Full length article

Trolling the trolls: Online forum users constructions of the nature and properties of trolling

Bryn Alexander Coles [a,*], Melanie West [b]

[a] Department of Psychology and Counselling, Newman University, Bartley Green, Birmingham, B32 3NT, United Kingdom
[b] University Centre Shrewsbury/University of Chester, Guildhall (Room 211), Frankwell Quay, SY3 H8Q, United Kingdom

ABSTRACT

'Trolling' refers to a specific type of malicious online behaviour, intended to disrupt interactions, aggravate interactional partners and lure them into fruitless argumentation. However, as with other categories, both 'troll' and 'trolling' may have multiple, inconsistent and incompatible meanings, depending upon the context in which the term is used and the aims of the person using the term.

Drawing data from 14 online fora and newspaper comment threads, this paper explores how online users mobilise and make use of the term 'troll'. Data was analysed from a discursive psychological perspective.

Four repertoires describing trolls were identified in posters online messages: 1) that trolls are easily identifiable, 2) nostalgia, 3) vigilantism and 4) that trolls are nasty. Analysis also revealed that despite repertoire 01, identifying trolls is not a simple and straight-forward task.

Similarly to any other rhetorical category, there are tensions inherent in posters accounts of nature and acceptability of trolling. Neither the category 'troll' nor the action of 'trolling' has a single, fixed meaning. Either action may be presented as desirable or undesirable, depending upon the aims of the poster at the time of posting.

© 2016 Elsevier Ltd. All rights reserved.

1. Background

'Trolling' refers to a specific type on malicious online behaviour, intended to aggravate, annoy or otherwise disrupt online interactions and communication (Binns, 2012; Bishop, 2012a). Trolling is also a topical, important issue attracting an increasing amount of attention in the public eye (Bishop, 2014) Cases of individuals identified as trolls, targeting other high-profile individuals are not only being widely reported in the press, but are also forming the basis of court actions.

Trolling appears to be pervasive throughout online media, having been observed in such diverse locations as online magazines (Binns, 2012), social networking sites (Bishop, 2012a; Cole, 2015), online computer games (Thacker & Griffiths, 2012), online encyclopedia (Shachaf & Hara, 2010), online newspapers (Ruiz et al., 2011) and even on Government e-petition pages (Virkar, 2014).

Trolling can have serious consequences for both the perpetrators and the victims of such behaviours, not only in their online spaces, but also in their daily life (Binns, 2012). The presence of trolls in online spaces may serve to create a hostile online space, unwelcoming to new posters, inhibiting the development of online communities. Alternatively, the consequence of uncivil online behaviour may include the polarisation of opinions and beliefs within that online group, as people move to reject the subject of uncivil discourse, or more damagingly, to accept it as normative. This may result in harsher opinions or judgements being formed, acting as a mechanism for the maintenance of prejudicial attitudes (Anderson, Brossard, Scheufele, Xenos, & Ladwig, 2014). Amongst the more serious consequence of trolling is an increase in the risk of suicidal ideation and self harm amongst the victims of such behaviours (Bauman, Toomey, & Walker, 2013; Hinduja & Patchin, 2010). As such, trolling may be understood not only as an unpleasant behaviour, but as an unethical one which holds the potential to do great harm.

---

* Corresponding author.
E-mail address: a.coles@newman.ac.uk (B.A. Coles).

2. Introduction

2.1. Characterising the troll

A number of academics have offered definitions of 'trolling' alongside attempts to account for trolls behaviours (Hardaker, 2010; Shachaf & Hara, 2010). Trolling may be understood as the posting of subtly or unsubtly offensive messages in order to create offence, start an argument or lure the unwary into pointless debate (Binns, 2012). Trolling may also be understood as unconstructive messages designed to provoke a reaction, to draw targets (and others) into fruitless argument, and to disrupt the avowed purpose of the group gathering (Bishop, 2012a). Trolling may also be understood as repetitive, harmful actions which violate a websites terms of use (Shachaf & Hara, 2010). Trolls following this latter definition should be demonstrably active throughout all sections of the website community which they are attacking, rather than just limiting their activity to a limited subsection of the online space.

As with the off-line world, one of the facilitating features of malicious actions online is anonymity (Shin, 2008; Suler, 2004). Anonymity is thought to provide a facilitating condition for disinhibition, leading in turn to greater self-disclosure, deindividuation and the emergence of counter-normative behaviours (Bishop, 2013b). Similarly with off-line behaviours, this may not necessarily be connected with notions of deindividuation, but rather may be associated with a shift in ones sense of identity and self-salience, and a loss of self monitoring (Suler, 2004).

Suler (2004) argues that factors which may encourage counter-normative online behaviour include dissociative anonymity; invisibility; asynchronicity; solipsistic introjection; dissociative imagination and the minimisation of authority. When such conditions are met, people do and say online what they would not do or say in an offline environment. Taken to an extreme, this 'online disinhibition effect' may be termed 'toxic disinhibition'.

Hardaker (2010) explored users definitions of the phenomenon of trolling, drawing upon an extensive archive of data collected over a nine year period from a single forum. Hardaker (2010) collected instances of posters using the word 'troll' or mobilising various euphemisms, such as making reference to other members of the online group 'living under a bridge'. This investigation revealed that posters definitions of a troll typically contain four characteristics. Those of: deception (hiding ones motivations); aggression (attempting to rile other posters); disruption (disturbing the flow of interaction); success (if the trolls failed to provoke anyone, they were not considered successful).

However, not all trolling may necessarily be considered malicious. Interestingly, some authors have also noted that more positive definitions of trolling may exist. Bishop (2012a) describes 'kudos trolling', where users may post irrelevant information seemingly in good faith. The disruption of online interactions here may be considered an unfortunate and unintended consequence.

Herring, Job-Sluder, Scheckler, and Barab (2002) explored in depth the requirements of successful trolling, adopting a case study approach to describe the stages which a troll moves through. These stages include: outward manifestations of sincerity, laying the 'flame bait', and attempting to provoke others into engaging in futile arguments.

Personality variables have also been noted to play a role in trolling, with some authors noting that trolls may exhibit a 'dark tetrad' of sadism, psychopathy, machiavellianism and narcissism or may otherwise display symptoms of personality disorders (Bishop, 2013b; Buckels, Trapnell, & Paulhus, 2014; Suler, 2004).

Aside from personality characteristics or situational factors contributing to disinhibition, trolls have also been shown to be motivated by circumstantial factors such as boredom, attention seeking, revenge and the perception of their targets or online spaces as sources of entertainment (Shachaf & Hara, 2010).

2.2. Troll management

Strategies suggested for the management of trolls vary, according to the aims and sophistication of the online space in which the trolls are operating. At its simplest, users of online spaces are simply admonished 'do not feed the trolls' (Binns, 2012; Shachaf & Hara, 2010). Deprived of oxygen, these flamers are expected to quickly die down.

An alternative, more involved approach to troll management has been termed 'gamification'. This refers to the use of video game elements in non-gaming contexts (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011). Specifically, online spaces may have elements designed into them which are intended to discourage trolling behaviours. This may include a requirement to 'sign in' to the online space in order to be able to interact — an action which should also reduce the anonymity and deindividuation effects of the individual, as it would render them traceable. Alternatively, 'tokens' may be awarded for good online behaviour. These tokens may in turn allow members of that online space to engage in additional activities, privileges, or to simply 'collect a high score' (Binns, 2012; Bishop, 2012a).

However, requiring individuals to link all their online activities may have an unanticipated downside in inhibiting peoples freedom of expression, where there is a fear that what is said could be taken out of context, be misunderstood or otherwise misrepresented.

Yet another solution to the scourge of trolling may be to depend upon the use of moderators, to control what may be posted online. Moderators may clearly delineate the boundaries of civil behaviour (Binns, 2012; Bishop, 2012a). This may be achieved through the use of a network of moderators, who may work within the online community in order to establish norms of civil discourse which all members of the online space can conform to (Lampe, Zube, Lee, Park, & Johnston, 2014).

This however may rapidly become a labour intensive activity, if a large number of individuals begin posting to the online space, and each comment requires consideration before it can be published. Potential resolutions to this issue include distributing moderation duties amongst a broad selection of the online forum (Lampe et al., 2014), or by automating the detection of malicious posts. Software algorithms may be deployed which are able to monitor information posted online and automatically filter out unwanted online activity (Galán-García, de la Puerta, Gómez, Santos, & Bringas, 2014).

Another attempt to resolve the problem of trolling is legislative in nature. Trolling then becomes a criminal offence, and the rule of law may be exercised in order to prevent it (Bishop, 2013a; Butler, Kift, & Campbell, 2009; The Select Committee on Communications, 2014).

2.3. A discursive psychological approach to trolling

The potential solutions to the problem of trolling cited above all rely upon the notion that trolls are easily identifiable, and universally undesirable. However, such an assumption may be problematised. A number of authors have suggested that both trolls and the act of trolling may be divided into a number of sub-categories, such as 'flame trolls' (Bishop 2013a), 'hater' and 'snert' (Bishop 2012a). Bishop (2012a) Further identifies a category of troll whose actions do not represent an attempt to disrupt the flow of interaction online. Rather, contributions from 'kudos trolls' seem to represent a genuine attempt to contribute to the interaction. That disruption occurs is simply an unfortunate side-effect.

The case of the kudos troll may be taken as an example of the