social media where context collapse and searchability lead to the easy retrieval of information. As will be explored further below, analysis of the Swartz and Alcorn cases suggest that when a death by suicide becomes politicised, there is also an increased chance of blame being directed at those around the deceased. While breaching of guidelines that relate to language use such as "committing suicide" could, as we suggested earlier, be a consequence of users' lack of knowledge, the volume, and nature of, the breaches relating to the bereaved, and to stereotyping, suggests that we need to return to Ahmed's argument about how value (and emotion) comes to attach to some bodies and not others. We do this by considering empathy (and its absence).

GUIDING EMPATHY?

Empathy can be understood as an act of perspective-taking and an emotional response to others that involves imaginative work (Lamm & Silani, 2014; Brownlie & Shaw, 2018) but it is also a social and political *relation* shaped by material and discursive contexts (Pedwell, 2014). As such, it plays out, often in unexpected ways, in relation to the guidelines and the practices they seek to discourage or encourage.

Tweets do not need to be empathetic to be consistent with the guidelines. For example, those that mention only facts, or that are simply news headlines, are not *actively* empathetic but nevertheless may follow the guideline on accurate reporting.[10] Moreover, practices that guidelines encourage can lead to empathy being expressed in ways that are unhelpful. For instance, in the case of Amanda Todd, the tragedy of her death was often framed in terms of her physical appearance and through descriptions of her as an "angel". On first reading, these tweets appear empathetic but they also reinforce the restrictive gendered norms that may negatively influence young women's mental health. It is also possible for tweets not to meet the guidelines and yet *be* empathetic. For example, as we have seen, guidelines deter the over-simplification of potential causes but it is possible to breach this guideline empathetically: "Swartz was bullied by US government until he died. He helped stop SOPA. Don't forget him. Retweet this" (#119).

Avoiding insensitivity to family, friends or the community of the deceased is one area where empathy might be assumed to be unproblematically aligned with the guidance. Guidelines treat family, friends and "community" in undifferentiated ways; in practice, however, as we have seen, people have different alliances, and empathy towards different groups or individuals can be in conflict. In Leelah Alcorn's case, tweets that were hostile towards her parents (e.g. #8 "#RIPLeelah you don't deserve this awful treatment, even by your parents") contravene guidelines encouraging empathy towards those left behind. The focus of considerable anger on social media for their reported views about their child's transgendered identity, these tweets are nevertheless rooted in empathy towards the deceased. Conversely, empathy expressed towards Leelah's parents could be read as expressing anti-transgender rhetoric or a lack of empathy towards the deceased.

There was evidence of lack of empathy across all the cases, but it was expressed most actively in relation to Amanda Todd. Qualitative analysis makes clear that an increasing number

of tweets engaged in a highly gendered discourse of blame in the form of "slut-shaming".[11] Lack of empathy was often expressed through disparaging comments about the amount of coverage of Amanda Todd's death, through jokes and through direct personal attacks on the deceased. In contrast, tweets about Aaron Swartz predominantly highlighted his life achievements, or his young age, and avoided discussion of his appearance or sexual conduct. The differentiated response to the deaths of Amanda Todd, Aaron Swartz, and also Leelah Alcorn, can be made sense of through the politicisation of the latter two deaths, and the depoliticisation of Amanda Todd's death, which was predominantly framed in terms of her individual lack of judgment and morals.[12] The differentiated way in which we value and judge others, while not explicitly addressed within the guidelines, are far from unusual on social media, particularly in relation to young women (Bindesbøl Holm Johansen et al., 2019).

Indeed, much of the negative portrayal of deaths by suicide on Twitter seems to be qualitatively different to the insensitivity mentioned in many of the guidelines. In some of these, there is an implicit assumption that the default view about death by suicide will be, if not broadly empathetic, at least not openly antagonistic. This assumption, however, is not borne out by analysis of the Twitter data where lack of empathy was often related to individual users' highly negative views about suicide, including that it is the fault of the deceased or indeed sinful: "Everyone talks about Williams…almost certain he is in hell now #donttakeyourlife" (646).

In the case of Amanda Todd, many of the comments drew on a "just deserts" discourse which takes the form of the deceased "deserving" (or not) their death and being "undeserving" (or deserving) of our concern and attention. Accounts of individuals being devalued in this way can be read as lacking empathy towards those who have been bereaved, though attempts to reduce the amount of coverage of individual suicides is consistent, in outcome, if not in tone, with suicide prevention guidelines. Lack of empathy often took the form of expressed anger about the extent of discussion of a particular death by suicide and the framing that the life in question did not merit such grief, for example "This Amanda Todd story is really annoying! Why is she special? Kids kill themselves because of bullying all the time" (235). Phillips' (2011, para. 9) work on trolling of memorial sites suggests that some of this anger could be push-back "against a corporate media environment that fetishizes, sensationalizes and commoditizes tragedy". As we saw earlier, however, jokes also became more prevalent as users became desensitised through their viral coverage.

The above analysis suggests that the socio-political relations of empathy need to be placed at the heart of discussion about the role of guidelines in discussing death by suicide on social media. Doing so does not make the *application* of guidelines to pre-existing content any easier; if anything, it may muddy the waters further. Nevertheless, a more nuanced, messy engagement with empathy is necessary to understand how and why guidance violations occur "all over" social media and what the possibilities and limitations are of guiding digital content in a different direction.

Our analysis makes clear that we need to pay close attention to the complex ways in which collective emotion is shaped by the affordances[13] of social media. Avoiding over-simplification, over-identification and over-linkage of information has long been at the heart of mainstream media strategies to avoid the Werther Effect, but the networked, immediate and hyperlinked nature of Twitter, along with its inbuilt rewarding of emotional resonant messages, makes all these strategies if not redundant, then extremely difficult to enforce. At the same time, wider sociocultural discourses which shape which lives are valued (and deaths judged as grievable), make clear that empathy needs to be understood as a social and political practice in traditional and social media. Platform materialities may undermine traditional media approaches but they also expand the circulation of discourses about value (and emotion), and hence reinforce the significance of empathy.

CONCLUSION

We have sought to raise questions about how useful, applicable and meaningful the notion of guidelines is in relation to discussions on Twitter of deaths by suicide. Those that exist, and that we drew on in this research, grew out of traditional print journalism. In the first half of this article, we offered a quantitative overview of the breaching of seven guidelines in the context of five case studies, exploring how and why, some were more translatable to the online than others. The quantitative analysis suggested breaches are commonplace with particular patterns emerging over time depending on the specific case. The qualitative analysis showed that Twitter communication is not subject to the journalistic standards of mainstream media and hence is often deliberately transgressive and/or offensive. Our qualitative analysis made clear that we need to move beyond generic claims of contagious talk about death by suicide to identify the specific challenges of applying guidelines to Twitter content, including the difficulty of retrieving instances of breaching because of the specificity of the deaths involved as well as the fact that guidelines can contradict each other and have unintended consequences. At the same time, the analysis points to the challenges arising from the nature of the platform itself; for instance, its hyperlinked nature, scale and immediacy—features which can make it difficult to contain or moderate virtual judgements about the deceased.

The complexity of applying guidelines to existing content means that a focus on the broader educational or preventative role of guidelines becomes pressing: if and how they can potentially help users to understand, or to imagine better, the implications of what they tweet. While the guidelines we reviewed and then operationalised in the analysis did not refer directly to empathy (or its lack), in the latter half of the analysis it became clear its presence or absence is what shapes much of the discussion about death by suicide on Twitter. Our analysis also suggests, however, that empathy needs to be addressed sociologically: it is a multi-relational and politicised and, as such, no less complicated or "grey" an area to address. Accentuated by the polarised nature of much debate on Twitter, we saw how empathy can be expressed for some at the expense of others, reinforcing in its turn existing divisions and hierarchies, including, gendered ones. This, and the fact that empathetic expressions and their absence are notoriously difficult to detect and interpret on social media (Brownlie & Shaw, 2018), mean that expanding guidelines to mandate for empathy is likely to be less productive than using these documents to educate about empathy as a social and political relation.

Returning to Butler's (2004) observation that some lives are more valued than others and their loss deemed more grievable, we need to understand how value and emotions in relation to deaths by suicide on social media are shaped by user demographics, by platform affordances, and by the culture of particular online spaces, as well as by wider sociocultural beliefs about suicide. The content in this article is troubling in so far as it is often transgressive and unempathetic but also inevitably contested. It is not possible given the project design, to know for sure the intent of those tweeting and, as Phillips (2011) and others have observed, it is possible for even the most aberrant content to be reframed more positively. What is less contestable, however, is that how suicide is discussed on social media *matters*. Twitter gives its users, and us as researchers, access to the complex social–cultural processes through which lives are grieved and deaths are positioned as valuable and "deserving" of our attention, processes which are intensified in the context of suicide. In Ahmed's terms, it reveals the power relationships that shape which emotions come to be attached to what objects. Twitter discussions about these deaths are predominantly read by the young, a group whose mental health is increasingly a focus of concern. This project cannot speak authoritatively to how suicide is discussed on other social media platforms. However, this and our previous analysis of Twitter—as a space in which the "unsayable" can be said because of the diffuse character of the platform (Brownlie & Shaw, 2018)—suggests that platforms where users feel part of a cohesive or tight-knit social structure may encourage substantively different conversations about suicide. These

conversations are likely to look more inhibited and careful in their framing and phrasing, particularly where there may be social consequences for breaking discursive norms around mental health. As well as platform affordances such as anonymity or pseudonymity, the social norms engendered by who populates a platform and a user's likely relationships with them will be influential in how suicide is discussed and empathy displayed.

Our research resonates with Perrin and Woods' (2018) argument that the public space of social media is too complex and fluid to establish a "duty of care" through detailed legal rules. Their suggestion that social media companies need to ensure their spaces are safe enough to prevent or reduce potential harms, however, understandably begs further questions about *how* this should happen. Theoretically, Bauman (1993, p. 144) has suggested that neither obedience to rules nor "combustive sociality" leaves much room for empathy and the empirical findings presented here support going beyond broad references to contagion and calls for the automatic application of guidelines, produced for the most part in a pre-digital world, to ask sociologically informed questions about why people discuss death by suicide online in the ways that they do, and about the role of empathy (and its absence) in these discussions.

AUTHOR CONTRIBUTIONS

Julie Brownlie: Conceptualization (lead); writing – original draft (lead); formal analysis (equal), supervision (lead), methodology (equal) **Justin Chun-ting Ho**: Methodology (equal); formal analysis (equal); writing – original draft (supporting); conceptualisation (supporting); software (lead) **Nikki Dunne**: writing - review and editing (equal); conceptualisation (supporting); Writing – original draft (supporting); investigation (equal); writing – original draft (supporting); **Nichole Fernández**: formal analysis (equal); conceptualisation (supporting) **Tim Squirrell**: formal analysis (equal); methodology (equal); writing – original draft (supporting); conceptualisation (supporting); review and editing (supporting).

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study

ORCID

Julie Brownlie ⓘ https://orcid.org/0000-0002-8299-6483

ENDNOTES

[1] For a discussion of the limitations of these approaches on Instagram, see https://www.telegraph.co.uk/technology/2019/11/13/instagram-removed-nearly-10000-suicide-self-harm-images-day/.

[2] This debate was reawakened in the UK following the death of Molly Russell in 2017—see (1).

[3] Different criteria led to slightly more tweets being included in this analysis than in the original study (Karamshuk et al., 2017).

[4] We would particularly like to thank Dr Frances Shaw for her work on this original analysis.

[5] These were as follows: Samaritans, the Scottish National Union of Journalists, Time to Change, the WHO, the New Zealand Government, the Suicide Prevention Resource Center, the American Framework for Suicide Prevention, Mindframe (an Australian government initiative), SAVE (an American charity which also owns and runs guidelines at reportingonsuicide.org and bloggingonsuicide.org), and suicide.org.

[6] A notable exception were guidelines from the Suicide Prevention Resource Center, a US-based charity, in collaboration with a project called TEAM Up, which have social media-specific guidelines (TEAM Up, 2014).

[7] Overall, active lack of empathy, not absence of empathy, was coded. For example, tweeting just the facts surrounding a suicide, as in a news headline, shows an absence, but not an active lack, of empathy.

[8] As positive guidelines are far less developed (John et al., 2014) and there were in any case only a few best practice guidelines emerging out of our review work, we focus on practices to be avoided.

[9] Sites which host the same content at different locations, often out of the control of the original platform (i.e. Tumblr in this instance).

[10] Where content showing a lack of empathy was re-tweeted but the user disputed the original tweet, such tweets were not coded for lack of empathy.

[11] Slut shaming is "the act of criticizing women or girls for their real or presumed sexuality or sexual activity, as well as for looking or behaving in ways that are believed to transgress sexual norms" (Karaian, 2014, p. 15).

[12] The research team that produced and analysed the original Twitter dataset identified a similar pattern of lack of empathy in relation to Amanda Todd (Karamshuk et al., 2017).

[13] Bucher and Helmond (2018) usefully extend the definition of affordances beyond the actions of platform users made possible by technology to the relations between platforms and a range of users, such as advertisers and developers, and to include the question of what users do to the technology.

REFERENCES

Ahmed, S. (2014) *The cultural politics of emotion*. Edinburgh: Edinburgh University Press.

Bauman, Z. (1993) *Postmodern ethics*. Oxford: Blackwell.

Baur, B. & Ninan, R. (2019) *Bullied teen's video passes 17M views*. Available at: https://abcnews.go.com/US/bullied-teen-amanda-todds-video-passes-13m-views/story?id=17548856 [Accessed 14th June 2019].

Bindesbøl Holm Johansen, K.,Pedersen, B.M. & Tjørnhøj-Thomsen, T. (2019) Visual gossiping: Non-consensual 'nude' sharing among young people in Denmark. *Culture, Health & Sexuality*, 21(9), 1029–1044.

Boero, N. & Pascoe, C.J. (2012) Pro-anorexia communities and online interaction: Bringing the pro-ana body online. *Body & Society*, 18(2), 27–57.

Boyd, D., Ryan, J. & Leavitt, A. (2011) Pro-Self-Harm and the Visibility of Youth-Generated Problematic Content (osu.edu). *ISJLP*, 7(1), 1–32.

Brady, W., Wills, J., Jost, J., Tucker, J. & Van Bavel, J. (2017) Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences USA*, 114(28), 7313–7318.Available from: https://doi.org/10.1073/pnas.1618923114

Brownlie, J. (2011) 'Being there': Multidimensionality, reflexivity and the study of emotional lives. *British Journal of Sociology*, 62(3), 462–481.

Brownlie, J. (2018) Looking out for each other online: Digital outreach, emotional surveillance and safe(r) spaces. *Emotion Space and Society*, 27, 60–67.

Brownlie, J. & Shaw, F. (2018) Empathy rituals: Small conversations about emotional distress on twitter. *Sociology*, 53(1), 104–122.

Bucher, T. & Helmond, A. (2018) The affordances of social media platforms. In: Burgess, J., Poell, T. & Marwick, A. (Eds.) *The SAGE handbook of social media*. London and New York, NY: SAGE Publications Ltd, pp. 233–252.

Burgess, J., Mitchell, P. & Münch, F. (2018) Social media rituals: The uses of celebrity death in digital culture. In: Papacharissi, Z. (Ed.) *A networked self and birth, life, death*. New York, NY: Routledge, pp. 224–239.

Butler, J. (2004) *Precarious life: The powers of mourning and violence*. London: Verso.

CBC News (2012) *B.C. teachers have option of showing Amanda Todd video*. Available at: https://www.cbc.ca/news/canada/british-columbia/b-c-teachers-have-option-of-showing-amanda-todd-video-1.1249899 [Accessed 22nd October 2019].

Costa, E. (2018) Affordances-in-practice: An ethnographic critique of social media logic and context collapse. *New Media & Society*, 20(10), 3641–3656.

Crawford, K. & Gillespie, T. (2014) What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society*, 18(3), 410–428.

Döveling, K., Harju, A.A. & Sommer, D. (2018) From mediatized emotion to digital affect cultures: New technologies and global flows of emotion. *Social Media+ Society*, 4(1), 1–11.

Durkheim, E. (1915) *The elementary forms of religious life. A study in religious sociology*. New York, NY: MacMillan.

Fahey, R.A., Matsubayashi, T. & Ueda, M. (2018) Tracking the Werther effect on social media: Emotional responses to prominent suicide deaths on Twitter and subsequent increases in suicide. *Social Science & Medicine*, 219, 19–29.

Franzke, A.S., Bechmann, A., Zimmer, M., Ess, C. & The Association of Internet Researchers (2020) *Internet research: Ethical guidelines 3.0*. Available at: https://aoir.org/reports/ethics3.pdf [Accessed 20th March 2020].

Gerrard, Y. (2018) Beyond the hashtag: Circumventing content moderation on social media. *New Media & Society*, 20(12), 4492–4511.

Gibson, M. (2015) Automatic and automated mourning: Messengers of death and messages from the dead. *Continuum*, 29(3), 339–353.

Gillespie, T. (2018) *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, CT: Yale University Press.

Gonçalves, E., Araújo, A., Benevenuto, F. & Cha, M. (2013) Comparing and combining sentiment analysis methods. In: *Proceedings of the first ACM conference on online social networks (COSN '13)*. New York, NY: Association for Computing Machinery, pp. 27–38.

Gonçalves, E., Araújo, A., Benevenuto, F. & Cha, M. (2013) Comparing and combining sentiment analysis methods. In: *Proceedings of the first ACM conference on online social networks (COSN '13)*. New York, NY: Association for Computing Machinery, pp. 27–38.

Gould, M.S., Kleinman, M.H., Lake, A.M., Forman, J. & Midle, J.B. (2014) Newspaper coverage of suicide and initiation of suicide clusters in teenagers in the USA, 1988–96: A retrospective, population-based, case-control study. *Lancet Psychiatry*, 1, 34–43.Available from: https://doi.org/10.1016/S2215-0366(14)70225-1

Horne, J. & Wiggins, S. (2009) Doing being 'on the edge': Managing the dilemma of being authentically suicidal in an online forum. *Sociology of Health & Illness*, 31, 170–184.Available from: https://doi.org/10.1111/j.1467-9566.2008.01130.x

Hsieh, H.F. & Shannon, S.E. (2005) Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288.

Jamieson, P., Jamieson, K.H. & Romer, D. (2003) The responsible reporting of suicide in print journalism. *American Behavioral Scientist*, 46(12), 1643–1660.Available from: https://doi.org/10.1177/0002764203254620

Jane, E.A. (2017) 'Dude… stop the spread': Antagonism, agonism, and# manspreading on social media. *International Journal of Cultural Studies*, 20(5), 459–475.

Javid, S. & Wright, J. (2019) *Online harms white paper*. UK Government Department for Digital, Culture, Media & Sport, Home Office. Available at: https://www.gov.uk/government/consultations/online-harms-white-paper [Accessed 25th July 2019].

John, A., Hawton, K., Lloyd, K., Luce, A., Platt, S., Scourfield, J. et al. (2014) PRINTQUAL a measure for assessing the quality of newspaper reporting of suicide. *Crisis*, 35, 431–435.

Karaian, L. (2014) Policing 'sexting': Responsibilization, respectability and sexual subjectivity in child protection/crime prevention responses to teenagers' digital sexual expression. *Theoretical Criminology*, 18(3), 282–299.

Karamshuk, D., Shaw, F., Brownlie, J. & Sastry, N. (2017) Bridging big data and qualitative methods in the social sciences: A case study of Twitter responses to high profile deaths by suicide. *Online Social Networks and Media*, 1, 33–43.Available from: https://doi.org/10.1016/J.OSNEM.2017.01.002

Lamm, C. & Silani, G. (2014) Insights into collective emotions from the social neuroscience of empath. In: von Scheve, C. and Salmela, M. (Eds.) *Collective emotions: Perspectives from psychology, philosophy, and sociology*. Oxford: Oxford University Press, pp. 63–77.

Litt, E. & Hargittai, E. (2016) The imagined audience on social network sites. *Social Media + Society*, 2(1), 1–12.

Marchant, A., Brown, M., Scourfield, J., Hawton, K., Cleobury, L., Dennis, M. et al. (2020) A content analysis and comparison of two peaks of newspaper reporting during a suicide cluster to examine implications for imitation, suggestion, and prevention. *Crisis*, 41(5), 398–406.Available from: https://doi.org/10.1027/0227-5910/a000655

Markham, A. (2012) Fabrication as ethical practice: Qualitative inquiry in ambiguous internet contexts Information. *Communication & Society*, 15(3), 334–353.

McCosker, A. (2014) Trolling as provocation: YouTube's agonistic publics. *Convergence*, 20(2), 201–217.

Mueller, A.S. (2017) Does the media matter to suicide?: Examining the social dynamics surrounding media reporting on suicide in a suicide-prone community. *Social Science & Medicine*, 180, 152–159.

Niederkrotenthaler, T., Till, B. & Garcia, D. (2019) Celebrity suicide on Twitter: Activity, content and network analysis related to the death of Swedish DJ Tim Bergling alias Avicii. *Journal of Affective Disorders*, 245, 848–855.

O'Dea, B., Larsen, M.E., Batterham, P.J., Calear, A.L. & Christensen, H. (2017) A linguistic analysis of suicide-related twitter posts. *Crisis*, 38, 319–329.Available from: https://doi.org/10.1027/0227-5910/a000443

Papacharissi, Z. (2015) Affective publics and structures of storytelling: Sentiment, events and mediality. *Information, Communication & Society*, 19(3), 307–324. https://doi.org/10.1080/1369118X

Pedwell, C. (2014) *Affective relations: The transnational politics of empathy*. New York, NY: Palgrave Macmillan.

Perrin, W. & Woods, L. (2018) Reducing harm in social media through a duty of care. *LSE Media Policy Project Blog*. Available at: https://blogs.lse.ac.uk/mediapolicyproject/2018/05/10/reducing-harm-in-social-media-through-a-duty-of-care/ [Accessed 25th July 2019].

Pew Research Center (2019) *Sizing up twitter users*. Available at: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/04/twitter_opinions_4_18_final_clean.pdf [Accessed 26th May 2020].

Phillips, D.P. (1974) The influence of suggestion on suicide: Substantive and theoretical implications of the Werther effect. *American Sociological Review*, 39, 340–354.

Phillips, W. (2011) LOLing at tragedy: Facebook trolls, memorial pages and resistance to grief online. *First Monday*, 16(12). Available at: https://uncommonculture.org/ojs/index.php/fm/article/view/3168/3115 [Accessed 26th July 2019].

Robinson, J., Cox, G., Bailey, E., Hetrick, S., Rodrigues, M., Fisher, S. et al. (2016) Social media and suicide prevention: a systematic review. *Early Intervention in Psychiatry*, 10, 103–121.Available from: https://doi.org/10.1111/eip.12229

Rogers, R. (2013) *Digital methods*. Cambridge, MA: The MIT Press.

Sampson, T., Maddison, S. & Ellis, D. (Eds.) (2018) *Affect and social media: Emotion, mediation, anxiety and contagion*. London: Rowman & Littlefield International.

Scott, M. (2016) *Step-by-step guide to WordSmith*. Oxford: Oxford University Press. Available at: https://lexically.net/wordsmith/step_by_step_English7/overviewofkeywords.html [Accessed 2nd October 2019].

TEAM Up (2014) *Social media guidelines for mental health promotion and suicide prevention*. Available at: http://www.eiconline.org/teamup/wp-content/files/teamup-mental-health-social-media-guidelines.pdf [Accessed 3rd October 2019].

The Guardian (2019) The Guardian view on online harms white paper grey areas. *The Guardian*, 8 April. Available at: https://www.theguardian.com/commentisfree/2019/apr/08/the-guardian-view-on-online-harms-white-paper-grey-areas [Accessed 25th July 2019].

Thelwall, M. & Kappas, A. (2014) The role of sentiment in the social web. In: Von Scheve, C. and Salmella, M. (Eds.) *Collective emotions: Perspectives from psychology, philosophy, and sociology*. Oxford: Oxford University Press. pp. 375–388.

Tromble, R., Storz, A. & Stockmann, D. (2017) We don't know what we don't know: When and how the use of twitter's public APIs biases scientific inference. *SSRN Electronic Journal*, 1–26. https://doi.org/10.2139/ssrn.3079927

Ueda, M., Mori, K., Matsubayashi, T. & Sawada, Y. (2017) Tweeting celebrity suicides: Users' reaction to prominent suicide deaths on Twitter and subsequent increases in actual suicides. *Social Science & Medicine*, 189, 158–166.

World Health Organization (2019). *Suicide fact sheet (online)*. Available at: http://www.who.int/mediacentre/factsheets/fs398/en/ [Accessed 20th June 2020].

Zappavigna, M. (2012) *Discourse of Twitter and Social Media: How We Use Language to Create Affiliation on the Web*. London: Continuum.