Affective Polarisation and Emotional Distortions on Social Media

Alessandra Tanesini

Social media are, seemingly, a fertile ground for hate speech, misogyny, racism, and abuse. They might also be partially responsible for the emergence of "culture wars" in the UK but especially in the USA.[1] These "wars" would be characterised by hostile, and often hate-filled, disputes over numerous topics including police brutality, critical race theory and Black Lives Matter, feminism, and the rights of transgender women. These disputes are conducted on new and old media by members of sharply polarised groups. It is not the first time, however, that European and North American societies have been riven by deep, seemingly unbreachable, divisions. In the Early Modern period, for instance, Europe was devastated by a series of religious wars. Then like now, disagreements were hostile, full of anger and aggression. People thought of members of the other camp as beyond the pale, and were not afraid to make the contempt in which they held their opponents manifest (cf., Bejan, 2017).

Even though polarisation, hostility, and "culture wars" are not unique to contemporary circumstances, there are aspects of the current situation that make it different from historical episodes of deep social divisions in ways that require novel ameliorative strategies. More specifically, I argue in this paper that social networking sites (hereafter, SNSs), like Facebook, Twitter or Instagram, are technologies whose design features facilitate the triggering and mass contagion of group-based anger; that is, anger experienced by individuals because of perceived slights to their social identity. I argue that some strategies aimed to address the negative consequences of such expressions of anger should be targeted at the design features of these platforms rather than at encouraging users to cultivate virtues such as discreetness (Frost-Arnold, 2021) or care (Desmond, this volume; Vallor, 2016).

This paper consists of five sections. In the first I offer a brief survey of empirical results that strongly suggest that SNSs are essentially emotional environments where strong negative emotions can spread quickly and to many users. In section two, I offer an account of SNSs as emotion technologies that promote a highly charged emotional environment where intrinsic

[1] For example, a recent report by the Policy Institute at King's College London recommends several measures to guard against further polarisation on "culture wars" issues in the UK. These include "holding media and social media to better account for the role they play in this [polarisation] process" and enjoining "political leaders on all sides to cool things down rather than raise the temperature further" (Duffy et al., 2021). The report, however, also suggests that the media's contribution primarily consists in amplifying, and giving unwarranted prominence to, the polarised views of a tiny fraction of the population. That said, this amplification process might turn "culture wars" in the UK into a reality.

emotion regulation is significantly weakened, and people's emotions are more strongly modulated by other people and by the technology itself. I show that these features of social media promote a simplistic emotional outlook which is an obstacle to the development and maintenance of virtue. In section three I explain how SNSs cause deindividuation and promote group-based emotions, including group-based anger. Section four focuses explicitly on the mechanisms that facilitate this affective polarisation. In the final section, after a discussion of the positive value of some forms of anger, I argue that SNSs should not be designed to prohibit or suppress anger, but that its encouragement should also be avoided. I conclude with a suggestion about how this might be achieved.

1. SNSs as Emotional environments

SNSs are online platforms, like Facebook, LinkedIn, WhatsApp, Twitter and Instagram, that are designed to facilitate and promote social relationships. Users construct personalised profiles and establish connections of "friendship" or "followership" with others. [2] They are then able to communicate directly with individuals, via direct messaging, and to broadcast information publicly or to selected groups. Users are also able to view and navigate the contributions made by their connections and sometimes also by others within the network (boyd, 2011). There is now an established body of empirical research in information science, psychology, media studies and social science that strongly indicates that users' engagement with these sites is primarily affect- driven (Löwe & Parkinson, 2014; Papachrissi, 2015). SNSs provide an environment within which messages that communicate emotions spread faster and further than those that do not. Further, the amount of discussion generated by a message is directly proportional to the strength of the emotions it conveys (Chmiel et al., 2011).

In addition to evidence of generalised emotional engagement, several studies have also highlighted phenomena that are akin to emotional contagion (Kramer et al., 2014; Zollo et al., 2015). That is, users tend to experience the emotions conveyed by the messages with which they engage, and to spread these common emotional responses further by expressing them in their comments, shares or retweets. Emotional contagion occurs when a person, influenced by their observation of the emotion expressed by other people, experiences the same emotion as they have witnessed in others.[3] Emotional contagion is therefore a process that gives rise to emotional

[2] Some, perhaps most, of these relationships are transient and superficial. This is true, for instance, of many Facebook "friends". Nevertheless, all social connections on SNSs are relationships of some sort.

[3] There are various accounts of the mechanisms involved in this phenomenon including social appraisal theory according to which others' appraisal of an event as conveyed by their emotions is factored into one's evaluation of the same event playing a role akin to testimony. Hence, the resulting evaluation as expressed in one's emotion is

convergence. That is, to say it leads to the synchronisation of emotions among different people. At times, it also involves emotion regulation. The latter refers to any goal-directed conscious process, technique or strategy that influences which emotion a person has, how, when and for how long they experience it, as well as how (and whether) they give expression to their emotional experience (Gross, 2015, p. 5). Emotional contagion can be an interpersonal or extrinsic form of emotion regulation when one person regulates another's emotion by purposefully sharing with them their emotions with a view to encourage them to feel the same (Gross, 2015, p. 5). For example, a user can post a joyful message on a SNSs with the express purpose of cheering oneself and others up. Such activity is a strategy of emotion regulation that is both intrinsic (directed at oneself) and extrinsic.

Contrary to what one would expect from many dire warnings about the angry and hateful tenor of online communications the overall emotional tone of conversations online is, as a matter of fact, positive. For example, Kramer et al. (2014) found that among the Facebook posts they examined there were twice as many positive emotional messages as negative ones. This finding is what one would expect of platforms designed to multiply social connections. People do not like to engage with those they do not like or to discuss depressing topics. In addition, there are unspoken social norms against posting negative content (Waterloo et al., 2018). When negative content is shared online, it is more frequently communicated as a request for help or an offer of support to close connections by way of direct messaging (Bazarova et al., 2015; Ziegele & Reinecke, 2017). Therefore, positive messages are prevalent in the public channels of SNSs.

It does not follow, however, that anger cannot also be on the increase. The evidence shows that online communication is emotionally charged. It is thus entirely possible that emotions online are more commonly positive than negative, and yet anger and hostility is also on the rise. There is evidence that angry messages on SNSs are especially successful in spreading far and wide (Martin & Vieaux, 2015; Wollebæk et al., 2019). In addition, anger online has been shown to facilitate homophily effects (Song & Xu, 2019). That is, angry people online have a propensity to communicate mostly with those who share their anger, and to avoid those who do not. Importantly, even people who do not interact with each other often or at all in real life are susceptible online to emotional contagion when they view angry content. Joy, instead, is less often shared among strangers online (Fan et al., 2020).

2. SNSs as Emotion Technologies

likely to agree with the appraisals that have informed it. This is a process that would lead to the convergence of emotions (Bruder et al., 2014).

The previous section has highlighted that SNSs are spaces within which connections are emotionally charged, where emotional messages spread fast and wide by processes of emotional contagion leading to emotional convergence. Although the tenor of SNSs communication is prevalently positive, the social media are also a fertile ground for anger-fuelled emotional contagion. In this section I argue that we can make some progress toward understanding these phenomena by thinking of SNSs as emotion technologies. These are artifacts with features designed for the purpose of modulating and regulating emotions (Krueger, 2014; Krueger & Osler, 2019).

Emotion regulation is not a goal of the designers of SNSs. Their goal is to foster social connections.[4] However, even transient and superficial human relationships depend on emotions since the expression of emotions is one of the principal ways of communicating relationship-relevant information and of guiding behaviour. In short, emotions are an essential tool for the management and negotiation of human relations (Löwe & Parkinson, 2014; Parkinson et al., 2005).[5] For example, people show anger to indicate to another agent that their relationship is at risk (Löwe & Parkinson, 2014, p. 130). This emotional expression demands an emotional response such as an acknowledgment of fault that is conveyed in a demonstration of guilt. If this attribution of a communicative function to emotions is correct, then we can explain why humans are strongly motivated to express their emotions, rather than hide them (Goldenberg et al., 2020; Jakobs et al., 2001). We can also understand why we are predisposed to share some emotions when we see others express them (Rimé, 2009).

These considerations suggest that if promoting human connections is an important goal of those who design SNSs, they are successful only if they construct platforms whose features facilitate the communication of emotions. For this reason, the modulation and regulation of emotions is a proximate goal of SNSs design. It is therefore legitimate to think of them as emotion technologies. In the remainder of this section, I flesh out the notion of an emotion technology and focus on some of the design features of SNSs that explain why communication on these platforms is emotionally charged, biased in favour of positive emotions, and why emotional convergence is widespread. I also show that the same features also weaken the effectiveness of traditional strategies of intrinsic emotion regulation and promote the adoption of a simplistic and Manichean emotional outlook. These last two effects of the features of SNSs make them an especially fertile ground for anger. They also make these platforms environments where it is

[4] Their ultimate goals, however, might be commercial. Thanks to Orestis Palermos for reminding me of this fact.

[5] These considerations offer some support for Macnamara's (2015) view that emotions serve a communicative function.

much harder for ordinary human beings, possessing an ordinary amount of virtue, to behave virtuously.

Human beings have always modified their environments to facilitate the performance of various tasks. Some of these tasks are primarily cognitive such as finding the way to a destination or the arithmetical sum of two numbers. To make problems easier to solve, humans have invented strategies (e.g., carrying when doing sums) and artifacts (e.g., maps, the abacus and calculators). These cultural and material resources function as a kind of cognitive niche designed to scaffold human cognition (Sterelny, 2010). But humans have also modified their surroundings to facilitate the task of emotion regulation. We create affective niches, where it is easier to achieve and maintain a desired affective tone (Colombetti & Krueger, 2015; Krueger, 2014). These niches include artifacts such as music or ambient lights designed to modulate our moods and emotions (Krueger, 2019). Thanks to these devices the task of successfully regulating one's emotions is made easier because it is scaffolded by cultural and material technologies.

Given that the goal of SNSs is to foster the increase of social connections, and the fact that in humans social relationships are created and sustained partly by way of expressing emotions, we would expect successful social media to have design features that call for emotional engagement. We would also expect these features to favour emotional contagion so that emotions are disseminated far and wide, and to promote positive emotions since people do not like to interact with disagreeable people. Whilst there are differences between SNSs, many have features that are particularly suited to promote positive emotional engagement and emotional contagion. Three prominent such features are: reaction buttons that encourage selection from a limited range of emotional responses; the availability of alternative channels for direct messaging and public broadcasting; the speed and ease of communication to a large public.

Facebook, for instance, explicitly encourages emotional responses to status updates by providing users with a fixed menu of reaction buttons (like, love, care, amusement, surprise, sadness, and anger).[6] These buttons call for emotional, rather than reflective, reactions and thus contribute to making Facebook an environment where communication is shaped by affect. The same buttons also give prominence to those aspects of a post that make it more likely to be shared because of its novelty or newsworthiness (surprise), its ability to generate positive feelings (like, amusement, love and care) or its capacity to harness support (like, sadness and anger). Hence, this design feature favours emotional contagion. In addition, the buttons are primarily designed to foster positive, rather than negative, emotions as four out of the seven existing options signal pro-social attitudes.

[6] Response buttons also feature in Twitter, Instagram and Tik Tok.

Furthermore, Facebook response buttons are an example of what Alfano et al. (2018) have labelled 'top-down technological seduction'. The platform design invites the readers to unthinkingly accept that the appropriate reaction to a post is one, and only one, among those pre-selected by Facebook. Users are thus taken down a path where there is no place for emotional complexity or ambiguity, and where the range of available emotional responses is rather limited. Users' adoption of Facebook's choice architecture for them has at least four consequences. First, all the available options consist in the expression of an emotion. Second, the available options are limited in number and thus promote emotional convergence. Third, the options are presented as exhaustive and mutually exclusive. Hence, users are led into the adoption of a Manichean emotional outlook where it is impossible for love to be tinged by anger, or vice-versa. It is an outlook that promotes simplicity, clarity, a "with us or against us" mentality; it has no space for ambiguity or complexity. [7] Fourth, the platform gives prominence to anger as one among the standing available responses.

When users are seduced into the adoption of a simplified emotional outlook, they risk losing the ability to appreciate the complexities of human relationships and to respond emotionally to others in the right way. Thus SNSs, because of some of their features, obstruct the development and maintenance of virtues since these involve the capacity to appreciate emotionally the moral features of often complex situations.

Several SNSs also have different channels that separate public broadcasting from direct messaging (Bazarova et al., 2015). This design feature facilitates emotional engagements by creating private channels where one is able to share bad news with intimate friends, whilst maintaining an upbeat persona in the public broadcasting channel. [8] This characteristic of the platforms facilitates both emotional engagement and an overall positive tone in its more public facing channels. The opportunity to broadcast one's message far and wide in the more public channels is a great enabler of emotional convergence on a massive scale as Kramer et al. (2014) detected on Facebook.

Finally, SNSs make communication easy and near instantaneous (Baym & boyd, 2012). Because posting requires little effort, and sharing is even easier, contributing content on SNSs can be done on the impulse of the moment without giving it much thought. The speed of communication ensures that these contributions can reach other people almost instantaneously. In this way exchanges that would take some time face-to-face are accelerated so that they can be conducted over shorter time frames. The compression of the timeframe of conversation

[7] This is an aspect of what Nguyen (2021) has labelled the 'gamification' of communication.

[8] There is evidence for instance that negative emotions are more prominent on WhatsApp than in other more public platforms (e.g., Facebook) (Waterloo et al., 2018).

facilitates emotional engagement since one feels compelled to respond immediately to an immediate response. The speed of communication also partially disables some common strategies of intrinsic emotion regulation. For instance, we try to slow down and "count to ten" to avoid giving expression to a negative emotion, we adopt similar delay strategies to dampen down the intensity of emotions. The speed of on-line communication makes it harder to deploy these strategies successfully. Given the role of intrinsic emotional regulation in promoting continence as the ability to resist urges that one does not endorse, this feature of SNSs constitutes another obstruction to the development of virtue.

The speed of online communication can potentially facilitate the escalation of angry exchanges, where anger is met with anger, that elicits more anger in response (Martin & Vieaux, 2015). It further contributes to the creation of emotional cascades (Alvarez et al., 2015). Whilst anger is generally a short-lived emotion, online communication could prolong its duration as instantaneous responses from others continually re-kindle the initial anger.

The speed and ease of communication on-line in addition to facilitating emotional engagements also encourages massive emotional convergence because it promotes emotional contagion at scale. Previously, mass emotional contagion was only possible when large crowds gathered. Whilst this phenomenon could already take place at concerts held in large stadia, sporting events, religious ceremonies and mass protests, it is potentially turned into an everyday and global occurrence by the speed of online communication. It is now possible for people located in different continents to experience simultaneously the same emotion triggered in part by one's knowledge that others feel in the same way as oneself. These events of mass emotional contagion have the potential to foster collective action across national boundaries and irrespective of location.

To summarise, SNSs are emotion technologies that call for emotional engagements, facilitate pro-social emotions, and promote emotional contagion at scale. Design features responsible for these properties of the platforms include reactions buttons that seduce users to respond emotionally and to adopt a limited emotional palette, the availability of both private and public channels of communication, and the speed and ease of on-line communication. Whilst these features are designed to promote positive emotions, they can also encourage anger by presenting it as a standing possible response, by partially disabling intrinsic emotion regulation, and by compressing the time frame of communicative exchanges.

3. Social identities, Group-based emotions and SNSs

Communication on-line also proceeds under conditions of relative anonymity. Users can, for instance, construct pseudonymous profiles. However, even when people's profiles feature information that identifies them, communication on some SNSs (but not others) mostly proceeds by text, emoticons, and pictures that do not often represent the users themselves. Hence, it is extremely rare to see a person's selfie on Twitter, Reddit, or Telegram. It is, however, more frequent on Facebook and positively ubiquitous on Instagram. When communication proceeds in contexts where the individuality of users is not prominent due, for example, to the absence of visual cues reminding one of the faces of each person with whom one is in dialogue, there is a tendency for users' social identities to become more salient, than their individuality. This phenomenon is known as deindividuation effect (Spears & Postmes, 2015). It results in increased conformity with behaviours that are socially acceptable for members of one's social group. That is, it facilitates acting in accordance with stereotypes.

Deindividuation effects are partially responsible for the prevalence of pro-social emotions online. Since disruptive and negative behaviour is usually frowned upon, conformist users refrain from acting and expressing emotions that are generally disapproved. That said, the relative anonymity of users that makes them less individually identifiable by members of other social groups, also promotes behaviour that, whilst it is judged acceptable by members of one's own social group, is disapproved of by outsiders (Spears & Postmes, 2015). For example, if swearing is accepted by members of one's own social group but disapproved by others, the relative anonymity of the internet promotes an increase in sweary contributions by those whose group approves of swearing. Thus, although anonymity in computer-mediated communication increases conformity to social norms adopted by one's own social group, when norms might vary among groups, it also enables engaging in group-stereotypical behaviour that is disapproved by those outside of one's own group. Hence, anonymity should not be understood as primarily a cause of disinhibition and loss of accountability. To sum up, conformism promoted by anonymity is one of the reasons why the tone of communication on SNSs is usually positive. It can, however, be ugly and antagonistic if such behaviour is, on occasion, stereotypically acceptable for members of a given social group.

Either way, when computer-mediated communication occurs in the absence of prominent visual cues of a person's individuality it promotes self-categorisation as a member of some social group.[9] The increased salience of social group membership has important consequences on the nature of the emotional expressions which, as I have argued above, are ubiquitous online. More

[9] Or at least this is true of social identities that are not visible. It is at least possible that the absence of visual cues dampens, rather than enhances, the salience of identities such as gender or race that are tied to observable characteristics.

specifically, it promotes the experience of so-called group-based emotions. In turn, the prevalence of group-based emotions strengthens the tendency to self-categorise as group member but also to identify more strongly with the values, interests, and commitments of the group (Livingstone et al., 2011).[10]

Group-based emotions are emotions experienced by individuals as members of social groups. For example, the anger experienced by a woman because she is not taken seriously is an instance of group-based anger if it involves an evaluation of someone's actions as a slight that is inflicted upon her because of her gender identity. Hence, as in this example, group-based emotions can be experienced by individuals who are alone. What is distinctive of group-based anger is the evaluation that the perceived slight is inflicted upon one because of one's membership in a group, rather than because of individual characteristics of the person or of her situation (Goldenberg et al., 2020). Group-based emotions are capable of converging by means of emotional contagion. We would expect this phenomenon to be especially prevalent on those SNSs that facilitate emotional engagement whilst depriving users of cues of their individuality.

There is a two-way relationship between self-categorisation and self-identification as group member, and group-based emotions. Prior self-categorisation as member of a social group contributes to how we emotionally appraise situations. Conversely, the experience of a group-based emotion facilitates classifying oneself as members of a group, and investing that categorisation with importance, and thereby identifying oneself with the social group to which one belongs.

Anger, in particular, is triggered when one experiences the actions of some members of a different social group to have slighted the social group to which one belongs (Mackie et al., 2016; Mackie et al., 2004).[11] But the converse is also true. One might come to categorise oneself as a member of a subgroup upon realising that others who belong to the same social group as oneself and share some additional sub-group defining feature experience the same group-based anger as oneself (Livingstone et al., 2011).[12] For example, a person who categorises as a woman and is angry because of some disrespectful behaviour directed at women, upon discovering that other

[10] See Spears (2011) for the distinction between categorising oneself as member of a group and identifying with the group by investing one's membership with significance.

[11] Emotional responses are modulated by the social context. However, Mackie et al. (2016, p. 151) are mistaken in their claim that anger is experienced when the offending out-group is not powerful, and to suggest that fear is the response to a slight by a powerful outgroup. It is perfectly possible to experience both anger and fear at the same time. Being slighted by a powerful outgroup plausibly triggers both anger and fear.

[12] The study was conducted with so called 'minimal groups'. These are made up groups that are created in the experiment. Livingstone et al. (2011) told participants who were all students that they were inductive reasoners. They then told these participants that other inductive reasoners were angry. As a result, angry participants were readier to classify themselves as inductive reasoners.

older women are also angry, if older might also be readier to categorise herself as an older woman than she was prior to knowingly share a group-based emotion with other older women. In addition, group-based anger intensifies identification with salient social groups. That is, individuals who experience group-based anger because of the actions of members of outgroups, invest their own group membership with more significance so that it becomes a more important part of who they are (Kessler & Hollbach, 2005). But group-based emotions do not just promote self-categorisation and identification but also preparedness for collective activity (Livingstone et al., 2011; van Zomeren et al., 2004). In short, experiencing group-based emotions leads people to invest their group membership with more meaning, and thus to become more committed to act in defence of the interests of their group.

4. Hostile Identities, SNSs, and Anger

The two-way relationship between social identification and group-based emotions might be partially responsible for some forms of affective polarisation, where members of different social groups strongly dislike or even hate each other, irrespective of whether their disagreement are genuinely substantive (Hannon, 2021).[13] Group-based emotions such as anger are facilitated by the emotional tenor of SNSs and by deindividuation effects. These emotions in turn intensify the significance with which one invests one's social identity. The enhanced salience of social identity further promotes the experience of even more intense group-based emotions (Mackie et al., 2004). In addition, witnessing the group-based emotions of those with whom one identifies, especially when these emotions are intense, informs one's emotional appraisal of the situation leading to emotional contagion among those who self-categorise, and identify, as members of the same social groups. We should thus expect online environments to be places where some forms of emotional contagion spread among members of a social group but not across different social groups.

In the previous section I have argued that SNSs promote the adoption in users of a simplistic and Manichean emotional outlook where it is not possible for anger to be tinged by love or by sadness. This combines with those features that amplify social identification online and the intense group-based emotions that are often associated with it. We should expect that individuals caught in this dynamic to develop strong emotions, but we should also expect these emotions to be global emotions like hate or contempt.

[13] Hence, affective polarisation is rather different from polarisation about belief which, in one of its many senses, occurs when people faced with the same evidence come to hold opposing views. The different mechanisms involved in these cases are discussed by Talisse (2019) and (Shackel) (this volume).

Anger is not a global emotion. One may be angry at someone for something that they have done, and at the same time care for them because of something bad that has happened to them. Thus, anger has both a target (the person or persons with whom one is angry), and a focus (their action or feature that one appraises as deserving to be met with anger). One can experience at the same time another emotion directed at the same target but with a different focus. Contempt is instead a global emotion because its focus is the whole person who is the target of the emotion (Bell, 2013). This emotion signals that the relationship is beyond repair. Because of its global character, it is not possible to experience both contempt and some positive emotion toward the same target at the same time.

The simplistic emotional outlook promoted by SNSs is one in which one cannot easily experience more than one emotion at same time about the same person. The consequent atrophy of emotional nuance, especially in the context of communication with virtual strangers, would seem to promote the experience of emotions that are global, such as contempt, or at least both extreme and without qualification. In short, the simplification of users' emotional palette, when combined with emotional contagion and strong social identification, is, I suspect, one important cause of affective polarisation on SNSs.

The argument so far has indicated that SNSs have features that promote emotional engagement, emotional contagion and increased social identification. I have noted that the overall emotional tenor of social media is positive but that there are several reasons why negative emotions, and especially anger, also thrive. First, anger is at least on Facebook one of the pre-selected standing reactions. Second, SNSs speed up communication making it hard to deploy successfully some forms of intrinsic emotion regulation standardly adopted to inhibit the expression of anger. Third, SNSs facilitate emotional contagion on a massive scale and thus the transmission and prolongation of anger once it emerges. Fourth, the relative anonymity of communication on some SNSs promotes social identification that facilitates group-based emotions. In turn, the experience of these emotions increases social identification which promotes the experience of even more intense group-based experiences. This phenomenon contributes to the segmentation of users into social groupings each of which is subject to emotional contagion. When this feature of SNSs communication is combined with its promotion of a Manichean emotional outlook, the separation of users into strongly emotive social groups provides fertile ground for affective polarisation where member of differing social groups develop a strong dislike for each other. Finally, the relative anonymity of on-line communication also facilitates stereotypical behaviour including behaviour that is frowned upon by society at large but is tacitly approved by members of one's given social group. In contemporary Western societies angry and hostile behaviour toward women is a stereotypically acceptable expression of some forms of masculinity.[14] It is thus not a surprise that misogynistic angry messages proliferate online.[15]

There are two further design features of SNSs that play a significant role in making these platforms fertile grounds for the expression of anger: algorithm-driven personalisation and public broadcasting that is responsible for context collapse. The first of these two features has been the topic of intense study since it is usually singled out as among the most significant causes of filter bubbles online (Pariser, 2011). SNSs, such as Facebook, have proprietary algorithms which, based on a user's track record of responses such as likes, clicks and other engagements, select which posts appear in that user's news feed. Personalisation is described by Alfano et al. (2018) as a 'bottom-up' technological seduction. The technology learns from the user's past behaviour to serve them more of what they have previously engaged with. In this manner users are segmented into niches of like-minded individuals that primarily interact only with each other. This aspect of personalisation exacerbates homophily as the propensity to interact only with those with whom we agree that is a common human tendency but one that is made worse by anger (Song & Xu, 2019). Personalisation would thus make it more likely that users come across the angry posts of others with whose emotional appraisal they are likely to agree. That is, SNSs give access to content that these users would not otherwise have sight of, and which is likely to be anger-triggering for them, because it has been anger-triggering for others who are like them.

The second design feature of SNSs is their inclusion of channels for public broadcasting. I have mentioned this feature above when I contrasted it with direct messaging that allowed for more private expressions of negative emotions. There I highlighted the role of public broadcasting in enabling emotional contagion. Here I focus on another aspect of this design feature: context collapse (boyd, 2011). When users broadcast content using one of the SNSs public channels, their content potentially reaches multiple audiences and makes it impossible for one to tailor one's message to a specific audience. In face-to-face contexts, we would not convey the same information to close friends, mere acquaintances, parents and work colleagues.[16]

Communications on SNSs facilitate the mashing up, or collapse, of these social contexts which we might wish to keep separate. One of the effects of this loss of the ability to tune one's

[14] Such behaviour is a primary manifestation of misogyny understood as hostility directed at women.

[15] Overall, in the US at least, men are more likely to experience online threats and name-calling, but the harassment to which women are subjected tends to be more severe and have deeper effects (Pew Research Center, 2017). A study commissioned by Amnesty International revealed that online harassment has lasting impact on women in numerous countries (Amnesty International, 2017).

[16] Desmond (this volume) might be thinking of the same phenomenon when he argues that messages broadcast on SNSs retain the feel of communications that are attuned to an audience despite being publicly communicated. He points out that different norms of trust govern private and public conversations, and that users online are especially vulnerable to having their trust betrayed.

message to one's audience is the increased likelihood of misunderstandings since messages reach people who do not know the messenger well (Frost-Arnold, 2021). Some of these misunderstandings might easily trigger angry responses.

Further, context collapse increases the risk of conscious or inadvertent violation of others' privacy. Partly because of its still relative novelty, users have not developed clear and firm conventions about sharing content and tagging pictures. Promiscuous sharing and tagging can make it easier for people to be targeted by malicious users, or have their content communicated to people whom they might have legitimately wished to keep in the dark (Frost-Arnold, 2021). In short, context collapse creates conditions that favour behaviour that might do speakers an injustice by violating their privacy and making them more vulnerable to harm. Since anger is the common response to these actions, context collapse creates conditions where anger triggering events are more likely.

It should now be clear that the character and distribution of anger on SNSs has many causes. It is inadvertently facilitated by many of those features of these platforms that are intended to facilitate social connections.[17] It is also parasitic on social norms that predate the SNSs and make some stereotypical aggressive behaviours acceptable for members of some social groups. Anger online therefore is partly continuous with the kinds of hostility, hate and social divisions that have often characterised Western societies, but it also exhibits novel features such as its ability to spread fast and at a massive scale, its disablement of intrinsic emotion regulation, and its formation because of top-down technological pre-selection. For these reasons, I suspect, people who are reasonably thoughtful in their face-to-face encounters are more easily seduced into becoming on-line angry bullies.

5. Amelioration

My discussion has so far consciously avoided questions about the value of anger. In this final section, I address this issue before making a brief ameliorative suggestion to address some of the more problematic expressions of anger on-line.

It might be thought that anger and affective polarisation are always bad for democratic communities since angry and polarised individuals are unlikely to try to compromise with those they oppose. This conclusion is unwarranted. Perhaps, one should welcome the fact that SNSs have facilitated the expression of anger, at least in some cases. To see this, one first needs to realise that not all anger is the same. Sometimes anger is a fitting and proportionate response to

[17] That said, a whistle-blower has recently alleged that Facebook has been aware of the divisive effects of some of these features and chosen to do nothing because anger keeps users engaged, and engagement brings advertising revenue (Whitwam, 2021).

slights and wrongs. On other occasions anger is not apt or fitting because it is a reaction that falsely appraises as a wrong or slight some action when it is not. In addition, showing fitting anger might at times be justified.[18] It might for instance be required by self-respect (Srinivasan, 2018). It might also be necessary to create an effective political community in the fight against grave social injustices (Cherry, 2021; Lepoutre, 2018). Hence, although anger can sometimes be misplaced, there are situations in which it is righteous and virtuous.

Perhaps, then, anger on social media should be welcome. The SNSs have offered an opportunity to the less powerful to voice their anger, and to resist oppression. SNSs' promotion of group-based anger and outrage might also have played some role in helping opponents of tyrannical regimes to create communities capable of collective action. These formations have not typically been very effective in achieving their goals (Tufekci, 2017). Nevertheless, the group-based emotions of members of these communities of resistance were often virtuous. Hence, one might wish to endorse the roles of SNSs in facilitating the expression and diffusion of anger as a way of giving a voice to the powerless and enabling the fight against injustice.

That said, there are expressions of anger online that are clearly not fitting and are morally unjustifiable. Some of these, such as those involving misogynistic anger and hate directed at women online, are continuous with behaviour that can occur in face-to-face situations. These expressions are however exacerbated by the relative anonymity of some social media that promotes increased conformism to stereotypical behaviour. It is also true that SNSs facilitate misplaced anger because of its promotion of a Manichean emotional outlook. Hence, there is little doubt that this increase of anger on SNSs needs addressing. Whatever remedies one proposes, however, one should avoid having a silencing effect on the virtuous anger of those who struggle against inequality. One should also not lose sight of the fact that SNSs weaken intrinsic emotion regulation and thus obstruct a strategy frequently used to achieve continence. These platforms are somewhat addictive (Vaidhyanathan, 2018, chs. 1-2). They also have features seduce us into adopting simplistic and distorted emotional outlooks. These considerations suggest that interventions aimed at individuals' characters requiring them to become more responsible, while valuable, might be of limited efficacy, and must be supplemented with measures that target the design features of the platforms themselves. These measures are likely to make SNSs less successful at establishing and maintaining social connections, and thus ultimately at generating advertising revenue. They are thus unlikely to be implemented without external intervention in the shape of regulation.

[18] See D'Arms and Jacobsen (2000) for the distinction between fittingness as accurate evaluation conveyed by the emotion and justification as moral propriety of experiencing the emotion.

Nevertheless, some small interventions could, for instance, have a positive impact on Facebook's tendency to promote a simplistic emotional outlook without making it less successful as a platform that promotes social connections. I have already argued above that Facebook's reaction buttons are a form of top-down technological seduction that erodes users' ability to experience emotional complexity and ambiguity. Their introduction in 2016 is in my view a retrograde step. To understand why, one needs to be clear about the functions served by the original "like" button and the reasons why users asked for more options.

In face-to-face encounters there is some conversational pressure not to ignore speakers when they address us.[19] In ordinary circumstances it is simply rude not to nod or respond when someone attempts to engage us in conversation. Some of that pressure survives, even though in a reduced form, on SNSs where users seek others' acknowledgement of their posts, whilst their friends feel a sense of obligation to respond especially if the content of the post is important. However, given the vast number of posts users shift through, they need a way to signal quickly that they have paid attention. Facebook's "like" button served this function (Sumner et al., 2017). It would have served it better, if the term "like" had not also potentially conveyed an endorsement of the content of the post. It is this implication that made its use awkward when the original post conveyed bad news or described an injustice. For this reason, users asked for a broader range of pre-selected emotional responses. But, as I argued above, this modification promotes emotional outlooks that should be resisted because it makes users less able to experience complex and ambiguous emotions even when they are the responses that fit the situation. This problem can be avoided if the reaction buttons are eliminated in favour of a single differently named button whose exclusive function is to convey that one has paid attention. It would be the Facebook equivalent of a head nod.[20]

Admittedly, my practical proposal concerns only one platform. However, I submit that other similar suggestions can be developed each tailored to the different features of SNSs. These engineering solutions would not on their own solve the problems of affective polarisation, inappropriate anger, and hostility online. They would, however, succeed in transforming social media platforms into environments that are less hostile to virtuous communication, including virtuously angry messages.[21]

[19] On this underexplored aspect of the norms of conversation see part one of Goldberg's (2020).

[20] This approach is to be preferred to the proliferation of response buttons that are not considered as mutually exclusive. Since reactions to posts need to be quick, the number of options for users must be limited to avoid trawling through endless possibilities. Hence, even if a broader range of emotions are enabled by additional buttons, users are still forced into the adoption of a limited palette of emotional responses.

[21] Thanks to Jon Webber and Orestis Palermos for their helpful comments.

References:

Alfano, M., Carter, J. A., & Cheong, M. (2018). Technological Seduction and Self-Radicalization. *Journal of the American Philosophical Association, 4*(3), 298-322. doi:10.1017/apa.2018.27.

Alvarez, R., Garcia, D., Moreno, Y., & Schweitzer, F. (2015). Sentiment cascades in the 15M movement. *EPJ Data Science, 4*(1). doi:10.1140/epjds/s13688-015-0042-4.

Amnesty International. (2017, 20 November). Amnesty reveals alarming impact of online abuse against women. Retrieved from https://www.amnesty.org/en/latest/news/2017/11/amnesty-reveals-alarming-impact-of-online-abuse-against-women/

Baym, N. K., & boyd, d. (2012). Socially Mediated Publicness: An Introduction. *Journal of Broadcasting & Electronic Media, 56*(3), 320-329. doi:10.1080/08838151.2012.705200.

Bazarova, N. N., Choi, Y. H., Schwanda Sosik, V., Cosley, D., & Whitlock, J. (2015). *Social Sharing of Emotions on Facebook*. Paper presented at the Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15, Vancouver, BC.

Bejan, T. M. (2017). *Mere civility: disagreement and the limits of toleration*. Cambridge, Massachusetts: Harvard University Press.

Bell, M. (2013). *Hard Feelings: The Moral Psychology of Contempt*. New York: Oxford University Press.

boyd, d. (2011). Social Network Sites as Networked Publics: Affordances, Dynamics, and Implication. In Z. Papacharissi (Ed.), *A networked self: identity, community and culture on social network sites* (pp. 39-58). New York: Routledge.

Bruder, M., Fischer, A., & Manstead, A. S. R. (2014). Social appraisal as a cause of collective emotions. In C. v. Scheve & M. Salmella (Eds.), *Collective emotions: perspectives from psychology, philosophy, and sociology* (First edition. ed., pp. 142-155). Oxford: Oxford University Press.

Cherry, M. (2021). *The Case for Rage: Why Anger Is Essential to Anti-Racist Struggle*. New York: Oxford University Press.

Chmiel, A., Sienkiewicz, J., Thelwall, M., Paltoglou, G., Buckley, K., Kappas, A., & Holyst, J. A. (2011). Collective emotions online and their influence on community life. *PLoS ONE, 6*(7), e22207. doi:10.1371/journal.pone.0022207.

Colombetti, G., & Krueger, J. (2015). Scaffoldings of the affective mind. *Philosophical Psychology, 28*(8), 1157-1176. doi:10.1080/09515089.2014.976334.

D'Arms, J., & Jacobsen, D. (2000). The Moralistic Fallacy: On the 'Appropriateness' of Emotions. *Philosophy and Phenomenological Research, 61*(1), 65-90.

Desmond, H. (2022). Reclaiming Privacy and Care in the Age of Social Media.

Duffy, B., Hewlett, K., Murkin, G., Benson, R., Hesketh, R., Page, B., . . . Gottfried, G. (2021). *"Culture wars" in the UK*. Retrieved from https://www.kcl.ac.uk/policy-institute/assets/culture-wars-in-the-uk.pdf

Fan, R., Xu, K., & Zhao, J. (2020). Weak ties strengthen anger contagion in social media. *arXiv:2005.01924 [cs.SI]*. Retrieved from https://arxiv.org/abs/2005.01924

Frost-Arnold, K. (2021). The Epistemic Dangers of Context Collapse Online. In J. Lackey (Ed.), *Applied epistemology* (pp. 437-456). New York: Oxford University Press.

Goldberg, S. C. (2020). *Conversational Pressure*. Oxford: Oxford University Press.

Goldenberg, A., Garcia, D., Halperin, E., & Gross, J. J. (2020). Collective Emotions. *Current Directions in Psychological Science, 29*(2), 154-160. doi:10.1177/0963721420901574.

Gross, J. J. (2015). Emotion Regulation: Current Status and Future Prospects. *Psychological Inquiry, 26*(1), 1-26. doi:10.1080/1047840x.2014.940781.

Hannon, M. (2021). Political Disagreement or Partisan Badmouthing? The Role of Expressive Discourse in Politics. In E. Edenberg & M. Hannon (Eds.), *Political Epistemology*.

Jakobs, E., Manstead, A. S. R., & Fischer, A. H. (2001). Social context effects on facial activity in a negative emotional setting. *Emotion, 1*(1), 51-69. doi:10.1037//1528-3542.1.1.51.

Kessler, T., & Hollbach, S. (2005). Group-based emotions as determinants of ingroup identification. *Journal of Experimental Social Psychology, 41*(6), 677-685. doi:10.1016/j.jesp.2005.01.001.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America, 111*(29), 10779. doi:10.1073/pnas.1412469111.

Krueger, J. (2014). Emotions and the Social Niche. In C. v. Scheve & M. Salmella (Eds.), *Collective emotions: perspectives from psychology, philosophy, and sociology* (pp. 156-171). Oxford: Oxford University Press.

Krueger, J. (2019). Music as affective scaffolding. In R. Herbert, D. Clarke, & E. Clarke (Eds.), *Music and Consciousness 2: Worlds, Practices, Modalities* (pp. 55-70). Oxford: Oxford University Press.

Krueger, J., & Osler, L. (2019). Engineering affect: emotion regulation, the internet, and the techno-social niche. *Philosophical Topics, 47*(2), 205-232. doi:10.2307/26948114.

Lepoutre, M. (2018). Rage inside the machine. *Politics, Philosophy & Economics, 17*(4), 398-426. doi:10.1177/1470594X18764613.

Livingstone, A. G., Spears, R., Manstead, A. S. R., Bruder, M., & Shepherd, L. (2011). We feel, therefore we are: emotion as a basis for self-categorization and social action. *Emotion, 11*(4), 754-767. doi:10.1037/a0023223.

Löwe, I. v. d., & Parkinson, B. (2014). Relational emotions and social networks. In C. v. Scheve & M. Salmella (Eds.), *Collective emotions: perspectives from psychology, philosophy, and sociology* (First edition. ed., pp. 125-140). Oxford: Oxford University Press.

Mackie, D. M., Maitner, A. T., & Smith, E. R. (2016). Intergroup emotions theory. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination, 2nd ed.* (pp. 149-174). New York, NY, US: Psychology Press.

Mackie, D. M., Silver, L. A., & Smith, E. R. (2004). Intergroup Emotions: Emotion as an Intergroup Phenomenon. In L. Z. Tiedens & C. W. Leach (Eds.), *The social life of emotions* (pp. 227-245). New York, NY, US: Cambridge University Press.

Macnamara, C. (2015). Reactive Attitudes as Communicative Entities. *Philosophy and Phenomenological Research, 90*(3), 546-569. doi:10.1111/phpr.12075.

Martin, R. C., & Vieaux, L. E. (2015). 9. The Digital Rage: How Anger is Expressed Online. In G. Riva, B. K. Wiederhold, & P. Cipresso (Eds.), *The psychology of social networking* (Vol. 2, pp. 117-127). Berlin and Boston: De Gruyter.

Nguyen, C. T. (2021). How Twitter gamifies communication. In J. Lackey (Ed.), *Applied Epistemology*. New York: Oxford University Press.

Papachrissi, Z. (2015). *Affective Publics: Sentiment, Technology and Politics*. Oxford: Oxford University Press.

Pariser, E. (2011). *The filter bubble: what the Internet is hiding from you*. New York: Penguin Press.

Parkinson, B., Fischer, A. H., & Manstead, A. S. R. (2005). *Emotion in Social Relations: Cultural, Group, and Interpersonal Processes*. New York and Hove: Psychology Press.

Pew Research Center. (2017). *Online Harassment.* Retrieved from Washington, D.C.: https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/

Rimé, B. (2009). Emotion Elicits the Social Sharing of Emotion: Theory and Empirical Review. *Emotion Review, 1*(1), 60-85. doi:10.1177/1754073908097189.

Shackel, N. (2022). Uncertainty Phobia and Epistemic Forbearance in a Pandemic. *Philosophy.*

Song, Y., & Xu, R. (2019). Affective Ties That Bind: Investigating the Affordances of Social Networking Sites for Commemoration of Traumatic Events. *Social Science Computer Review, 37*(3), 333-354. doi:10.1177/0894439318770960.

Spears, R. (2011). Group Identities: The Social Identity Perspective. In S. J. Schwartz, K. Luyckx, & V. L. Vignoles (Eds.), *Handbook of identity theory and research*. New York: Springer Science+Business Media.

Spears, R., & Postmes, T. (2015). Group Identity, Social Influence, and Collective Action Online: Extensions and Applications of the SIDE Model. In S. S. Sundar (Ed.), *The Handbook of the Psychology of Communication Technology* (pp. 23-46). Malden and Oxford: Wiley Blackwell.

Srinivasan, A. (2018). The Aptness of Anger. *Journal of Political Philosophy, 26*(2), 123-144. doi:10.1111/jopp.12130.

Sterelny, K. (2010). Minds: extended or scaffolded? *Phenomenology and the Cognitive Sciences, 9*(4), 465-481. doi:10.1007/s11097-010-9174-y.

Sumner, E. M., Ruge-Jones, L., & Alcorn, D. (2017). A functional approach to the Facebook Like button: An exploration of meaning, interpersonal functionality, and potential alternative response buttons. *New Media & Society, 20*(4), 1451-1469. doi:10.1177/1461444817697917.

Talisse, R. B. (2019). The Problem of Polarization. In *Overdoing Democracy* (pp. 95-128). New York: Oxford University Press.

Tufekci, Z. (2017). *Twitter and tear gas: the power and fragility of networked protest*. New Haven ; London: Yale University Press.

Vaidhyanathan, S. (2018). *Antisocial media: how facebook disconnects US and undermines democracy*. New York: Oxford University Press.

Vallor, S. (2016). *Technologies and the Virtues: A Philosophical Guide to a Future Worth Wanting*. New York: Oxford University Press.

van Zomeren, M., Spears, R., Fischer, A. H., & Leach, C. W. (2004). Put your money where your mouth is! Explaining collective action tendencies through group-based anger and group efficacy. *Journal of Personality and Social Psychology, 87*(5), 649-664. doi:10.1037/0022-3514.87.5.649.

Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2018). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media & Society, 20*(5), 1813-1831. doi:10.1177/1461444817707349.

Whitwam, R. (2021). Whistleblower: Facebook Is Designed to Make You Angry. Retrieved from https://www.extremetech.com/internet/327855-whistleblower-facebook-is-designed-to-make-you-angry

Wollebæk, D., Karlsen, R., Steen-Johnsen, K., & Enjolras, B. (2019). Anger, Fear, and Echo Chambers: The Emotional Basis for Online Behavior. *Social Media + Society, 5*(2), 1-14. doi:10.1177/2056305119829859.

Ziegele, M., & Reinecke, L. (2017). No place for negative emotions? The effects of message valence, communication channel, and social distance on users' willingness to respond to SNS status updates. *Computers in Human Behavior, 75*, 704-713. doi:10.1016/j.chb.2017.06.016.

Zollo, F., Novak, P. K., Del Vicario, M., Bessi, A., Mozetic, I., Scala, A., . . . Quattrociocchi, W. (2015). Emotional Dynamics in the Age of Misinformation. *PLoS ONE, 10*(9), e0138740. doi:10.1371/journal.pone.0138740.