Avoiding conflict and minimising exposure:

Face-work on Twitter

Michael James Walsh[1]
University of Canberra, Canberra, Australia

Stephanie Alice Baker
City University of London, London, UK

Abstract

In a context where Twitter has come under criticism for enabling and encouraging hostile communication, this article explores how users adopt a 'Twitter face' when navigating interactions on the platform. Extending Goffman's observation that face-work is applicable to both immediate and mediated interaction, this article provides a novel application of face-work on Twitter. Reporting on data from an online questionnaire completed by general Twitter users, we explore how uncivil interaction is experienced on the platform and the interaction strategies users employ to protect their face. We examine how interactions on the platform can lead to a ritual break-down that generate forms of alienation arising from aggressive uses of face-work. We contend that attempts to enhance Twitter as a medium by limiting and restricting particular interactions are ultimately attempts at shaping Twitter's affordances. In analysing user experience, our discussion considers how incivility is responded to and how the platform encourages users to engage in the avoidance components of facework, while simultaneously inhibiting the easy adoption of its restorative dimensions. While both dimensions of face-work are vital, the downplaying of restorative aspects of face-work arguably undermines Twitter's efforts to encourage inclusive interactions across the platform.

Keywords

Face-work, Goffman, health, incivility, interaction, social media, Twitter

Introduction

Twitter has become a leading social media technology. The company promotes itself as an inclusive space to engage in a global conversation comprised of a diversity of voices: 'people around the world come to Twitter to share ideas and have conversations' (Twitter 2019a). Twitter is a micro-blogging site that contrasts with other online social networks, in that it affords brief text-based communication (messages comprised of 280 characters or less) and social connectedness through sharing content with open social networks (Murthy 2013:8). Unlike Facebook, Twitter's design involves no 'technical requirement' or 'social expectation of reciprocity' (Marwick and boyd 2011:142). The platform's default mode of public expression has implications for how people interact. The platform mandates a concise character limit, which encourages brief, impulsive communication and exchange (Jaidka et al. 2019). Twitter has faced increasing criticism regarding the rise of hostility and abuse users encounter on the platform (Lawson 2018; Matamoros-Fernández 2017). In response to these concerns, a 'healthy conversations' strategy was launched by the company seeking to investigate and 'measure' conversational health on Twitter (Twitter 2018a, 2018b). Given Twitter's attempts at curtailing incivility, in this article we draw on Goffman's theory of face-work to examine the nature of hostile interaction on the platform. We suggest that users interacting on the site adopt a type of 'Twitter face': a mode of face-work navigating the interaction order of Twitter. As Goffman (1963:28) contends, we 'have party faces, funeral faces, and various kinds of institutional faces' that respond and adapt to the interaction requirements of different situations.

---

[1] Corresponding author: Michael James Walsh, Faculty of Business, Government & Law, University of Canberra, Building 11, Kirinari St, Bruce, Canberra, ACT 2601, Australia.

We find users respond to hostile interactions on Twitter by using a series of strategies including partitioning their activities through multiple accounts, using protected tweets, blocking, steering clear of certain topics and users who may represent threats to face or by partially engaging with others and self-censoring. While Goffman's theory of face-work was established prior to the advent of social media, we contend his ideas provide researchers with 'a set of tools which can be developed to understand social media, including Twitter' (Murthy 2012:1066). Extending Goffman's theory of face-work, we examine how hostile interaction is managed by users and propose it represents an adaptation of the conventions concerning face-work. In order to demonstrate the role face-work plays on Twitter, we organise the paper as follows. First, we frame Twitter as a social media technology in the broader context of navigating unpleasant user experience online. Next, we discuss Goffman's notion of face-work and how his theory has been drawn on to date, before outlining the research methodology and presenting an analysis of user responses to hostile interactions by drawing on an online questionnaire. Our contribution extends ritualistic understandings of incivility on Twitter, highlighting the role face-work plays in managing online interactions.

Twitter, we have a problem: navigating unpleasant user experiences

Towards the end of 2017, Twitter was under considerable public pressure to improve content moderation on the site due to the prevalence of abuse encountered by users and allegations of inaction from the company. Along with Twitter's growth and scale, various controversies regarding a culture of incivility and abuse emerged. A prominent example included the case of US actress and comedian, Leslie Jones, whose experiences of targeted abuse on Twitter, brought public attention to the way Twitter elided responsibilities around user safety, while exposing how digital platforms 'and celebrity culture become spaces for ideological battles over race, sexuality, and gender' (Lawson 2018:819). There were also signs that ordinary Twitter users were subject to abuse on the platform. In 2017, one study suggested that four-in-ten adults from the US had personally experienced harassment or abusive behavior online (Pew 2017). Perhaps, due to the affordances of the platform, Twitter users tend to be less inhibited in their speech and have less space for deliberation, which encourages more uncivil interaction than other popular social networks (Oz et al. 2018:3414). In addition, trolling activities continue to be present on the site, with trolls understood as internet users who employ antisocial behavior to provoke and irritate other users, often using aggressive or profane language to derail the typical evolution of online discussions (Fornacciari et al. 2018:258). Trolling in this respect goes beyond a singular intimidating act directed at a user, that carries another level of aggression, seeking to disrupt more accommodating interactive practices (Groshek and Cutino 2016:3). This can be seen as a consequence of the deindividuation synonymous with online discussions in addition to the deficit in conversational cues and the speed with which a comment can go viral, fomenting a perfect environment for incivility to proliferate (Chen 2017:64).

As a consequence of growing public criticism of the platform's role in the spread of misinformation and abuse, Twitter was on the precipice of substantial reputational damage and risked driving users and advertisers from the platform. In response to claims the site facilitated problematic and abusive user behaviour (Pew 2018b), Twitter launched a 'healthy conversation' strategy aimed to address and investigate abusive encounters and the general coarsening of interactions on the platform (Twitter 2018a, 2018b). The initiative sought to improve users' experience, by encouraging constructive conversations through the adoption of health-based metrics. These new metrics regarding the 'health' of interactions on Twitter were to be used to assess the status of conversation on the platform and feed into future improvement and iterations of Twitter's user interface into the medium to longer term. However, these corporate responses that seek to restore the civility of Twitter interactions in some respects fail to leverage the practices Twitter users engage in managing hostile interactions presently.

While research has been conducted in relation to the use of Twitter as an instrument for tracking health-related information online (Lee et al. 2014; Murthy 2013), this article explores how users navigate this platform through a consideration of face-work, examining the strategies adopted by users to manage incivility and how these are afforded by Twitter's platform design. Affordances describe the ways in which a platform is engineered to afford and constrain certain behaviors and interactions (Davis and Chouinard 2016:241). While Twitter's affordances enable and encourage users to engage in the avoidance components of face-work, we contend that the platform simultaneously inhibits the restorative dimensions of face-work. Both dimensions of face-work are vital. However, because one element is downplayed, this undermines Twitter's efforts to encourage inclusive practices across the platform. In analysing user experience, our discussion considers the strategies users employ to respond to incivility and how attempts at engineering the platform for conversational health would benefit from understanding the practices Twitter users currently adopt to navigate hostile interactions.

Understanding the role of face-work, incivility and affordances

Defining face-work

'Face' was originally defined by Goffman (1955) as the verbal and non-verbal behaviours allowing for the establishment of 'a line' of contact between two persons. Through this line of contact, the individual provides an assessment of the unfolding interaction in which they are situated, including how they perceive themselves in living up to the line—or projection of self—they have adopted (Goffman 1967:5). 'Face-work', therefore, is the way individuals shape their behaviour in response to contact with others to fashion the impression of self, projected during an interaction. When we have or maintain face, the line adopted is one where the projected image of self remains consistent and is 'supported by judgements and evidence conveyed by other participants, and that is confirmed by evidence conveyed through impersonal agencies in the situation' (Goffman 1967:6-7). As Collins (2008:339) explains, 'Individuals try to maintain their own face, their claim to be what they put themselves forward as being (at least in that situation), and they help others do the same'.

Alternatively, when one loses face, what occurs is a faltering in the interactional equilibrium; communication is conveyed with the intent or effect of damaging the performance of an interlocutor through a 'crack' or some such evidence countering the prevailing projection one is fostering. In such situations, face-work is the active management undertaken by individuals to counteract any potential incidents that may represent disruptions to the current expressive order (Smith 2006:51; Gottschalk 2018:41). When face is threatened, face-work is undertaken to restore the equilibrium within the situation. This can be undertaken by either interlocutor during the interaction when compensatory efforts are typically made to undertake the task of restoring face (Goffman 1967:27).

Face-work is a ritual interaction designed to enable a consistency of face. It is identified as falling into two overarching variants: avoidance and corrective processes (Smith 2006:51; Radford et al. 2011:435) or as Gottschalk (2018:41) identifies them as 'preventative' and 'restorative' processes. Avoidance or preventative processes involve a range of moves that seek to reduce potential hazards directed towards the face of an individual, by presenting the self with modesty, self-deprecation, a lack of seriousness, self-censoring, performing courtesies and tactful blindness, avoiding risky encounters, graciously withdrawing, using go-betweens, and avoiding topics that would show up inconsistencies (Gottschalk 2018:41). On the other hand, the corrective or restorative dimensions come into being when face-threating occurs and avoiding is no longer possible, where individuals are anticipated to perform restorative face-work (Gottschalk 2018:41). This can be identified as 'efforts like apologies to make good the face-threatening act that has occurred' (Smith 2006:51) and usually will proceed along the following track: first, a *challenge* is encountered where an action presented threatens an interlocutor's face; then an *offering* of a genuine apology is provided; an *acceptance* then follows were the interlocutor is anticipated to accept the apology provided; and finally *thanks* are then also offered which demonstrates a sense of gratitude that the apology was accepted (Gottschalk 2018:41). This provides a 'model for interpersonal ritual behaviour, but a model that may be departed from' (Goffman 1967:22). In the following, we argue that a key dimension of face-work on Twitter emphasises the avoidance dimensions of face-work at the expense of these restorative and corrective variants. The result is that Twitter, and the interactional practices undertaken by participants on the platform, represent an unbalanced variant of face-work.

Although using face-work to consider Twitter interactions may appear anachronistic given the internet was in embryonic form when Goffman conceptualised this theory, his insights into the impact on interaction and the materiality of technologies is significant (Pinch 2010:411). Moreover, we concur with Radford at el. (2011:431) when they state that 'face-work has the potential to greatly increase the understanding of interpersonal dynamics in computer-mediated communication realms'. In this sense the internet removes the temporal and geographic restrictions on communication, dramatically reducing the costs of knowledge exchange (Towne and Herbsleb 2012:98) enabling face-work's digital adaptation. There is also precedence for our use of face-work given other applications of Goffman's ideas to digital life (see Baker & Walsh 2018; Benediktsson et al 2015; Gottschalk 2018; Grønning and Tjora 2018; Murthy 2012; Tjora 2011; Walsh & Baker 2017; Walsh & Clark 2019; Walsh forthcoming). Of particular significance to our argument, Brownlie and Shaw (2019) offer a conceptual extension of Goffman's approach by considering how 'supportive interchanges' (Goffman 1971) work as a form of digital 'empathy ritual' that allow users to affirm each other's 'platform performance' to validate and acknowledge a generic relationship between doer and recipient on Twitter. Additionally, Sharkey et al. (2012) deploy the concept of face-work in online help forums for young people who self-harm, examining the way face is used to increase understanding and supportive interactions. When it comes to online environments like Twitter, there remains a dearth of social cues that increase the potential for misinterpretation while simultaneously affords a level of anonymity unavailable in face-to-face interactions (Groshek and Cutino 2016:2). This has implications for face-work that must be considered.

In some contexts, aggressive uses of face can come to override the prevailing collaborative ethos, departing from accommodating forms of face-work that is suggestive of more uncivil variants. The interaction ethos of a police interrogation or the participation of a defendant in a court case represent examples where interaction has a greater potential to become coarse, with deference to face far less likely (David et al. 2017; Roach and Mack 2015). In these situations, a 'winner' will use face in an interactional game by introducing favourable information about themselves while providing unfavourable information about the other to demonstrate they can handle themselves better than their adversaries (Goffman1967:12-13). These situations are comprised of 'aggressive interchanges' where information is conveyed either explicitly or implicitly to damage an interlocutor through a 'crack', implying the 'initiator is better at footwork than those who must suffer their remarks' (Goffman1967:12-13). Consequently, different faces are required to navigate and respond to the prevailing interactions across different encounters. As Goffman suggests:

> We have party faces, funeral faces, and various kinds of institutional faces, as the following comments on life in prison suggest: 'Every new inmate learns to dog-face, that is to assume an apathetic characterless facial expression and posture when viewed by authority' (Phillips 1950, as quoted in Goffman 1963:28).

If there are specific faces for different situations, extending this logic we contend the existence of a 'Twitter face'. This is not a literal face, but rather a version of face that is configured around platform affordances and interactions. Twitter-face is attributed to textual cues at the expense of embodied variants, given that online we tend to remain 'mute, invisible, and literally out of touch' (Gottschalk 2018:42). Just as Goffman demonstrates how inmates are careful to refrain from presenting overly welcoming visages, on Twitter the notion of a Twitter-face is accompanied by face-work practices that similarly adopt a defensive position. In this environment, users respond to the hostile and combative interactional ethos by adopting strategies that either elide negative sentiment or actively avoid it through modifying how one engages with other Twitter users. Significantly, we contend face-work strategies observed on Twitter accentuate the avoidance components of face-work as opposed to the restorative dimensions, as compared to Goffman's original conceptualisation. Through emphasizing the avoidance components of face-work rather than its restorative dimensions, this has consequences for the nature of interaction on Twitter with respect to experiences of incivility. The significance of this is that during conflictual interaction, user behaviour and Twitter's own platform affordances tend to undermine attempts to reduce hostile interactions.

Incivility and social interaction

While Goffman's discussion of face-work never invokes the concept of incivility, there are important conceptual linkages (Smith 1997:60; Papacharissi 2004). One of Goffman's most well-known concepts, 'civil inattention', comes in some respects closest to signifying the importance of behaviour affording an individual interactional dignity. Civil inattention occurs when individuals remain unknown to one another, yet nonetheless demonstrate—usually through fleeting eye contact—they do not constitute a special curiosity and withdraw from the demands of focused attention (Goffman 1963:84). The concept is exemplified in the way strangers in a public setting fleetingly look at one another, size the other up, but then glance downwards in a show analogous to the passing of vehicles dipping their lights for one another (Smith 2006:38; Goffman 1963:84). This interaction can be scarcely perceptible and delicately nuanced (Smith 1997:60) and therefore represents one of 'the slightest of interpersonal rituals' (Goffman 1963:84). While almost undetectable, it allows for the possibility of organising everyday life in a way that would otherwise be impossible, rendering co-presence possible without co-mingling, awareness without engrossment and courtesy without conversation (Lofland 1989:462). The act of affording another individual civil inattention communicates to them that they have 'no reason to suspect the intentions of the others present and no reason to fear the others, be hostile to them, or wish to avoid them' (Goffman 1963:85). In this respect, there is a level of regard provided, offering the basis of an equal footing in civic life.

In exploring experiences of incivility online, Papacharissi (2004:262) argues it is important to distinguish between civility and impoliteness. Invoking Goffman's account of interaction, Papacharissi indicates passions and robust conversations can upend polite conversation even though there may be various attempts 'in ensuring that events do not occur which might effectively carry an improper expression' (Goffman 1971:40). Papacharissi suggests it is important not to conflate civility with interpersonal politeness because this understanding can eschew the democratic nature of heated and forthright conversation (Papacharissi 2004:260). This distinction is pertinent to online interactions, 'where anonymity makes it easier for individuals to be rude, although not necessarily uncivil' (Rowe 2015:128). Eliding excessive adherence to conversational norms can also be part of face-work: 'Too much perceptiveness or too much pride, and the person becomes someone who is thin-skinned, who must be treated with kid gloves, requiring more care on the part of others than he may be worth to them' (Goffman 1967:40). Papacharissi's reading of Goffman highlights the delicacy of interaction and

the spontaneous dimensions that can clash with notions of politeness, but not—as she argues—with notions of civility. In rendering these differences between politeness and uncivil interaction, she suggests that a distinction should be drawn that defines 'politeness as etiquette-related, and civility as respect for the collective traditions of democracy' (Papacharissi 2004:260). In a somewhat distinct vein, Chen (2017:6) contends incivility can be defined as interaction exhibiting one of three main dimensions: 'insulting language or name-calling; profanity; and a larger category that encompasses stereotypes, and homophobic, racist, sexist, and xenophobic terms that may at times dip into hate speech'. These dimensions are said to exist on a 'continuum of aversive speech that both violates what is considered normal in conversation and also has the potential to cause harm' (Chen 2017:6). Incivility though undesirable, is nonetheless part of the human experience that we must tolerate to 'have the type of engaged electorate that informs a strong democracy', but simultaneously Chen (2017:4) notes, 'some forms of incivility are so harsh that they offer no contribution to public discourse'. These tendencies regarding uncivil interaction are exacerbated when transposed online. Unlike when these interactions occur face to face, 'a blistering online taunt lasts longer and potentially exposes the target to embarrassments on a larger scale' (Chen 2017:64). Therefore, using face-work to consider interaction on Twitter is valuable in understanding our civic interactions 'because face refers to our management of our public identity' (Papacharissi 2004:263).

Affordances

In addition to the theory of face-work and notions of incivility, literature on technological affordances further demonstrates how social media shapes social interaction. Affordances refer to 'the range of functions and constraints that an object provides for, and places upon, structurally situated subjects' (Davis and Chouinard 2016:241). They describe the possibilities for action between a technology and user, enabling or constraining potential behavioural outcomes (Evans et al.2017:36). The character limit on Twitter, for example, can be seen to afford brief, impulsive social exchange, given the absence of context afforded by Twitter's character limit. We can identify communication designs, such as increases to the character limits for tweets, as a type of change that seeks to intervene in an established activity to provide possibilities for interaction and simultaneously constrain and remove other possibilities for interaction (Aakhus 2007:114). Other affordances include the type of information a user is required to share when interacting on a platform. For example, researchers compared posting to a website, to posting on a Facebook page where users are identified with their account information. In this case, Rowe (2015:132) found 'the occurrence of uncivil… comments are significantly more common on the website version of the Washington Post, where users are able to maintain their anonymity, compared to the Facebook version of the Washington Post'. This finding suggests that the affordance of being identifiable impacts civility **online** with anonymity associated with uncivil behaviour.

Another affordance relevant to face-work concerns the length and terseness of communication. As Murthy (2012:1069) notes, brevity is a distinguishing aspect of communication on Twitter. We can get a sense of how this constraint on communication represents a type of affordance prefiguring interaction by comparing the social media sites, Twitter and Facebook. As Oz et al. (2018) suggest, the length of posts, as well as de-individuation (the feeling of being 'freed' from one's own identity) on Twitter, encourages users to feel less inhibited in their speech compared to Facebook, which the authors suggest affords more deliberative discussion. Brevity affords more civil communication with research suggesting Twitter's decision to double 'the length of tweets contributed to less uncivil political discussions and more deliberative political discussions' (Jaidka et al. 2019:363). Even the type of hardware one utilizes to interact on the platform (e.g. a mobile device, compared to a desktop computer), represents a type of affordance impacting the likelihood of uncivil or impolite tweets, where more caustic tweets are more likely to originate from mobile devices compared to fixed web devices (Groshek and Cutino 2016). Moreover, users posting from mobile phones tend to use negative language more frequently, possibly due to the way in which the device shapes their engagement with others (Murthy et al. 2019:834). Here, the role of affordances and the way they 'create a scaffold through which artifacts request, demand, allow, encourage, discourage, and refuse' (Davis and Chouinard 2016:246) particular types of interaction is critical, when we consider how this plays out in terms of the face-work of Twitter users.

Methods

To understand how users respond to hostile encounters online, we examined user experiences and the interaction strategies adopted to navigate Twitter. Our aim was to ascertain how users manage potential conflicts and disputes on Twitter. To achieve this, we deployed a questionnaire to analyse user experiences to produce qualitative data to 'understand meaning-making, placing technology use into specific social contexts, places, and times' (Marwick 2013:119). While eager to explore the more objective markers of these hostile experiences,

we were aware that incivility can also include 'whatever is taken as offensive, impolite or crude because human subjects impose meaning on actions' (Smith, Phillips King 2010:11). In this regard, user perspectives that provide a sense of context, interpretation, meaning and intention all play a role where uncivil interaction is experienced, and these dimensions are all the more important in situations where communication channels are attenuated to solely textual information.

For this study, we devised an online questionnaire aiming to examine participant experiences of engaging with other users on Twitter.[2] Participants were recruited on social media using the snowballing technique with data collected from July to September 2019. During this period, 99 respondents completed the questionnaire. The sample we gathered from this process was, therefore, a convenience sample rather than representative of the population. While we were not aiming to achieve representativeness, our approach sought to determine if there were patterns or common approaches in responding to hostile encounters in order to ground our theoretical discussion. Our intention was to explore individual experiences and examine whether user strategies navigating the platform were shared among participants. The study includes closed and opened-ended responses with participants providing extended responses to questions (an abbreviated version of the questionnaire is provided in the appendix). The topics in the questionnaire included: demographic information, Twitter use, experiences of uncivil interaction and finally attitudes towards removing users from the social networking site (the topic of deplatforming however is not considered in this article due to space constraints). Once the responses to the questionnaire were collected, thematic analysis was used to identify the major themes provided by participants.

The sample comprised a relatively even split between female and male respondents: 55 per cent were female, 43 percent were male, with the remainder identifying as non-binary. The age distribution of participants was also somewhat evenly spread with the exception of respondents in the youngest age bracket (4 per cent for 18-24 years of age, just over 20 per cent were aged between 25-34, just over 20 percent for those aged 35-44 and the same for those aged 45-54, almost 24 per cent for those aged 55-64, with 65+ coming in at just over 6 per cent of the sample). Overwhelmingly, those who participated in the questionnaire identified as white (90 per cent), with just over 6 per cent identifying Asian (the remainder of participants did not denote their cultural background). Given our Twitter followers lived in Australia this resulted in our sample residing mostly in Australia (90 per cent), with most working (close to 92 per cent), across various industries, especially education (almost 30 per cent of the sample), Government and Administration (13 per cent) and Business and Financial (7 per cent).

The limitations of this study should also be considered. Given the number of participants, we were unable to systemically compare responses across different demographics of Twitter users. Indeed, future research could follow our suggested argument concerning the role of face-work to determine if it is contingent on demographic criteria, such as age or gender, for example. Nevertheless, where possible, we have attempted to include relevant demographic information about respondents to contextualise their experiences. We found Twitter users adopt several interactional practises that correspond with Goffman's notion of face-work. However, we also note at the time of data collection we were not seeking to empirically capture face-work practices. This analysis therefore is one that retrospectively applies the concept of face-work to illuminate the practices Twitter users undertake and do so in order to return to the small interchanges happening in digital public spaces to remind ourselves of the emotional gamble of sharing online (Brownlie and Shaw 2019:107). Moreover, while amplification is a significant contributor to hostility online, our focus is on the strategies employed by individual Twitter users to manage incivility rather than mapping the social effects of amplification on the platform. Notwithstanding these limitations, we suggest these do not negate our attempts to illuminate the interactional dynamics of Twitter users.

Face-work on Twitter: Strategies for avoiding uncivil interactions

While Twitter enables social interaction among users, the platform also facilitates aggressive interaction that aims to diminish the face of others. Although avoidance and remedial practises are considered critical for face-work, our analysis suggests avoidance practices are far more common on the micro-blogging site. In other words, face-work on Twitter appears unbalanced. It is this emphasis on avoidance practices users adopt that provides some insight into the interaction ethos on Twitter. In our analysis we found that Twitter users articulate several interactional moves to navigate incivility on the platform.

Multiple accounts and protected tweets: responses to Twitter's affordances

---

[2] The project 'Twitter Health and Deplatforming' received ethics clearance from the University of Canberra ethics committee.

One approach used to manage consistency of face is the use of multiple Twitter accounts. Unlike Facebook, which discourages users from creating more than one personal account—it is against Facebook's community standards to maintain more than one personal account (Facebook 2019)—Twitter allows users to create multiple accounts (Twitter 2019b). This affordance is significant given over one third of our respondents indicate they possess more than one account. This process can allow users to segregate audiences, so those who witness the performance associated with one account will not be able to witness—in a non sequitur fashion—the individual in another role (Goffman 1959:137). Thus, switching between accounts is one way to achieve greater control of the scheduling of one's platform performance (Brownlie and Shaw 2019:106-7). The motivations identified for those using multiple accounts tend to fall into two main categories. The first is users' desire to distinguish their personal Twitter activity from their professional personas. As one respondent reports:

> Separation of professional & personal. I rarely use my personal account though (mostly use when wanting to join in with twitter conversations about popular reality shows!) (Female aged 25-34)

Or, as another indicates:

> I have one for posting in my more professional capacity and one for more frank views (Female aged 55-64).

The other main reason participants cited for possessing more than one account was to enable greater propriety in managing interactions:

> A particular name occurred to me and I wanted a second account for it. I'm often on both because I was harassed on one, and created an alternative one to get away from the harassers. (Male aged 55-64)

The preceding statements highlight the use of multiple accounts as a strategy to avoid harassment. However, this motivation also appears as a strategy to access tweets from accounts some users have been blocked from viewing and therefore, as one responded indicated, multiple accounts can be used 'to get around blocked tweets' (Male aged 35-44).

Another approach adopted to manage face and avoid uncivil interaction is by adopting to approve each person who can view an account by enabling 'protected' tweets (Jin 2013:816). When users 'protect' their tweets this renders the user's account secluded and means content cannot be retweeted or viewed by those who have not been given access to view the protected account. In contrast, when an account is open, the user's tweets are public by default and the public is invited to become investors in the information provided (Jin 2013:816). Most respondents indicate they tweet without 'protecting' their profiles: 94 per cent of our respondents indicated that their accounts are public, with a minority opting to control who can follow their tweets. This figure also corresponds with other reported data; for example, Pew's study suggests 13 per cent of users protect their Twitter accounts (Remy 2019). This represents a dimension of Twitter's affordances that assists in managing face via segregating and controlling who the user accords audience status. Given protecting an account enables the user to accept or refuse another user to view their account, this process disables the possibility of interacting with a person nonratified, thus protecting the face of a user more discreetly. This is suggested when probing participant motivations regarding the concealment of their tweets; respondents suggest they have concerns about maintaining privacy and because they have 'personal views not meant to be shared outside of my circle' (Female aged 25-34). Others indicate they protect their tweets because:

> I have social anxiety. I never post tweets. I use twitter to view what others are saying and don't participate in conversations (Male aged 45-54).

Or, as explained by another respondent:
> I switch between the two depending on where I'm at with life (Male aged 35-44).

While Twitter is public by default, these user preferences for protected tweets are revealing of the interactional ethos on the platform and a desire to manage face. It is also suggestive of what Marwick and boyd (2014:1045) argue is sometimes misunderstood about sharing on social media; sharing online does not render users ungoverned by conceptions of privacy. Privacy remains vital, much like acts of managing privacy in other public contexts (Cahill et at 1985; Goffman 1973;); users of social media take similar steps to accomplish privacy in networked publics (Marwick and boyd 2011) and along with this manage the extent to which face is publicly exposed. While most participants perceive Twitter as a public medium that allows any post to proliferate across its platform, those who protect their tweets do so as a way of managing their exposure to the undefined audiences characteristic of Twitter. This is because Twitter facilitates 'context collapse', flattening multiple audiences into one, rendering it difficult to adopt a singular identity and therefore compelling users to manage overlapping audiences to strategically conceal information (Marwick and boyd 2010:122). Protecting

tweets and using multiple accounts as discussed in this section represents two important approaches of managing face-work, highlighting its avoidance dimensions; users aim to sidestep the hazards associated with the open and public nature of the platform. We now turn to another overt measure: blocking users.

Blocking: responses to others

A common response to uncivil interaction is to ignore the action that seeks to provoke a corresponding reaction. Smith's (2001:171) study of runners' exercising found the most prevalent response towards uncivil interaction is not to dignify the original remark (see also Gardner 1980). For many Twitter users, this disregard represents one main variant of this approach, but here blocking a user's account furthers cements the eliding of hostile interjection, as it abruptly halts the possibility of any additional interaction. It is also in one important way different from the example above as users are excluded from interacting with one another. Whereas in the case of Smith's study, ignoring fails to provide the offender with a means to engage in an interactional game. Interestingly, while blocking users is enthusiastically reported by participants as one way to cease interactional transgressions, it is not extensively reported, compared to the techniques discussed below. For example, when asking participants if they have strategies for avoiding conflict on Twitter one responded suggested:

> I block unpleasant people and have filters to avoid unpleasant content (Male aged 45-54).

Another indicates they may even block in advance of an anticipated interaction:

> Blocking trolls and idiots - sometimes pre-emptively - eg journo's who are more trolls (Male aged 35-44).

Blocking enables users to cease communicating with an interlocutor's account without the need for any leave-taking action. The potential to abort interaction in this manner is an important dimension of Twitter's affordances that shape how interactions of a hostile nature might be encouraged. The ability to block may increase the possibility of aggressive face-work, as blocking remains part of the interaction strategies that emphasize avoidance dimensions, at the expense of engaging restorative face-work. Either party has the option of unilaterally blocking the other and this renders the need to cooperate through interaction diminished. This is because blocking does not involve any leave taking rights and presumes a terminal ceasing of personal relations (Adato 1975). Blocking, while discouraging less subtle variants of face-work, has the potential to empower the user by enabling them to neutralise explicit threats to face. This move represents a blunter instrument, users no longer must work together to conduct face-work, overriding the collaborative ethos constituting co-present face-work where blocking is unavailable as a strategy.

Avoiding topics, users and salvos: responses to others

While blocking remains ever present, there are several other ways respondents highlight how they seek to minimise the potential for conflict. Most respondents (80 per cent) indicate that they employ strategies to avoid conflict on Twitter, with open-ended responses offering insights. In this case, averting salvos aimed at causing offence or aimed to goad another user into a cycle of aggressive face-work was evaded. For example, one participant noted that they, "Avoid impulsive and reactionary participation unless it is mindful and adds value. I.e. - facts and evidence" (Female aged 45-54). Another indicated some users might warrant being avoided because of how they present themselves through their profile information, signifying the user might be difficult to interact with:

> If users have cartoon profile pictures or lots of emojis in their bio (or any obvious marker that they belong to troll communities) I won't bother arguing or responding. I also try not to 'pile on' to a tweet that I disagree with, as that seems unnecessary. I will block occasionally (Female aged 25-34).

Another suggests:

> SIMPLE - Don't get involved in the first place, and think about what you are posting or retweeting before doing so (Male aged 25-34)

These acts users identify represent avoidance practices or a type of pre-emptive non-response whereby users eschew acts likely to threaten face (Smith 2006:51). As Goffman (1976:15) reminds us, the best way for a person to prevent threats to face is to evade 'contacts in which these threats are likely to occur… in many societies, members know the value of voluntarily making a gracious withdrawal before an anticipated threat to face has had a chance to occur'. In practicing disregard towards topics or users that appear hazardous, participants demonstrate a keen awareness to head problematic interaction of at the pass, emphasizing, once again, the protective aspects of face-work.

Partial involvement and passive consumers: responding to other users through self-censoring

Respondents also adopted face saving postures by curtailing how they interact on Twitter. Users describe the practice of either 'broadcasting' monologically or simply 'liking' content; replying to users or engaging in conversation is eschewed. The range of interactional possibilities is limited with users deliberately adopting a partial involvement strategy to minimise exposure to interacting directly with other users. For example, one user indicates: "I don't respond to anything and only broadcast, but I am ready to abandon Twitter again. It is so lame" (Male 45-54). Another suggested that they:

> On my personal account I prefer to limit my interactions to 'liking' to avoid unwanted attention and hostility, as my views are highly liberal and I am not interested in engaging in the toxicity I have witnessed from trolls and the extremely conservative. I generate content only on my professional account, primarily for the purpose of networking (Female aged 35-44).

Another strategy identified by participants was the selection, considered curation and moderation of content posted. Unlike the partial involvement strategy outline above, users in this mode of face-saving, upload content but do so in a considered manner; engaging only in certain types of conversations (if at all) and actively deliberating how a tweet or a reply may be perceived from the perspective of other users. As Gottschalk (2018:41) suggests, 'self-censoring, performing courtesies and tactful blindness, avoiding contact or risky situations…staying off topic and away from activities that would reveal inconsistencies in the lines participants are performing' are adopted as preventative face-work strategies. The aim is to avoid the potential for being drawn into conflict by considering how a message is presented. For example:

> I tend to rephrase anything I think could be even slightly controversial as a question instead of a flat statement. I also assume at first blush that a nasty/abrupt comment from someone could in fact be a miscommunication or poor use of language—you will be surprised how often that is the case—so things don't escalate (ie give benefit of the doubt). If someone looks like they want a fight, I will normally just walk away, either muting or blocking (Female aged 45-54)

Some users even suggest that they might provide a response, but then remove it to avoid escalations: "Sometimes I'll type out a response and then delete it. Escalating the situation, or giving an assailant the satisfaction of a reply, is rarely productive" (Female aged 45-54). Responses of this kind indicate an active consideration with respect to formulating comments prior to replying, especially where there is the potential to damage the face of an interlocutor. This approach aligns with the avoidance dimensions of face-work in that topics that increase the likelihood of threats to face are dodged (Smith 2006:51). Some users will even go so far as to adopt a 'reasonability test', suggestive of a type of selection and deliberation process. Users therefore approach their interactions with users on Twitter to consider the ideas or facts in a genuine, open-minded manner and concede where appropriate. For example:

> If I sense that someone isn't going to engage in good faith, and will misrepresent what I post in a way to bolster their own position, then I may not engage with them. I've had bad experiences of this before, so aware of what can happen when someone is determined to strawman you (Male aged 35-44).

This strategy was commonly identified, as another user explains:

> Only respond to negative comments on my own content. Diffuse aggression in these comments but not escalating, but also clearly explaining with evidence why the comment is wrong/misguided. Also, I reflect, and if I think the negative comment is actually justified, I will concede the point and explain as required (Female aged 35-44).

There are several moves available to users in navigating the interaction order of Twitter as the above suggests with various approaches to manage face and avoid threats to it. Our analysis suggests interaction on Twitter skews towards a more hostile ethos and this corresponds with an overcorrection emphasising the avoidance practices. This includes blocking users (sometimes pre-emptively), avoiding particular topics or partially interacting online, all at the expense of more restorative approaches. Though we should note, as suggested by the previous comments, there are nonetheless some attempts at a restorative or collaborative position, however, these were not reported as frequently, presumably because they require a greater level of cooperation from interlocutors that was less common. Users therefore rely on the defensive approaches as our respondents suggest. This is because such defensive practices can be undertaken with greater ease in the ambiguous informational context that Twitter represents. This also makes sense when we consider face-work on twitter departs somewhat from the original face-to-face corrective cycle of interaction, further placing Twitter users on a limb should they wish to seek to restore fractured interactions. As Goffman notes:

An important departure from the standard corrective cycle occurs when a challenged offender patently refuses to heed the warning and continues with his offending behavior, instead of setting the activity to rights. This move shifts the play back to the challengers. If they countenance the refusal to meet their demands, then it will be plain that their challenge was a bluff and that the bluff has been called. This is an untenable position; a face for themselves cannot be derived from it, and they are left to bluster. (1967:22-23)

Avoiding the risk of escalating hostilities, the defensive track is adopted. Users evade the potential costs of further exposing the vulnerability of their face to others who are less likely to play a role in collaborative face-work.

Conclusion

In exploring how incivility is responded to by users on Twitter, we reveal some of the practices of face-work and dimensions of social interaction in this context. We have provided insights into some of these practices employed by users to navigate the platform. We have also sought to understand how Twitter encourages particular modes of social interaction. Critically, we found there is an overemphasis on the avoidance components of face-work in lieu of restorative and redemptive capacities. Nonetheless, Twitter users have been shown to adopt and extend face-saving practices when exposed to hostile encounters, illuminating the interaction order present on Twitter. By communicating on Twitter, we put ourselves up for public display and expose our faces to more hostile forms of communication. Conceived in this way, the face-work practices examined speak to the collective organisation and curation of face that is compelled in order to navigate the public nature that Twitter as a social media platform assumes. In a context where Twitter seeks to shape the nature of interaction on its platform through the notion of conversational health, considerations of face-work and the implications that stem from enabling and disabling its various forms should be further considered. Given Twitter's public nature, it appears that attempts at developing conversational health metrics—while aimed at reducing hostility and increasing user engagement—need to adopt a wider perspective factoring in the affordances of the platform. Additionally, user behaviour as it interacts with these affordances represent vital dimensions, requiring consideration when engineering redemptive and restorative Twitter mechanisms into the future. Given that social media combines elements of broadcast media with face-to-face communication, where social contexts are collapsed, greater awareness of the nature of platform affordances and processes of face-work should be incorporated into mitigation practices aimed at combating uncivil interaction. After all, the strategies examined in this article, rather than signifying categorically distinct moves in our interaction rituals, represent important continuities in the varied nature of face-work that find emergent forms of expression online.

References

Aakhus M (2007) Communication as Design *Communication Monographs* 74(1):112-117. DOI: 10.1080/03637750701196383

Adato A (1975) Leave-taking: A study of commonsense knowledge of social structure. *Anthropological Quarterly* 48(4):255-271.

Baker, S. A., and M. J. Walsh. 2018. "'Good Morning Fitfam': Top Posts, Hashtags and Gender Display on Instagram." New Media & Society 20 (12): 4553–4570. doi:10.1177/1461444818777514.

Benediktsson M, Alexander D, Bermeo J, Contreras J, Kingston B, Harper W, Henkin J, Lopez F, Wagenheim R, and Williams W (2015) Hybrid strategies: Allocating involvement in the digital age. *Symbolic Interaction* 38(3):331–351. https://doi.org/10.1002/symb.160.

Brownlie J, Shaw F (2019) Empathy rituals: Small conversations about emotional distress on Twitter. *Sociology* 53(1):104–122.

Cahill S, Distler W, Lachowetz C, Meaney A, Tarallo R, and Willard T (1985) Meanwhile backstage: Public bathrooms and the interaction order. *Urban Life* 14(1): 33–58.

Chen, G (2017) *Nasty Talk: Online Incivility and Public Debate*. Cham, Switzerland: Palgrave Macmillan.

Collins R (2008) *Violence: A Micro-sociological Theory*. Princeton, New Jersey: Princeton University Press.

David GC, Rawls AW, and Trainum J (2017) Playing the interrogation game: Trust, coercion and confession in police interrogation. *Symbolic Interaction*. 41(1):3–24.

Davis J and Chouinard JB (2016) Theorizing Affordances: From Request to Refuse. *Bulletin of Science, Technology & Society* 36(4): 241–248.

Evans S, Pearce K, Vitak J and Treem J (2017) Explicating Affordances: A Conceptual Framework for Understanding Affordances in Communication Research *Journal of Computer-Mediated Communication* 22:35-52. https://doi.org/10.1111/jcc4.12180

Facebook (2019) 'Can I create multiple Facebook accounts?' Help Center, accessed 6/11/19 https://www.facebook.com/help/975828035803295

Fornacciari P, Mordonini M, Poggi A, Sani L and Tomaiuolo M (2018) A holistic system for troll detection on Twitter *Computers in Human Behavior* 89 (2018):258–268.

Gardner CB (1980) 'Passing by: Street Remarks, Address Rights and the Urban Female', Sociological Inquiry, 50:328–56.

Goffman E (1959) *The Presentation of Self in Everyday Life*. New York: Anchor Books.

Goffman E (1963) *Behaviour in Public Places: Notes on the Social Organization of Gatherings*. New York: Free Press of Glencoe.

Goffman E (1967) *Interaction Ritual: Essays on Face-to-Face Behavior*. Chicago, IL: Aldine. Publishing Company

Goffman E (1971) *Relations in Public: Microstudies of the Public Order*. New York: Basic Books.

Gottschalk S (2018) *The terminal self: Everyday life in hypermodern times*. Abingdon, Oxon: Routledge.

Grønning I and Tjora A (2018) Digital absolution: Confessional interaction in an online weight loss forum *Convergence: The International Journal of Research into New Media Technologies* 24(4):391–406.

Groshek J and Cutino C (2016) Meaner on Mobile: Incivility and Impoliteness in communicating contentious Politics on Sociotechnical Networks *Social Media + Society* 2 (4):1-8. https://doi.org/10.1177/2056305116677137

Jaidka K, Zhou A and Lelkes Y (2019) Brevity is the Soul of Twitter: The Constraint Affordance and Political Discussion *Journal of Communication* 69(4):345–372.

Jin, SA (2013) Peeling back the multiple layers of Twitter's private disclosure onion: The roles of virtual identity discrepancy and personality traits in communication privacy management on Twitter *New Media & Society* 15(6) 813–833.

Lawson C (2018) Platform vulnerabilities: Harassment and misogynoir in the digital attack on Leslie Jones. *Information, Communication & Society* 21:6, 818-833, DOI: 10.1080/1369118X.2018.1437203

Lee J, DeCamp M, Dredze M, Chisolm M, and Berger Z (2014) What are health-related users tweeting? A qualitative content analysis of health-related users and their messages on twitter. *Journal of Medical Internet Research*. 16(10).

Lofland L (1989) 'Social Life in the Public Realm: A Review', *Journal of Contemporary Ethnography* 17(4):453–482.

Marwick A (2013) 'Ethnographic and Qualitative Research on Twitter.' In Weller, 1 K., Bruns, A., Puschmann, C., Burgess, J. and Mahrt, M. (eds), Twitter and Society. New York: Peter Lang, 109-122.

Marwick A and boyd d (2010) I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society* 13(1):114–133.

Marwick A and boyd d (2011) To see and be seen: Celebrity Practice on Twitter. *Convergence: The International Journal of Research into New Media Technologies* 17(2):139–158.

Marwick A and boyd d (2014) Networked privacy: How teenagers negotiate context in social media. *New Media & Society* 16(7):1051–1067.

Matamoros-Fernández A (2017) Platformed racism: The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society*, 20:6, 930-946, DOI: 10.1080/1369118X.2017.1293130

Murthy T (2012) Towards a Sociological Understanding of Social Media: Theorizing Twitter *Sociology* 46(6):1059–1073.

Murthy T (2013) *Twitter: Social Communication in the Twitter Age*. Cambridge: Polity.

Murthy T, Bowman S, Gross A and McGarry M (2015) Do we tweet differently from out mobile devices? A study of language differences on Mobile and Web-based twitter platforms. *Journal of Communication* 65:816–837. https://doi.org/10.1111/jcom.12176

Oz M, Zheng P and Chen GM (2018) Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes, *New Media & Society* 20(9):3400-3419 https://doi.org/10.1177/1461444817749516

Radford ML, Radford GP, Connaway LS, et al. (2011) On virtual face-work: An ethnography of communication approach to a live chat reference interaction. *Library Quarterly* 81(4):431–453.

Remy E (2019) How public and private Twitter users in the U.S. compare — and why it might matter for your research. *Pew Research Center*. <https://medium.com/pew-research-center-decoded/how-public-and-private-twitter-users-in-the-u-s-d536ce2a41b3> date accessed 7.2.20.

Roach S, Mack K (2015) Performing authority: Communicating judicial decisions in lower criminal courts. *Journal of Sociology*. 51(4):1052-1069.

Rowe I (2015) Civility 2.0: A comparative analysis of incivility in online political discussion, *Information, Communication & Society*, 18(2):121-138. DOI:10.1080/1369118X.2014.940365

Sharkey S, Smithson J, Hewis E, Jones R, Emmens T, Ford T, and Owens C (2012) Supportive interchanges and face-work as 'protective talk' in an online self-hard support forum. *Communication and Medicine*. 9(1):71-82

Smith G (1997) Incivil attention and everyday intolerance: Vicissitudes of exercising in public places *Perspectives on Social Problems* 9:59-79

Smith G (2001) Techniques of neutralization, techniques of body management and the public harassment of runners. in Cunningham-Burley S., Backett-Milburn K. (eds) *Exploring the Body*. Palgrave Macmillan, London

Smith G (2006) *Erving Goffman*. Abingdon, Oxon: Routledge.

Smith P, Phillips T and King R (2010) *Incivility: The Rude Stranger in Everyday life*. Cambridge: Cambridge University Press.

Papacharissi Z (2004) Democracy online: Civility, politeness, and the democratic potential of online political discussion groups *New Media & Society* 6(2):259-283.

Pew (2017) "Online Harassment 2017" < https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/>

Pew (2018a) "Crossing the Line: What counts as Online Harassment?" <https://www.pewinternet.org/2018/01/04/crossing-the-line-what-counts-as-online-harassment/>

Pew (2018b) "Social Media Use in 2018" https://www.pewinternet.org/2018/03/01/social-media-use-in-2018/

Pinch T, (2010) The Invisible Technologies of Goffman's Sociology from the Merry-Go-Round to the Internet. *Technology and Culture* 51 (2):409–424. doi:10.1353/tech.0.0456.

Tjora AH, (2011) Invisible Whispers: Accounts of SMS Communication in Shared Physical Space *Convergence: The International Journal of Research into New Media Technologies* 17(2) 193–211.

Towne B, and Herbsleb J (2012) Design Considerations for Online Deliberation Systems *Journal of Information Technology & Politics* 9(1):97-115.

Twitter (2018a) "Measuring Healthy Conversation" <https://blog.twitter.com/en_us/topics/company/2018/measuring_healthy_conversation.html

Twitter (2018b) "A healthier Twitter: Progress and more to do" <https://blog.twitter.com/en_us/topics/company/2019/health-update.html>

Twitter (2019a) 'Our 2018 Inclusion and Diversity Report' https://blog.twitter.com/en_us/topics/company/2019/ianddreport.html

Twitter (2019b) 'How to manage multiple accounts' Help Center, accessed 6/11/19 https://help.twitter.com/en/managing-your-account/managing-multiple-twitter-accounts

Walsh, M. J., and S. A. Baker. 2017. "The Selfie and the Transformation of the Public–private Distinction." Information Communication & Society 20 (8): 1185–1203. https://www.tandfon line.com/doi/full/10.1080/1369118X.2016.1220969

Walsh, M.J., & Clark, S.J. (2019). Co-present conversation as "socialized trance": Talk, involvement obligations, and smart-phone disruption. Symbolic Interaction, 42(1), 6–26.

Walsh, M.J. (Forthcoming). About 'face': Reconsidering Goffman's theory of face-work for digital culture. In Michael H. Jacobsen and Greg Smith (eds.), The Routledge International Handbook of Goffman Studies (Routledge, forthcoming).

Appendix: Abbreviated version of online Twitter conversation questions

- How many of your twitter followers have you met in person? Response options: All of them/ Most of them/ About half of them / A quarter of them/ A few of them/ None of them

- Do you feel twitter has become an unfriendly space? Response options: yes/no
  - Please describe why you responded this way? (textbox response)

- Do you tweet publicly, or do you have protected tweets? Public tweets—the default setting on twitter—are visible to anyone, whether or not they have a Twitter account, whereas, protected tweets are only visible to your Twitter followers)? Response options: Public tweets/ protected tweets

- Do you have multiple twitter accounts? Response options: yes/no
  - If you have multiple twitter accounts, please tell us why? (textbox response)

- When you come across content on twitter that your strongly disagree with how do you typically respond? (textbox response)

- Do you engage in conversations with other users who you strongly disagree with? Response options: Yes—I engage in conversations with those disagree/ No—I do not engage in conversations with those disagree
  - If you do interact with other users that you disagree with how you manage these interactions? (textbox response)

- What do you do to minimize the extent to which others might disagree with content you post through twitter? (textbox response)

- Do you think the way twitter is currently organized allows users to engage in conversations in a respectful manner? Response options: yes/no response
  - Why? (textbox response)

- Have you blocked users of twitter? Response options: Yes, all the time/ Yes, occasionally/ No, although I have been tempted/ No, I have never felt the need to block another user
  - Why have you blocked users in the past? (please describe what interactions lead you to take this action) (textbox response)

- Do you feel safe when using twitter? (textbox response)

- Do you believe Twitters attempts at managing user behavior has been successful or unsuccessful? Response options: It has been successful/ It has been minimally successful/ It has been unsuccessful/ Don't know
  - Why? (textbox response)