Looking out for each other online

Looking out for each other online: Digital outreach, emotional surveillance and safe(r) spaces

Julie Brownlie

School of Social and Political Science, 22 George Square, Rm1.4, University of Edinburgh, Edinburgh EH8 9JU, Scotland, United Kingdom

A B S T R A C T

This article is concerned with what it means to think of online spaces as emotionally safe or safer. It does this by looking at the sharing of emotional distress online and the role of organisations in identifying and proactively engaging with such distress. This latter type of digital engagement is analytically interesting and rendered increasingly feasible by algorithmic developments, but its implications are relatively unexplored. Such interventions tend to be understood dualistically: as a form of supportive digital outreach *or* as emotional surveillance. Through an analysis of blog data about a Twitter-based suicide prevention app, this article attempts to understand the tensions and also the potential points of connection between these two meanings of 'looking out for each other' online. From an avowedly sociological, relational and emotional perspective, it tries to offer a more nuanced account of what it might mean to share emotional distress 'safely' online.

This article engages with what it means to think of an online space as emotionally safe or, recognising the impossibility of guaranteeing safety completely or indefinitely, saf*er*. Drawing on a case study of Radar, a Twitter based suicide prevention app introduced and then withdrawn by a leading UK charity, it examines how and why people share emotional distress on Twitter and the implications for organisations who choose to engage proactively with such distress. Such engagement has tended to be framed as either surveillance *or* harm-preventing 'outreach'; as expressions, for instance, of digital capitalism and moral entrepreneurship (Reeves, 2017) *or*, more positively, digital caring and 'digital professionalism' (Ellaway et al., 2015). Yet the ensuing debates are rarely grounded in research on experiences of such practices (Robert et al., 2015). Bringing these two framings into conversation is important precisely because organisations thinking of outreach need to understand practices *already* happening in different online spaces.

Creating dialogue requires careful exploration of differences and also potential shared concerns. In the case of suicide prevention apps, the latter include anxieties about risk: anxieties drive the expansion of surveillance in general (Crawford, 2014) but they are particularly powerful in the context of suicide prevention because of the fatal consequences of 'signs' being overlooked. Practices to identify such signs can be thought of as a form of emotional surveillance. Adapting Lyon's (2007: 14) definition of surveillance, emotional surveillance can be understood as 'focused, systematic and routine attention' being paid to information that appears to be relevant to our emotions 'for purposes of influence, management, protection or direction'. Surveillance is also emotional, however, in the sense that Crawford (2014) notes above: that there are affective consequences for both those doing the surveillance and being surveilled. The consequences for the latter are familiar from the literature on surveillance but surveillers, too, can be 'haunted by a very particular kind of data anxiety: that no matter how much data they have, it is always incomplete'.

The overall aim here is not, therefore, to add to existing critiques of the Radar app (Lee, 2014; Reeves, 2017), but to inform practice and conceptual debates about 'safe sharing' of emotions online. This involves moving beyond a focus on individuals, and their rights to privacy (though see Bernal (2014) for an account of privacy as a communal value), towards a more *emotional* and *relational* approach to understanding online sharing. By this, I mean considering pre-existing relationships online as well as relationships between those doing, and being on the receiving end, of the outreach/surveillance. This approach also means engaging with what people believe they are doing by expressing emotions in particular spaces: keeping safe on Twitter, for instance, may involve the expression of apparently unsafe or risky feelings.

Such situated analyses of what is being termed here digital outreach are necessary, as it is meaningless to speak of outreach or the sharing of emotional distress online *in general* (Hines, 2015). Before outlining the methodology for the study, the Radar app is first positioned in the

E-mail address: Julie.Brownlie@ed.ac.uk.

context of recent work on digital outreach and on the sharing of emotions and online spaces more generally.[1]

1. Online distress and digital outreach

While it is not always possible to clearly distinguish between the two (Tucker and Goodings, 2017), there is increasing recognition of the significance of both informal emotional support on social media (Eggertson, 2015) and digital outreach, a proactive form of support facilitated by organisations such as voluntary or government bodies. The challenges of this latter type of 'digital professionalism' (Ellaway et al., 2015) are debated but, as Facebook's recent announcement about the use of algorithms to identify suicidal users makes clear (Kelion, 2017), it is increasingly presumed that social networking platforms have the potential to 'promote positive change' in relation to mental health (Inkster et al., 2016) and to offer digital 'safe spaces'.

At the same time, however, there are concerns that such outreach can become a form of 'health surveillance' (French and Smith, 2013; Lupton, 2012). Regardless of whether or not it is framed as emotional surveillance, the complexities of how selves are presented, on and off-line (Murthy, 2012), make the interpretation of online content by those engaged in online support difficult. This is particularly evident in relation to suicide (Mok et al., 2016). Reflecting wider dichotomous views of the Internet, it's role in relation to suicide is often perceived as Manichean: a struggle between dark/death and light/life (Robert et al., 2015). Social media is increasingly part of this context (Christensen, 2014). There is evidence, on the one hand, that informal online suicide communities can be valued by participants for their mutuality (Baker and Fortune, 2008); on the other, of suicidal postings not being acknowledged or of forums maintaining or even amplifying suicidal feelings (Mok et al., 2016). Similar concerns about how online sharing may reinforce or exacerbate risky behaviours have also been noted in relation to other practices (Cantó-Milà and Seebach, 2011).

Reflecting the ambiguity of emotional expression, the task of distinguishing between harmful and helpful content in these contexts is complex: apparently 'dangerous' text can act as a deterrent, be life-affirming or empowering, and be read differently over time (Mars et al., 2015). Suicide prevention apps (resources designed to support a person in distress) are increasingly a part of this complex digital landscape. They range from social media interventions[2] to smart phone self-help apps[3] (Aguirre et al., 2013). Facebook has been the platform most closely associated with these developments[4] but, in 2014, Samaritans, a charity that provides emotional support to people experiencing distress, launched a Twitter-based app, Radar,[5] under the tagline, 'turn your social net into a safety net'[6]. This allowed registered users to be alerted when a Twitter account they followed included messages that might suggest depressed or suicidal thoughts. Although media and other responses to the app were initially positive, it proved increasingly controversial and was withdrawn nine days after its launch. To understand this, we need to look beyond digital outreach to what it means to share emotions in online spaces and how such spaces come to be thought of as safe or safer.

2. Materiality, relations and emotions in and across (online) space

Notions of 'space', including 'safe spaces', are key to the sharing of emotion online. An understanding of space as not innate but constituted and reconstituted through our actions and interpretations is long-standing (Bondi, 2005; Cronin, 2014) and has been applied to online settings (Marino, 2015). However, despite initial framings that digital space might involve the collapsing of time and the overcoming of materialities, the relations through which *all* space is produced remain stubbornly material (Massey, 2005). The 'new' material turn has placed matter at the heart of analysis of space, including online, reminding us, as Lehdonvirta (2010: 885) put it, that the online world is less an 'open frontier' than a 'built-up' area. This materiality is relevant not just because, as Fayard (2012) notes, virtual space involves 'a lot of stuff', hard and software, but because all users of such spaces are embodied and embedded in particular places (Hines, 2015) and because online spaces have socio-material consequences. In other words, space – online or otherwise – is constituted through, but also shapes, social relations (Lefebvre, 1991): it can keep people in (their) place or help them move. Not surprisingly, then, online space is often made sense of through material metaphors. This is true, too, of the sharing of emotion in online space, though the metaphors used in this context speak to the more ethereal aspects of offline space: atmosphere (Tucker and Goodings, 2017), intimacy (Michaelsen, 2017) or ambiance (Thompson, 2008).

The relational emphasis on understanding space, noted in the above discussion of materialities, is often framed in an online context in terms of sharing. Despite being under-conceptualised (Kennedy, 2015), sharing is key to analysing online space because the Internet is both a space where sharing happens and is made up of spaces constituted *through* such sharing. Because the online realm is made up of relational spaces, it is also emotional. In addressing the sharing *of emotion* specifically, I am concerned with both the meanings identified above: that is, to give and receive emotion, and to do so jointly within the same space – but also with a third, less researched dimension, that is, to have the same understanding of an emotion (or why that emotion was expressed) as another. Research on the role of technological affordances, social relations and norms on sharing across different social media suggest a complex interplay between the three (Bucholtz, 2013). To understand this interplay involves going beyond what it means technically to be a Twitter user/follower (Bruns and Moe, 2014) to examine the meaning Twitter users give to the sharing of emotion and how this sharing constitutes, as much as is mediated by, space.

2.1. Constituting safe(r) spaces

When people talk about safe spaces, notions of the relational and emotional are present but not always foregrounded. Safe spaces are 'imaginary construction [s]' (Stengel, 2010:524) that involve complex boundary work in relation to the imagined 'unsafe' (Rosenfeld and Noterman, 2014). The impossibility of wholly or indefinitely achieving exclusion or inclusion ensures that the boundaries and the spaces created are always porous, contestable and shifting and for this reason it makes sense to refer to safer rather than safe spaces.

Discourses about 'safety' in relation to mental health, race, class or sexuality (Haber, 2016), for instance, are sometimes framed around the exclusion of others and, at other times, through inclusion – for instance, the idea of a safe space for *all*. The idea of being 'safe to' express oneself, without repercussion and perhaps in contestation of dominant discourses, is core to the creation of a space as safe but cannot always be separated from the idea of being 'safe from'. This speaks to a long history of the public sphere as a space of surveillance and exclusion but also, especially for marginalised groups, as a space of radical potential, a 'haven' away from the oppressions of the private sphere (Haber, 2016). Online spaces, because of their relative anonymity and ease of access, have been seen as potentially offering increased access to safe(r) spaces, though these same features can also increase the risk of feeling

---

[1] While, of course, it is possible for practices to be digital and not involve the Internet, to reflect common usage when discussing social media, digital is used here synonymously with online. In order to avoid reinforcing digital dualisms (Jurgenson, 2011), however, the description 'virtual' has been avoided, except when referencing other authors' use of the term.

[2] See http://www.durkheimproject.org/.

[3] http://www.prevent-suicide.org.uk/stay_alive_suicide_prevention_mobile_phone_application.html.

[4] See http://www.intheforefront.org/forefront-and-facebook-launch-suicide-prevention-tool.

[5] http://www.samaritans.org/news/samaritans-radar-announcement.

[6] http://www.samaritans.org/sites/default/files/kcfinder/branches/branch-96/files/Samaritans%20Radar%20Walkthrough.pdf.

unsafe.

Safe(r) spaces depend then not just on who is included or excluded and what is expressed but on how those who are present *listen* and respond. Reflecting a wider neglect of listening in the social sciences (Brownlie, 2014), much research on digital sharing still tends to focus on expression (Crawford, 2009). Yet listening to/reading emotional expression in public online spaces, including Twitter, involves considerable emotional work. The therapeutic remains a dominant narrative shaping understandings of the 'right' amount of sharing online as offline (Illouz, 2008): too much information ('tmi') is problematised as leading to an over-exposed self. Conversely, undersharing raises concerns within social networks about those who 'go silent'. In both cases, listening is implicit but critical.

Some have suggested there are 'rules for sharing' experiences of emotional pain and suffering online (Sandaunet, 2008), and that these are space- and subject-specific (Hess, 2015). But the fluidity and complexity of sharing (through both expressing and listening) in moderated and unmoderated spaces (Skeggs and Yuill, 2015) suggests that 'rules' may be too rigid a description for what shapes such sharing. Instead, emotional reflexivity, the way in which emotions are drawn on to negotiate social life, might offer greater analytic purchase (Burkitt, 2014).

Radar's introduction into Twitter was a moment of contestation around the meaning of an online safe space: positioned by some as a form of digital outreach that could constitute a safe space but understood by others as breaching a pre-existing safe space. In what follows, I explore these framings through a focus on bloggers' reflexive accounts of what it means to be emotionally 'safe(r)' online, but I begin by outlining how these accounts were identified and analysed.

3. Researching Radar

The following analysis draws on a sample of blogs about Radar identified from a search of all tweets that contained #Samaritans Radar or keywords 'Samaritans' and 'Radar', identified via the Twitter API in the two weeks following the launch of the app.[7] This search produced 6540 tweets and the blogs drawn on here are the ones mentioned in the top 100 (ranked by number of retweets and favourites). The resulting 34 blogs were produced by a range of authors including academics, journalists, mental health writers and activists. Typically, they were not opinion pieces written by detached commentators but were from engaged Twitter users.

The project focused on blogs accessed *through* Twitter rather than communication *on* Twitter itself. While analysis of tweets might allow for exploration of reaction to Radar on Twitter, the question of how Twitter users' account for their sharing of emotion on this platform, in other words, their emotional reflexivity, is better accessed through their expansive and discursive blog writing. In other words, blogs offered Twitter users (and researchers) a degree of distance from their Twitter practice.

There are good analytical reasons, however, for accessing this blog sample *through* Twitter. First, the Twitter dataset, produced as part of a larger study on emotional distress and digital outreach, offered the possibility of a comprehensive sampling frame of active Twitter users writing on this topic. Though there are limitations to the Twitter API that mean neither the Twitter nor the blog dataset can be treated as complete, it would not have been possible to sample these blogs systematically via a general internet search engine. Second, sampling through Twitter allowed identification of blogs and opinions that were being actively engaged with – or at least publicised – by Twitter users following the Radar debate. These blogs might reasonably, then, be regarded as a 'long form' version of the Twitter debate about Radar. Finally, analysis of the Twitter dataset suggested, not surprisingly

perhaps, that themes relating to safety and the sharing of emotion emerged in both the Radar-related blogs *and* the larger dataset of tweets from which the blogs were identified. This offers some reassurance about the prevalence of these themes beyond the blog sample – again, this would not have been possible if the blogs had been identified through Google. As noted, however, my analytical concern here is not with quantifying discourses but rather with *how* those who supported and resisted the app framed their understanding, hence the decision to focus solely on blogs.

As with other forms of documentary analysis, blog analysis can involve structural, thematic and narrative dimensions (Elliot et al., 2016). In particular, blogs can be read as performances of narrative identity and as constituting particular communities (Schuurman, 2014). While a focus on such community formation could inform further analysis, the focus here is on mapping ways of writing and thinking about being safe (r). To this end, blogs were coded thematically and were then re-read with a focus on the use of metaphors. How metaphors shape the way we think has long been of academic interest and now includes a focus on the online (Lakoff and Johnson, 2011; Markham, 2013).

Ethically, no assumptions were made about the public nature of blogs unless the blogger's page was explicitly linked to an organisation, such as a newspaper (AOIR, 2012). In all other cases bloggers were contacted via email to advise them of the research project and to seek consent to quote from, and link to, their blog. Only where consent was given were blog data used in these ways.[8]

The analysis draws on a particular understanding of emotion. Much research on the sharing of emotions through social media has drawn on affect (Hillis et al., 2015). Understandings and definitions of affect vary widely: for some, it involves a visceral, bodily or 'gut' reaction beyond intent, though increasingly there is a move towards recognising its relational and discursive elements (Veletsianos and Stewart, 2016). Many of these framings of affect share a focus on its productivity, on what it *does* (Kennedy, 2015) and have, therefore, been useful in helping to think through how social media both engages and produces emotion (Hillis et al., 2015) In work focused on media affordances, however, how emotions are constituted by and constituting of particular relationships, as well as spaces, can get lost. To this end, the analysis which follows is informed by an understanding of our capacity for emotional reflexivity and of emotions as constituted, and made sense of, through relational processes (Brownlie, 2014; Burkitt, 2014).

4. Looking out for each other online: outreach versus surveillance

At first glance, analysis of the blogs suggested that the response to Radar could indeed be understood through the polarised framework of emphasising benefits (harm reduction through outreach) *or* risks (those associated with surveillance). The app was introduced on the basis that it could potentially save lives and, initially, was positively received. In the two week period after its launch, however, the metaphors appearing in the blogs shifted towards the discourse of surveillance. The announcement of the petition for the app to be withdrawn, for instance, described Radar as having been launched 'behind people's backs' on 'unsuspecting Twitter users'.[9] One blogger who writes about data protection[10], referred to the app through the metaphors of 'profiling' (http://bit.ly/2A9W9uc) while another, who writes on mental health and disability, referred to 'Orwellian' practices (http://bit.ly/2Bcoo8y). The peculiarly human nature of the work the app was being programmed to do – identify emotional distress – meant it was metaphorically positioned by some bloggers as cyborg-like, a cross-over between a 'suicide bot and concerned friend' (Hess, 2015). To a degree

[8] In just two cases, blog authors did not respond and their consent was, therefore not assumed.

[9] https://www.change.org/p/twitter-inc-shut-down-samaritans-radar. Links to blogs are embedded in the text; other links, like this one, are included in footnotes.

[10] Where possible, I have used bloggers' own descriptions of themselves.

this points to the particular ambiguity of the Radar app: introduced by an organisation and hence akin to traditional centralised surveillance practices, yet dependent on Twitter users signing up for the app to be reminded of what their followers have been up to, hence closer to peer to peer surveillance.

The app's ambiguous position was further accentuated by the commercial context in which it was developed (Mason, 2014). Drawing on metaphors of experimentation, one mental health blogger suggested Twitter users were being positioned as 'subjects for your tech' (http://bit.ly/2Bcoo8y), reflecting an unease with such developments even when, as in the case of Radar, they were packaged as 'big data for good'.

On first reading, then, the conventional (individualised) polarities associated with digital outreach, of reducing harm *or* increasing the risk of surveillance, are evident in discussion of the app. As the debate unfolded it became more polarised, a process accentuated by the adversarial nature of the platform. Yet it is also clear that both 'sides' acknowledged the significance of finding support online and both expressed anxieties about making Twitter a safe space. Re-reading the blog data through an understanding of safe(r) spaces as materially, relationally and emotionally achieved helps make sense of polarised views of how digital caring happens or is imagined to happen. Discourses, often metaphorically expressed, run through all these dimensions and, in what follows, I consider online safe(r) space discretely through each. This, though, is a heuristic separation as they are in practice co-constituting.

5. Constructing the materiality of Radar space

Twitter, like all space, is constantly under construction, produced through everyday relationships, practices, emotions, materials and the discourses that revolve around all of these. In launching Radar and in subsequent statements, Samaritans sought to introduce the app as offering Twitter users a 'second chance'[11] to see tweets from someone they knew who might be struggling to cope. In other words, they sought to make the technical case that the tweets that triggered app alerts were *already* appearing in the Twitter feeds of the app subscribers.[12]

Bloggers' response to this justification, however, reflected a by now well-established critique of the dualistic understandings of public/private developed through work on the nature of visibility on social network sites (Marwick and boyd, 2014). In particular, some bloggers described how they constituted different spaces *within* 'public' Twitter through the strategic placing in tweets of the @ sign or a period in order to restrict or expand their potential audience (for an explanation of how this works in practice, see http://bit.ly/1fkfKbI). In this extract, for instance, a blogger who writes on digital privacy explains that such practices allow him and others to use the Twitter space in ways that are less public than those deployed by celebrities such as Stephen Fry:

> Not all Stephen Fry. Not all tweets equally visible – can make more intimate and more public by using @ and also by using '.'. (http://bit.ly/2BqLrNR)

Reflecting a tendency, noted earlier, to perceive digital space through physical metaphors (Haber, 2016), Samaritans in their naming of the app and through their core metaphor of Radar as a 'safety net' sought to make sense of digital space through material offline practices and objects. Similarly, offline *place* metaphors were important for how bloggers attempted to understand Twitter as a digital safe space. Place has long been associated with understandings of safeguarding: 'the basic character of dwelling is safeguarding' (Heidegger, 1978:352 cited in Martin, 2017) and this is also the case with online dwellings. In the following extracts, bloggers are trying, in particular, to think through

the potential for privacy on Twitter. One blogger, an academic, turns to offline surveillance practices (CCTV) to do so, while another turns to offline places (the pub, the dinner table) and communicative practices (whispering):

> my office window looks out on a public street – whatever people do there is public. There would still, though, be privacy issues if I installed a video camera in my window to tape what people did outside. (http://bit.ly/2BxoUOF)

> You can have an intimate, private conversation in a public place – whispering to a friend in a pub, for example […]. Chatting around the dinner table when you don't know all the guests – where would that fit in?' (http://bit.ly/2BqLrNR)

Crucially, though, these place metaphors are also a means of imagining *who* is listening (Litt and Hargittai, 2016). Like those who have written about online 'places' as having 'residents' and 'clusters of friends and colleagues' (White and LeCornu, 2011 cited in Veletsianos and Stewart, 2016), bloggers constitute Twitter, not just as a material and discursive space but also as a *relational* one.

6. Safe(r) relations

From the outset, Radar was conceived by Samaritans as a relational tool; a means for all of us to look out for each other online. This is consistent with a shift towards the democratisation of helping: the idea that one does not need to be trained to offer help to others (Lee, 2014). Those who initiated Radar had in mind a diffuse but benign audience: in other words, the 'attitude of the whole community' (Mead, 1934) was imagined to be supportive. In an update following the launch, Samaritans suggested that the app could be aimed at those 'who are more likely to use Twitter to keep in touch with friends and people they know'.[13] One blogger, who writes about technology, also suggested that the app might work better if it was restricted to reciprocal arrangements with people 'who the at-risk person is following back – so theoretically only friends' (http://bit.ly/2A9YdSY).

Different platforms offer different understandings of safe relations and this focus on friends is closer to the generalised 'attitude' of the Facebook rather than Twitter community. Facebook involves known networks, and indeed most research on emotional support has been focused on the platform for this reason (Burke and Develin, 2016). Yet there are powerful norms that make it difficult to seek emotional support on Facebook (Buehler, 2017) and, for some, its very interconnectedness makes it a less than safe space for sharing emotional distress. Affordances of other platforms may mean that sharing is experienced as potentially less stigmatising with feelings expressed, for instance, through images and through second (Instagram) or 'throwaway' (Reddit) accounts (Andalibi, 2017) in a way that is not about maximising visibility.

Twitter followers are not necessarily friends in the sense of intimates; nor are followers who are friends, and *are* intimates, necessarily able or willing to take on the responsibility of 'looking out' for others. Indeed, there are significant risks in assuming friends and/or followers are in a position to reply or, as a blog below from a mental health service user suggests, that their response would be useful or welcomed. Indeed, they might in fact have a 'chilling' effect (Hess, 2015):

> It's not just having the information, it's being able to *do* something useful with it. If it just flags up that you need to try and connect with and support a person, that's one thing, but what if people who know very little about mental health sign up and wake up to a worrying email? Would they have enough info to call police/ambulance?

---

[11] http://www.samaritans.org/how-we-can-help-you/supporting-someone-online/samaritans-radar.

[12] http://www.samaritans.org/how-we-can-help-you/supporting-someone-online/samaritans-radar#4nov.

[13] http://www.samaritans.org/how-we-can-help-you/supporting-someone-online/samaritans-radar#30oct.

SHOULD they – would the person welcome this? (http://bit.ly/2Bcoo8y)

One blogger who writes on disability and mental health expressed concern that app users might be potentially deluged by 'ghost patterns' (Crawford, 2014): 'if you use [Radar], you're going to get a hell of a lot of spam, unless you only follow a few close friends' (http://bit.ly/2zEvAOz). Such anxieties, however, might also be about whether apps can capture the relational nuances of friendships and about the authenticity of those friendships that depend on the app: would a *real* friend need technological help to become aware of a known other's emotional upset?

> If individuals need to be poked to remember to check on someone, they very much might not be the type of person someone [….] wants to talk to. (http://bit.ly/2AchWiH)

There are ways of displaying friendship, online and offline, that shore up authenticity; and, conversely, friendship norms can be breached through inappropriate sharing or use of technology (Bazarova, 2012):

> Friends are people who although I have only met them via Twitter care enough about me to check in with me before they go to bed and when they wake up, who states openly, "I am really concerned about you, how can I help?" They don't need an app to record what I have been saying while they were asleep or at work, because they simply message me as soon as the[y] wake/return asking how I am and telling me they are thinking of me. (http://bit.ly/2Bcoo8y)

The clash of norms arising here is from the use of an algorithm[14] in relation to the emotional support and specifically about the automation of friendship: the speeding up, delegation and mechanisation of what it means to be a friend. In the context of online dating, this is what Peyser and Eler (2016) refer to as the 'tinderization' of emotion. Research is beginning to emerge on the limitations of 'conversational agents' such as Siri in responding to distress (Miner et al., 2016) and for Wachter-Boettcher (2016) this is a reflection of the political economy of social media, 'an industry willing to invest endless resources in chasing "delight" but not addressing pain'. These criticisms are with the friend-like sensitivity of the app but the bloggers' concerns are about what using such an app in the context of friendship says about the nature of that relationship.

Other bloggers were focused less on friendships than on types of relationships seen as 'unsafe' because they are *not* supportive. These include digital bystanders; those who choose *not* to respond to tweets about distress. Illustrating the complexity of ethics of care in digitally mediated relations, these may be people who do not want to help but they may also include those who tactfully 'disattend' to messages believing they are not intended for them (Brake, 2014: 45). Either way, lack of acknowledgment of emotional distress could leave those who have tweeted about their distress with the impression, as one mental health blogger described it, 'that nobody cares enough to respond' (http://bit.ly/2zEvAOz).

The public space of Twitter can also be inhabited by those who are the antithesis of 'friends' or even 'digital bystanders': those whom some bloggers referred to as 'stalkers' and 'abusers'. As with safe spaces, safe relations are defined by their opposite. Drawing on metaphors which speak to vulnerability and fear of violation, some bloggers suggested that in the 'wrong hands', the app acts like an advertisement for our 'empty homes' or 'lost children'. This speaks to the porousness of safe(r) spaces: the inability to separate out a safe space constituted by well-intentioned use of the app from other spaces, on and offline, that are, as the mental health blogger below notes, potentially less benign:

> What if your stalker was a follower? How would you feel knowing your every 3am mental health crisis tweet was being flagged to people who really don't have your best interests at heart, to put it mildly? (http://bit.ly/2zrS8hv)

Radar had been introduced as a way of listening and being empathic. But not all listeners are benign and, as a result, some bloggers feared that being listened *in on*, rather than listened *to*, might lead some to withdraw from the informal support that exists from simply being on Twitter. For these Twitter users, apps such as Radar, are disruptive of a pre-existing 'positive' community formed in a 'natural, human way'. In describing this sense of community, the metaphors used are system related – Twitter as 'an online ecosystem'. Positioned as 'ad hoc', this relational space is understood by some as being built up through individuals' sharing distress and thus can be thought of as an 'authentic space'. It is a community then in the sense that it involves a fusion of weak and strong ties (Gruzd et al., 2011) and, crucially, has an affective component: sharing emotions constitutes a sense of 'we ness' or 'digital togetherness' (Marino, 2015).

While this imagined space is finely calibrated, constituted by the honesty of what is said, significantly, there is not necessarily an expectation of direct intervention from those listening. This speak to a particular understanding of emotional expression (and listening) in public spaces which is at odds with the understanding that informed the development of the app, outlined in this extract from Samaritans:

> People often tell the world how they feel on social media and we believe the true benefit of talking through your problems is only achieved when someone who cares is listening.[15]

Awareness of the app, however, meant that some bloggers imagined Twitter as less safe and the app as silencing rather than amplifying. This was a view expressed by some of those who were already known to be using Twitter for emotional support, such as the mental health blogger below. Those not publicly using Twitter in this way may, of course, have very different views but they are less likely to be known about:

> It will cause me harm, making me even more self aware about how I present in a public space, and make it difficult for me to engage in relatively safe conversations online for fear that a warning might be triggered to an unknown follower. (http://bit.ly/2AchWiH)

Bloggers discussed the possibility of switching Twitter accounts to 'private' to avoid the algorithm but this recalibration of the imagined Twitter space and its relationships, would make it, for some, more akin to Facebook's relational space(s) with its attendant expectations of distress being shared in particular (undemanding) ways (Buehler, 2017). In the final section, understandings about emotional expression embedded in the above accounts of safe(r) spaces are analysed more closely.

7. Emotions and safe(r) spaces

As Pedersen and Lupton (2016) have noted there is still surprisingly little focus on the role of emotions in the sizeable body of literature on digital sharing and surveillance. In the specific context of sharing emotional distress, Radar was premised on the belief that there is such a thing as a digital cry for help and that it can be identified by algorithm as such. Indeed the app, as noted, was positioned as a second chance to hear such online cries. Radar was intended to create a sense of safety by increasing awareness of those in difficulty and indirectly a sense of obligation to act caringly towards them. Metaphorically, as described by a journalist below, the app was intended to catch people on the precipice:

---

[14] Computerized (or computationally enacted) systems apparently outside our control (Introna, 2016).

[15] http://www.samaritans.org/how-we-can-help-you/supporting-someone-online/samaritans-radar#30oct.

The point of Radar, however, is to catch people on the edge of the cliff face who would be truly grateful of a helping hand to pull them back up, but for whatever reason, didn't feel they could say this outright.[16]

The assumption here is that the meaning of expressed emotion is shared. This is the third understanding of sharing emotion I introduced earlier: that two or more people share an understanding of why an emotion is expressed. In the case of emotional distress, however, those who express emotion and those who then hear or read it are not necessarily sharing the same understanding. Expression of emotional distress can be a cry for help but it can also have and achieve other ends.

A cathartic model of self-expression, for instance, suggests that emotional expression is a form of release (Solomon, 2008). Other perspectives focus on emotional expression as self-actualisation or as a means of making sense of our emotional experiences by sharing them. Bennett (2017) has suggested that there is also a symbolic model for understanding emotional expression: that emotions, like poetry, are an end in themselves - a way of speaking to whom one is (Murthy, 2012). What is missing from some of these psychological and philosophical frames, however, is a strong enough sense of the relationships and contexts within which the sharing of emotions takes place. In their work on academics who share personal information online, for instance, Veletsianos and Stewart (2016) note this group tend to do so with clear intent and a strong awareness of social context.

Some of the Radar-dataset bloggers, such as the mental health blogger below, made it clear that they would not consider it safe for them to express their distress on Twitter in a way that could be interpreted as a cry for help:

No matter how suicidal I've been, I would never, ever tweet something like 'I can't go on' or 'I'm so depressed'. […] I also don't want to cause a fuss, so being aware of specific trigger words that may cause a fuss would make me consciously not use them. Furthermore, I am private about my mental health […] I don't want my health problems to be noticeable. (http://bit.ly/2AchWiH)

This practice of public tweeting with the intent of not being noticeable or 'making a fuss' returns us full circle to the question of what constitutes a safe 'public' online setting, and why people choose to express and listen in such spaces. For some there is safety in being one of many when disclosing pain. This can be thought of as *quiet* public disclosure, constituting and constituted by diffuseness and a dispersed sense of emotional connection. This is as potentially public as those who 'life stream' but is driven less by a desire for visibility per se than, as the mental health blogger below suggests, 'mutual witnessing and display' (Warner, 2002: 13, cited in Haber, 2016: 393). A way, in other words, to 'welcome care' from 'known' networks (Veletsianos and Stewart, 2016)

It's a place I can express myself without worrying that I will upset relatives, where I can say the unsayable because other people understand. (http://bit.ly/2Bcoo8y)

This saying of the unsayable is a way of feeling differently in digital spaces which are often regulated towards optimism (Pedersen and Lupton, 2016) or what Michaelsen (2017) framed as 'a better future of emotional sameness'.

Some bloggers suggested that it was the 'processing' work of the algorithm which made their messages (and hence their emotions), public and potentially unsafe, not the act of tweeting or expressing emotion *in itself*. For these bloggers, the interplay between the material, relational and emotional positions the algorithm as part of an online

interaction order (Mackenzie, 2016). These concerns tune into anxieties about emotional surveillance and about whether or not we can control information about ourselves, including about our emotions (Brandimarte et al., 2013). This brings us to a key distinction between outreach and surveillance: the notion of consent. Becoming aware that one is part of an infrastructure of 'noticing' online that one has not directly consented to can lead to an unease with, and a reaction against, the reading of such information, not least because it may not speak to *who* people believe they 'really' are (Ball et al., 2016). Though others, of course, may position information gained through emotional surveillance as authentic exactly because it happens outside of platform users' full control or even knowledge (Horning, 2016).

8. Conclusion

It is in the nature of safe(r) spaces that they are contestable, open to both boundary maintenance and change. Digital safe(r) spaces can be even more porous and hence boundary skirmishes in relation to them are potentially even more fraught. This article, through exploring the reaction to Radar, has been concerned with what makes an online space feel *emotionally* safe(r).

Those concerned with safeguarding or digital outreach and those concerned with the guarding of safe spaces against surveillance both write about Twitter as a material, emotional and relational space, but they constitute it differently (Lien and Law, 2011). Working with metaphors helps us to understand what and why things matters to people and is revealing of how the Radar app processed and also produced, a great deal of emotion: strong emotions linked to notions of territoriality, belonging and safety. So while there is value in thinking about the online in terms of what we *do* (Jurgenson et al., 2018) rather than as a place or space, all space, as argued earlier, is constituted and for many users of online platforms, doing and space are interconnected. To respond to the sharing of emotions, including cries for help, in a way that is disembedded from and does not acknowledge their spatial context creates upset, and this can be read in the metaphors of intrusion and violation that permeate the blog discussions. Focusing on the material, emotional and relational context of safe spaces, however, also reveals points of connection particularly in relation to valuing peer relationships and the benefits of sharing of emotional distress.

Those concerned with both surveillance and outreach, however, also risk positioning emotional distress as there for the seeing/hearing: depending on one's point of view, ready to be caught in a 'safety net' or turned into 'sensitive data' by the app. The analysis presented here suggests emotions are not so easily read or categorised. For some bloggers, the sharing itself is what keeps the distress at bay; for others, there may well be no obvious emotional content to their sharing and yet they *are* distressed. Even if distress is correctly identified, however, there are still questions about how online proximates and strangers should be cared for. The analysis points to concerns about the automation of support online and also to the claim that some users may leave online public spaces such as Twitter if they believe these spaces to be 'monitored'.

For digital outreach to be enabling rather than undermining of existing peer support, therefore, several issues need to be considered by organisations. These include a need to be explicit about organisational assumptions about what constitutes support and how consent to be supported is sought and given online; being aware of the potential for outreach to be *both* silencing and amplifying and of the sensitivity of the tipping point between keeping a digital eye out and emotional surveillance; and recognising the specific risks of using algorithms in relation to the most human of activities - responding to emotional pain. While only implicitly present in this article, trust runs through all these considerations and is core to what constitutes feeling 'safe(r). Drawing on the meaning of safe(r) spaces repositions the sharing of emotions as being about more than the commercial interests of social media platforms or the norms relating to Twitter, and highlights that people share

---

[16] https://www.theguardian.com/technology/2014/oct/30/suicide-samaritans-radar-will-save-lives-twitter.

distress or emotions because of their sense of trust in relationships they believe they have or desire to have. Any organisation engaging with digital outreach needs to understand how their actions can either shore up or breach such trust, and hence vitally affect people's sense of being safe.

A dualistic focus on outreach *or* surveillance in relation to emotional distress online reduces possible responses to either direct intervention or silence – digital Samaritan or bystander - missing other nuanced practices, including what I have referred to here as quiet public disclosure of emotional pain. Moving beyond dichotomous accounts of online sharing of emotional distress creates more ambiguity but should be an important first step for organisations in working out the fine line between digital caring and surveillance and what it might mean, in both senses, to look out for each other online.

Acknowledgments

References

Andalibi, N., 2017. Sensitive self-disclosures, responses, and social support on instagram: the case of #depression. In: CSCW'17 Proceedings, pp. 1485–1500.

Aguirre, R.T., McCoy, M.K., Roan, M., 2013. Development guidelines from a study of suicide prevention mobile applications (apps). J. Technol. Hum. Serv. 31 (3), 269–293. https://doi.org/10.1080/15228835.2013.814750.

AOIR, 2012. Ethical Decision-making and Internet Research 2.0: Recommendations from the AoIR Ethics Working Committee.

Baker, D., Fortune, S., 2008. Understanding self-harm and suicide websites. Crisis J. Crisis Interv. Suicide Prev. 29 (3), 118–122.

Ball, K., Di Domenico, M., Nunan, D.I., 2016. Big data surveillance and the body-subject. Body Soc. 22 (2), 58–81. https://doi.org/10.1177/1357034x15624973.

Bazarova, N., 2012. Public intimacy: disclosure interpretation and social judgments on facebook. J. Commun. 62 (5), 815–832. https://doi.org/10.1111/j.1460-2466.2012.01664.x.

Bennett, C., 2017. Expressive Needs: Symbols and Self-realisation in Post-Kantian Philosophy (Unpublished manuscript). .

Bernal, P., 2014. Samaritans Radar: Misunderstanding Privacy and 'publicness' (Unpublished manuscript). .

Buehler, E.M., 2017. Seeking emotional support on facebook. Social Media and Society 3 (3).

Bondi, L., 2005. Troubling space, making space, doing space. Group Anal. 38 (1), 137–149.

Brake, D., 2014. Sharing Our Lives Online: Risks and Exposure in Social Media, first ed. Palgrave Macmillan, Basingstoke, Hampshire.

Brandimarte, L., Acquisti, A., Loewenstein, G., 2013. Misplaced confidences: privacy and the control paradox. Social Psychological and Personality Science 4 (3), 340–347. https://doi.org/10.1177/1948550612455931.

Brownlie, J., 2014. Ordinary relationships. A sociological study of emotions, reflexivity and culture. Palgrave MacMillan, Basingstoke.

Bruns, A., Moe, H., 2014. Structural layers of communication on twitter. In: Weller, K. (Ed.), Twitter and Society. Peter Lang, New York, pp. 3–14.

Bucholtz, I., 2013. Diffused Intimacy: trust and self-disclosure in online relationships. Media Stud. 1, 23–39. Retrieved from. https://medialnistudia.files.wordpress.com/2013/06/ms_2013_1_stat2.pdf.

Burke, M., Develin, M., 2016. Once More with Feeling: Supportive Responses to Social Sharing on Facebook CSCW '16. February 27-March 02, 2016, San Francisco, CA. .

Burkitt, I., 2014. Emotions and Social Relations. Sage, Los Angeles.

Cantó-Milà, N., Seebach, S., 2011. Ana's friends. Friendship in online pro-ana communities. Socio. Res. Online 16 (1), 1. http://www.socresonline.org.uk/16/1/1.html.

Christensen, H., 2014. Social media. The new e-mental health tool. Australian Psychological Society Accessed at. https://www.psychology.org.au/Content.aspx?ID=5848.

Crawford, K., 2009. Following you: disciplines of listening in social media. Continuum 23 (4), 525–535.

Crawford, K., 2014. The Anxieties of Big Data. The New Inquiry Retrieved from. http://thenewinquiry.com/essays/the-anxieties-of-big-data/.

Cronin, A.M., 2014. Between friends: making emotions intersubjectively. Emotion, Space and Society 10, 71–78.

Eggertson, L., 2015. Social media embraces suicide prevention. CMAJ: Canadian Medical Association Journal 187 (11), E333. http://dx.doi.org/10.1503/cmaj.109-5104.

Ellaway, R., Coral, J., Topps, D., Topps, M., 2015. Exploring digital professionalism. Med. Teach. 37 (9), 844–849. https://doi.org/10.3109/0142159x.2015.1044956.

Elliot, H., Squire, C., O'Connell, R., 2016. Narratives of normativity and permissible transgression. Forum for Qualitative Research 18 (1) Retrieved from. http://discovery.ucl.ac.uk/1478200/.

Fayard, A.-L., 2012. Space matters, but how. In: Leonardi, P.M., Nardi, B.A., Kallinikos, J. (Eds.), Materiality and Organizing: Social Interaction in a Technological World. Oxford University Press, Oxford.

French, M., Smith, G., 2013. ' Health' surveillance: new modes of monitoring bodies, populations, and polities. Crit. Publ. Health 23 (4), 383–392.

Gruzd, A., Wellman, B., Takhteyev, Y., 2011. Imagining twitter as an imagined community. Am. Behav. Sci. 55 (10), 1294–1318. https://doi.org/10.1177/0002764211409378.

Haber, B., 2016. Queer facebook. Digital sociality and queer theory. In: Daniels, J. (Ed.), Digital Sociologies. Polity Press, UK.

Hess, A., 2015. How Reddit Is Changing Suicide Intervention. Slate Magazine.

Heidegger, M. 1978 [1954]. Building, dwelling, thinking. In: Farell Krell, D. (Ed.), Martin Heidegger: Basic Writings. Routledge, London.

Hillis, K., Paasonen, S., Petit, M., 2015. Networked Affect. MIT Press, Cambridge, Massachusetts.

Hines, C., 2015. Ethnography for the Internet. Bloomsbury, London.

Horning, R., 2016. Mass Authentic. The New Inquiry Retrieved from. http://thenewinquiry.com/blogs/marginal-utility/mass-authentic/.

Illouz, E., 2008. Saving the Modern Soul: Therapy, Emotions, and the Culture of Self-help. University of California Press, Berkeley, CA.

Inkster, B., Stillwell, D., Kosinski, M., Jones, P., 2016. A decade into Facebook: where is psychiatry in the digital age? The Lancet Psychiatry 3 (11), 1087–1090. https://doi.org/10.1016/s2215-0366(16)30041-4.

Introna, L.D., 2016. Algorithms, governance, and governmentality. On governing academic writing. Science Technology and Human Values 41 (1).

Jurgenson, N., 2011. Digital Dualism versus Augmented Reality. Accessed at. https://thesocietypages.org/cyborgology/2011/02/24/digital-dualism-versus-augmented-reality/.

Jurgenson, N., Horning, H., Molokow, A., King, S., 2018. Extremely Online. The Internet Is Not a Place but a Genre. Accessed at. http://reallifemag.com/issue-extremely-online/.

Kelion, L., 2017. Facebook Artificial Intelligence Spots Suicidal Users. Retrieved from. http://www.bbc.co.uk/news/technology-39126027.

Kennedy, J., 2015. Conceptual boundaries of sharing. Inf. Commun. Soc. 19 (4), 461–474. https://doi.org/10.1080/1369118x.2015.1046894.

Lakoff, G., Johnson, M., 2011. Metaphors We Live by, first ed. Univ. of Chicago Press, Chicago [u.a.].

Lee, N., 2014. The lancet technology: trouble on the radar. Lancet 384 Retrieved from. http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(14)62267-4.pdf.

Lefebvre, H., 1991. The Production of Space. Basil Blackwell, Oxford.

Lehdonvirta, V., 2010. Online spaces have material culture: goodbye to digital post-materialism and hello to virtual consumption. Media, Culture & Society 32 (5), 883–889.

Lien, M., Law, J., 2011. 'Emergent aliens': on salmon, nature, and their enactment. Ethnos 76 (1), 65–87. https://doi.org/10.1080/00141844.2010.549946.

Litt, E., Hargittai, E., 2016. The imagined audience on social network sites. Social Media + Society 2 (1). https://doi.org/10.1177/2056305116633482.

Lupton, D., 2012. M-health and health promotion: the digital cyborg and surveillance society. Soc. Theor. Health 10 (3), 229–244.

Lyon, D., 2007. Surveillance Studies: an Overview. Polity, Cambridge.

Mackenzie, D., 2016. How Algorithms Interact: Goffman's 'Interaction Order' in Automated Trading. Retrieved from. http://www.sps.ed.ac.uk/_data/assets/pdf_file/0009/199863/IntOrder27.pdf.

Markham, A., 2013. Metaphors of Internet: Tool, Place, Way of Being (The 2003 Version). Retrieved 9 May 2016, from. http://www.markham.internetinquiry.org/2013/12/metaphorsofinternet/.

Mars, B., Heron, J., Biddle, L., Donovan, J., Holley, R., Piper, M., Gunnel, D., 2015. Exposure to, and searching for, information about suicide and self-harm on the Internet. J. Affect. Disord. 185, 239–245. https://doi.org/10.1016/j.jad.2015.06.001.

Marino, S., 2015. Making space, making place: digital togetherness and the redefinition of migrant identities online. Social Media and Society 1 (2).

Martin, D., 2017. Curating space, choreographing care: the efficacy of the everyday. In: Bates, C. (Ed.), Care and Design: Bodies, Buildings, Cities. Wiley-Blackwell, London.

Marwick, A., boyd, d, 2014. Networked privacy: how teenagers negotiate context in social media. N. Media Soc. 16 (7), 1051–1067. https://doi.org/10.1177/1461444814543995.

Mason, B., 2014. Samaritans Radar Row Won't Be the Last of its Kind. Retrieved from. http://www.theguardian.com/voluntary-sector-network/2014/nov/04/samaritans-radar-app-data-privacy?CMP=twt_gu.

Massey, D., 2005. For Space. Sage Publications, London.

Mead, G.H., 1934. Mind, Self and Society. Chicago University Press, Chicago.

Michaelsen, A., 2017. Feeling differently together: the it Gets Better Project as an unlikely intimate public of LGBT peers. Emotion, Space and Society 25, 144–149.

Miner, A., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., Linos, E., 2016. Smartphone-based conversational agents and responses to questions about mental

health, interpersonal violence, and physical health. JAMA Internal Medicine 176 (5), 619. https://doi.org/10.1001/jamainternmed.2016.0400.

Mok, K., Jorm, A., Pirkis, J., 2016. The perceived impact of suicide-related internet use. Digital Health 2 (0). https://doi.org/10.1177/2055207616629862.

Murthy, D., 2012. Towards a sociological understanding of social media: theorizing twitter. Sociology 46 (6), 1059–1073. https://doi.org/10.1177/0038038511422553.

Pedersen, S., Lupton, D., 2016. 'What are you feeling right now?' communities of maternal feeling on Mumsnet. Emotion, Space And Society 26, 57–63. https://doi.org/10.1016/j.emospa.2016.05.001.

Peyser, A., Eler, A., 2016. Tinderization of Feeling. The New Inquiry Retrieved 9 September 2016, from. http://thenewinquiry.com/essays/tinderization-of-feeling/.

Reeves, J., 2017. Of social networks and suicide nets: biopolitics and the suicide screen. Television and New Media 18.6, 513–528.

Robert, A., Suelves, J., Armayones, M., Ashley, S., 2015. Internet use and suicidal behaviors: internet as a threat or opportunity? Telemedicine And E-Health 21 (4), 306–311.

Rosenfeld, H., Noterman, E., 2014. Safe space: towards a reconceptualization. Antipode 46 (5), 1346–1365. https://doi.org/10.1111/anti.12089.

Sandaunet, A., 2008. A space for Suffering? Communicating breast cancer in an online self-help context. Qual. Health Res. 18 (12), 1631–1641. https://doi.org/10.1177/1049732308327076.

Schuurman, N., 2014. Blogging situated emotions in human–horse relationships.

Emotion, Space and Society 13, 1–8.

Skeggs, B., Yuill, S., 2015. The methodology of a multi-model project examining how facebook infrastructures social relations. Inf. Commun. Soc. 19 (10), 1356–1372.

Solomon, R.C., 2008. True to our feelings what our emotions are really telling us. Oxford University Press, Oxford.

Stengel, B., 2010. The complex case of fear and safe space. Stud. Philos. Educ. 29 (6), 523–540. https://doi.org/10.1007/s11217-010-9198-3.

Thompson, C., 2008. Brave New World of Digital Intimacy. The New York Times September 7, sec. Magazine. http://www.nytimes.com/2008/09/07/magazine/07awareness-t.html.

Tucker, I., Goodings, L., 2017. Digital atmospheres: affective practices of care in ele-friends. Sociol. Health Illness 39 (4), 629–642.

Veletsianos, G., Stewart, B., 2016. Discreet openness: scholars' selective and intentional self-disclosures online. Social Media and Society 1–11.

Wachter-Boettcher, S., 2016. Dear Tech, You Suck at Delight. Retrieved from. https://medium.com/@sara_ann_marie/dear-tech-you-suck-at-delight-86382d101575#.9nl8z93zy.

Warner, M., 2002. Publics and counterpublics. New York: Zone Books. Distributed by MIT Press, Cambridge, Mass.

White, D.S., LeCornu, A., 2011. Visitors and residents: a new typology for online engagement. First Monday 16 (9).