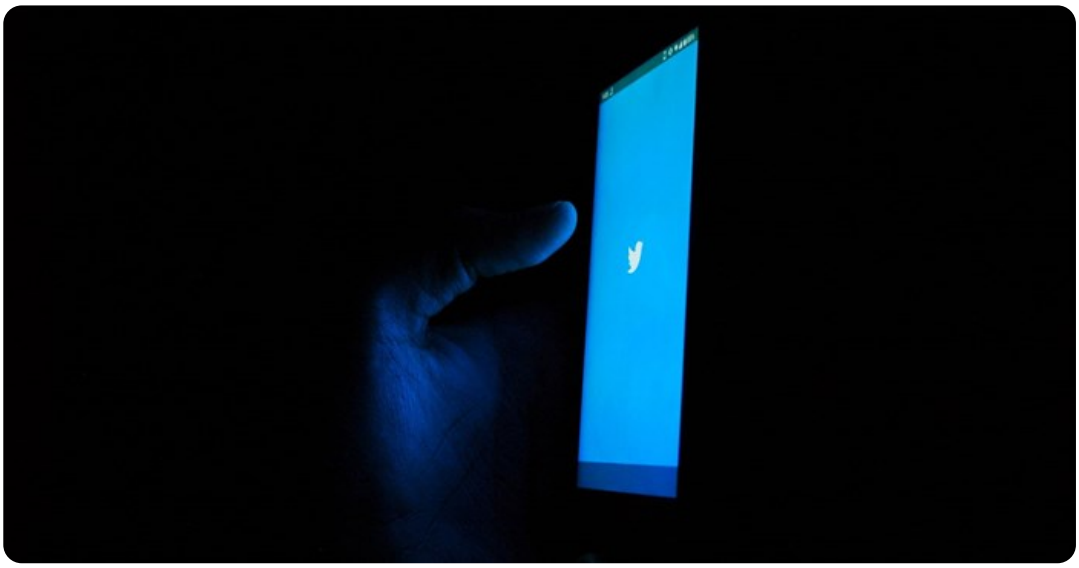
bizcommunity

Twitter to ask users to rethink abusive messages - a promising step towards 'slowcial media'

In an effort to reverse the flood of abuse on the platform, Twitter is rolling out a new feature whi will show a self-moderation prompt to users who compose replies that the platform's algorithms recognise to be abusive. The prompt effectively asks users to think twice before posting an abusive message.

By Martin Graff 24 May 2021



Source: www.unsplash.com

Because it compels users to rethink and reflect on abusive tweets, Twitter's new self-moderation prompt could be a promising step away from fast and furious social media posting and towards a more considere slow media – or "slowcial media".

Twitter's new feature

This isn't the first time Twitter has trialled and released 'nudges' aimed at addressing poor behaviour on th platform. In 2020, Twitter added misinformation labels to tweets in response to Covid-19 conspiracies circulating on the platform, which they say reduced the number of tweets quoting misleading information b 29%.

But online abuse continues to be a highly contentious issue for Twitter, with reports of celebrity abuse, aimed particularly at women, commonplace on the platform. Twitter's new self-moderation prompt aims to address this problem.

The new prompt has been trialled for select accounts and regions since May 2020. Twitter shared the results of this trial in a recent blog post, announcing that 34% of people who encountered the prompt revis their initial reply – or deleted it altogether. They also claim those who'd been prompted once composed a average of 11% fewer offensive tweets in the future.

Why the abuse?

Online behaviour is often characterised by a tendency to act in a less inhibited way than one might act offli as when users post abuse they'd not necessarily share in a face-to-face context. Research suggests this disinhibition stems from our feeling of anonymity and invisibility online – and the absence of any perceived authority to prevent us from misbehaving.

I've previously been involved in studies that investigated the different ways in which people seek validatior from posting on social media. We found that people were often prepared to manipulate posts to increase the degree of attention they received in the form of likes. They even reported blindly posting about issues they didn't necessarily agree with, explaining that they did this to boost their spirits or self-esteem.

All this seems to suggest that social media platforms are a unique environment where individuals post with little prior consideration as to whether that post could offend or upset others.

To understand why Twitter's new nudge – adding a little friction to the instantaneous process of posting a tweet – appears to be reducing abusive replies, we can look at what existing studies tell us about the sources of online abuse.

> *When things get heated, you may say things you don't mean. To let you rethink a reply, we're running a limited experiment on iOS with a prompt that gives you the option to revise your reply before it's published if it uses language that could be harmful.— Twitter Support (@TwitterSupport) May 5, 2020*

Slowcial media

Twitter's move to extend the time period we use to consider rushed and sometimes abusive replies ties ir with the work of psychologist Daniel Kahneman, whose book *Thinking Fast and Slow* argues we think in t different ways. Fast thinking requires little to no effort and takes place with a minimal degree of control, wh slow thinking is more thoughtful and reflective, and is associated with higher levels of concentration.

It's clear that both ways of thinking might determine what we post on social media. When we follow the impulse to post quickly, we're thinking fast and with less consideration. But when Twitter's algorithm makes us pause to stop and think, it may bring slow thinking into play.

Seeing as slow thinking is responsible for overseeing a person's behaviour, its activation in the sometimes frenzied environment of social media may prevent us from instantly venting our anger via fast thinking – ev if we feel justifiably aggrieved.

Convenience over concentration

Having said all of this, as humans, we do tend to seek the easiest and most economical route to our needs and wants. Therefore, it's possible that we may be reluctant to activate slow thinking – often the case wher we unthinkingly click through terms and conditions prompts.

Whether Twitter's "stop and think" prompt will work may also depend to some extent on how impulsive you are. Impulsiveness is characterised by a tendency to act without thinking too closely about one's actions, a can be measured using an impulsiveness test.

Finally, regardless of any new "stop and think" function on Twitter, other personality factors also drive people's desire to use social media in a toxic way, a behaviour often referred to as trolling. Typically, trolls

show a disregard for any pain or suffering inflicted on other people, which is often characteristic of a psychopathic and sadistic personality types.

So while granting users a second chance to rethink their abusive tweets might reduce online abuse, it's unlikely to be enough. There will still be those who won't take the chance to slow down and reflect, and others who press on with their abusive messages anyway – even after engaging their slow, measured system of thinking.

ABOUT THE AUTHOR

Martin Graff is a senior lecturer in Psychology of Relationships at the University of South Wales.