Brno Studies in English
Volume 39, No. 1, 2013
ISSN 0524-6881
DOI: 10.5817/BSE2013-1-1

Christopher Hopkinson

Trolling in Online Discussions: From Provocation to Community-building

Abstract
This paper focuses on the practice of trolling in online discussions. Working with a corpus taken from the websites of three British newspapers, it examines how users themselves define trolling; comparison with a previous study (Hardaker 2010) suggests that users' definitions of trolling may vary depending on the discussion topic. The paper then presents a qualitative pragmatic analysis of one discussion which was attacked by trolls. After examining how trolls announce their presence and attempt to provoke reactions from core community members, the article then moves on to discuss several salient aspects of the antagonistic facework used during the ensuing 'flame war'. Finally, the article turns to address the social dimension of trolling, outlining how a practice which is generally considered destructive can also paradoxically have constructive effects, helping to build new communities and strengthen existing ones.

Key words
Computer-mediated communication; discussion forum; genre; face; impoliteness; trolling; antagonism; discourse community

Introduction

In this paper I explore a certain type of antagonistic interaction occurring in online discussions, generally known as 'trolling'. This is a form of behaviour through which a participant in a discussion forum deliberately attempts to provoke other participants into angry reactions, thus disrupting communication on the forum and potentially steering it away from its original topic.

Although the newly emerged genre of online discussions has attracted considerable attention from researchers (e.g. Marcoccia 2004; Lewis 2005; Wan-

ner 2008; Angouri and Tseliga 2010; Kleinke 2010; Upadhyay 2010; Neurauter-Kessels 2011; Hopkinson 2012), the phenomenon of trolling has so far remained largely on the sidelines, tending to be discussed in popular media rather than in a research context. One exception is Hardaker (2010); drawing on an extensive corpus, she proposes a working definition of trolling based on four key characteristics: aggression, deception, disruption, and success.

This paper aims to build on Hardaker's work in two ways. Firstly, I take her characterization of trolling as a starting point, comparing it against the data found in a different corpus to reveal how the concept of trolling may vary depending on the topic of discussion. Secondly, I explore several key pragmatic aspects of trolling through an in-depth analysis of one online discussion that was targeted by trolls. I focus on the strategies of antagonistic interaction used by trolls and their opponents, and I discuss the implications of trolling behaviour for the wider discourse community.

1. Online discussions – characteristics of the genre

Since the early 1990s, the rapid development of computer-mediated communication (CMC) has stimulated the development of a wide range of new genres. Santini (e.g. 2007) develops a fluid typology of web genres; some are essentially electronic reproductions of earlier generic antecedents (reproduced/replicated genres), others are adaptations of pre-CMC genres (adapted/variant genres), while others represent more novel responses to the new possibilities offered by the medium (novel/emergent genres). Online discussions are a prototypical example of the latter type. Although some features of online discussions resemble the direct interaction found in face-to-face (FTF) conversation, in other ways the genre represents a fresh development, stimulated by the new possibilities offered by the medium. The following paragraphs briefly focus on two characteristic features of the genre with particular relevance for trolling – the prevalence of conflict-based interaction, and the existence of cohesive discourse communities.

A discourse of conflict

One of the defining aspects of online interaction, compared to FTF interaction, is the physical remoteness of communication. Participants are not in physical proximity; they do not see each other during the interaction. This factor plays an important role in conditioning the tenor of the discourse. The physical distance between participants may potentially have a dehumanizing effect; when involved in antagonistic interaction, it is easier to see one's opponent not as a real human being but as a mere character in a form of game. This in turn may lead to a heightened intensity of antagonism, as some participants feel licensed to behave

towards their opponents with a degree of aggression that they would generally avoid in face-to-face interaction.

This intensity of antagonism is further aggravated by the anonymity of participants in online discussions. Although participants are free to use their own names if they wish, in practice they do so very rarely, preferring virtual identities. Online anonymity appears to lower participants' inhibitions; Hardaker notes that "this anonymity can […] foster a sense of impunity, loss of self-awareness, and a likelihood of acting upon normally inhibited impulses" (Hardaker 2010: 224).

The combination of physical remoteness and anonymity thus creates an ideal environment for the aggressive, disruptive behaviour of trolls to flourish. Some studies, though not focusing specifically on trolling, have nevertheless identified certain typical features of online discussions which correspond closely with the patterns typically associated with trolling behaviour. Angouri and Tseliga (2010) have noted that online discussions are characterized by the rapid escalation of conflicts, as relatively mild disagreements frequently spiral out of control, degenerating into angry exchanges of insults known as 'flame wars'. ('Flaming' is a term commonly used by online discussion participants to refer to the practice of aiming personal insults at other posters.) This pattern represents a shift away from what Kleinke (2010) terms 'propositional disagreement' (targeting what an opponent has said) and towards 'personal disagreement' (targeting the opponent personally via direct attacks on his/her face). Closely related to the pattern described above is the observation by Lewis (2005) that online discussions are frequently characterized by topic decay; the first messages in the discussion usually address the topic directly, however the thematic line of the discussion soon tends to become fragmented as participants become sidetracked into multiple dialogues with each other. All these features clearly fit the typical pattern of trolling behaviour: an initial provocation, an angry response, and ultimately the 'hijacking' of the discussion, which drifts away from its original topic and disintegrates into a series of increasingly intense personal attacks.

Discourse communities

A second defining aspect of online discussions which has significant implications for trolling behaviour is the polylogic nature of the genre. The properties of the medium and the structure of online discussion sites make it possible for participants to engage in many-to-many communication to an extent that is not physically possible in FTF communication; a discussion can potentially involve hundreds of contributors. This in turn has implications for the tenor of the discourse, as the polylogic nature of the interaction stimulates the formation of online communities. Besides offering a space for users to construct and project their own individual identities, online discussion boards also enable participants to project a social identity, aligning themselves with the values of the community and engaging in acts of social bonding. Discussion boards tend to have a 'core'

community whose members share similar opinions and value systems – this can be conceptualized as the in-group. Posters who do not share the core community values, and who view themselves as dissenting voices, represent an out-group. Trolling behaviour represents an extreme case of dissent by out-group members, and (in its initial stages at least) it is generally targeted against the core community as a whole, rather than at specific individuals. As will be seen in Section 5 of this paper, trolling – though ostensibly a destructive, disruptive practice – can in fact have a constructive effect, helping to cement the interpersonal bonds among in-group members. If the community is attacked by a troll, its members may close ranks and mobilize in defence of the community, with members forming ad hoc alliances against the intruder which ultimately serve to strengthen the community's cohesion.

2. Material

This paper is based on two separate data sets. The first is a corpus of online discussions hosted on the websites of four British newspapers (*The Daily Express*, *The Daily Mail*, *The Daily Telegraph*, and *The Guardian*) in August 2011. The discussions all deal with the same topic: the riots that broke out in London and several other English cities in early August 2011. The corpus contains a total of 66 discussions, incorporating a total of 26,547 separate 'posts' (i.e. messages posted by participants). This first data set is used as a basis for the discussion in Section 3, which addresses the definition of trolling; being relatively extensive, it offers some scope for quantitative analysis.

The second data set consists of one single discussion extracted from the wider corpus – a discussion which was targeted and disrupted by trolls. The discussion in question was hosted on the website of *The Daily Express* under the title "Debate: Have police been too soft on the rioters?".[1] It contains 150 messages, totalling approximately 20,000 words. This much smaller data sample enabled a close qualitative analysis to be carried out, examining a single case of trolling behaviour and tracing the progression of the interaction from beginning to end. The advantage of this more in-depth approach is that it enables full consideration to be given to the context and consequences of one particular instance of trolling, thus offering a deeper insight into the pragmatic aspects of this type of behaviour. The results of this analysis are presented in Sections 4 and 5.

3. Towards a definition of trolling

Trolling is a conceptually fuzzy term. Judging from a cursory survey of the use of the word in online discussions, it clearly means different things to different people, and it is often applied indiscriminately to describe various types of negatively evaluated online behaviour. Surveying various definitions of trolling, mainly

taken from the media and popular literature, Hardaker notes that most of these definitions nevertheless share a certain area of common ground, which can be characterized as "the posting of incendiary comments with the intent of provoking others into conflict" (Hardaker 2010: 224). However, she points out that the surveyed definitions are intuitive and not based on the analysis of actual data. She therefore sets out to formulate a more data-driven definition of trolling, based on an analysis of approximately 2,000 user comments about trolling, which were extracted from an extensive initial corpus of online discussions. She arrives at the conclusion that trolling, as perceived by forum users, involves four main interrelated characteristics: aggression, success, disruption, and deception.

The first characteristic – aggression – involves "aggressive, malicious behaviour undertaken with the aim of annoying or goading others into retaliating" (Hardaker 2010: 231). The second characteristic – success – depends on whether or not the troll's provocation elicits the desired angry response. (Such a response is generally known by discussion forum users as 'biting'. The metaphor is drawn from fishing; the troll places bait in the water, and hopes that the fish will bite.) The third characteristic of trolling behaviour – disruption – involves the troll's desire to 'hijack' the discussion, leading to topic decay (Lewis 2005) as the participants are sidetracked away from the original topic to become embroiled in a series of intense personal attacks. The fourth characteristic – deception – is connected with the troll's projection of a false identity for purposes of disrupting the discussion; a troll is thus defined as "a CMC user who constructs the identity of sincerely wishing to be part of the group in question […] but whose real intention(s) is/are to cause disruption or exacerbate conflict for the purposes of their own amusement" (Hardaker 2010: 237).

However, as Hardaker acknowledges, this characterization may not be valid for all types of online discussion. For example, her data was taken from discussion groups about equestrian sports; it is reasonable to assume that other topics of discussion may be associated with different patterns of user behaviour, and that trolling may be perceived differently there. In order to provide a comparison with Hardaker's findings, I processed the corpus described in Section 2 above (i.e. the first data set, consisting of 26,547 separate posts) to identify all occurrences of the lexeme "troll" (including inflections) plus all derivatives and compounds. This yielded a total of 127 occurrences; though not comparable in size to Hardaker's data set, it nevertheless offers some scope for a basic quantitative analysis. These excerpted occurrences were then analyzed to determine what users actually mean when they describe a certain type of behaviour as trolling, or state that an opponent is a troll.

Comparing these results with Hardaker's observations, it is clear that all four elements of her definition also fit my own data to a certain extent. Nevertheless, there does appear to be a significant shift in the meaning of the word. In my data, the most frequent use of "troll" is as a generalized label for any participant who is perceived as an out-group member and who intrudes upon the in-group's discussion – an intrusion which is perceived as provocative by its very nature. (It should

be acknowledged that some posters in my corpus criticize the blanket use of the word "troll" to describe anybody who does not share the views of the majority community, and demand a more precise use of the term – in fact, a use of the term that would be more in accordance with Hardaker's proposed definition. However, such objections are relatively rare in my data.)

This difference between the findings from the two data sets can be explained by the markedly different topics involved, which affect the types of interaction found in the respective discussions. My data set, on the topic of the August 2011 riots, is dominated by a strongly antagonistic binary opposition between the in-group and the out-group; this binary structure underlies almost all of the messages in the corpus, and the majority of messages position their authors (whether explicitly or implicitly) as belonging to one or the other of these groups. The in-group/out-group distinction is drawn along political lines, depending on the political orientation of the newspaper in question. Thus, 44 occurrences of the word "troll" (out of a total 127) involve a collocation with a modifier which describes the political orientation of the out-group – on the one hand "right-wing trolls", "rightie trolls", "Tory trolls" etc., and on the other hand "left-wing trolls", "lefty trolls", "Labour trolls", "socialist trolling", and so on. Given that Hardaker's corpus is taken from discussion groups about equestrian sports, it is understandable that this highly polarized in-group/out-group structure does not apply there.

This prevailing use of the term "troll" in my data – to describe participants who are perceived as out-group members – reflects the expectations that appear to have developed among users of online discussion fora dealing with current affairs topics. Users of these fora generally seem not to expect a genuine debate among proponents of various competing views; instead the expectation is of a community consisting of like-minded people who (at least when discussing emotive topics such as the 2011 riots) come together to jointly vent their anger and frustration, and to assign blame to those whom they consider responsible. Against such an ideologically homogeneous background, the dissenting voice of an outsider strikes a highly discordant note, and such an outsider is usually ostracized and singled out for negative evaluation as an intruder.

At this point, one last observation should be made regarding definitions of trolling. Part of Hardaker's working definition (cited above) states that trolls aim to "cause disruption or exacerbate conflict *for the purposes of their own amusement*" (Hardaker 2010: 237; my italics). Put simply, trolls derive entertainment and enjoyment from their trolling. Of course, this potentially presents problems for the analyst, as it is often not possible reliably and unambiguously to assign intent to a speaker's utterance. Nevertheless, I would argue that motivation is such an essential element of this type of behaviour that it merits inclusion in any definition of trolling, at least in a tentative form. My corpus contains several instances in which self-confessed trolls openly explain their motivation. The following example shows one troll, addressing another out-group member with whom he/she has formed an ad hoc alliance against the core community, describing why he/she enjoys trolling discussions on the *Daily Express* website:

(1) I love winding them up with bits of pedantry. It makes my day go a little faster and I know they're not quite smart enough to realise that if they don't bother rising to the bait that I set, I'll just go away a little frustrated. But no, these morons and bigots and racists will always come out and have their say, they can't let anything lie, and there's always a simple way to make them look and feel a little more stupid.

It should also be mentioned that such entertainment and enjoyment is evidently not restricted to the trolls. Although many members of the core discourse community ignore trolls – refusing to 'bite' in order to deny trolls the satisfaction of success – others willingly become embroiled in flame wars, with these hostile exchanges often consisting of multiple turns.

There are two potential explanations for this behaviour. Firstly, flaming can be seen as form of verbal contest – a competitive activity in which participants engage as a form of game (cf. Chovanec 2006 on competitive verbal interaction in minute-by-minute online sports reports). The behaviour of trolls and their antagonists can be viewed as taking place within a game frame; there are certain parallels with e.g. Labov's (1972) observations on the practice of competitive ritualized insults in urban African American culture.

The second potential explanation for the prevalence of flaming between trolls and core community members is connected with the status of the individual as part of the wider discourse community. On many online discussion boards, combative and aggressive verbal behaviour appears to be perceived as a positive cultural value, enjoying prestige status; a core community member who mounts a robust attack on a troll will frequently be rewarded with praise and declarations of respect from his/her peers. Such behaviour may also serve to entertain other members of the community who, though not participating directly in the exchange, are present in the role of spectators. Posters are of course aware that they are performing to an audience; as Neurauter-Kessels (2011: 193-4) observes, "we are dealing here with an unrestricted public space on these online media sites and users are aware that they are operating in a public place and are faced with a potentially large and anonymous audience attending the speech event". In his study of impoliteness, Culpeper (2011: 234-5) identifies several sources of enjoyment that may be derived from watching this type of behaviour, including the thrill (emotional arousal) of observing impoliteness, a certain voyeuristic pleasure, and aesthetic enjoyment derived from the use of verbal creativity in impoliteness. Flaming thus represents a form of display behaviour, as posters assume the prestigious role of the entertainer and 'play to the gallery' in order to seek esteem from their peers. Trading insults can thus be viewed as a form of 'politic behaviour', defined by Watts as "that behaviour, linguistic and non-linguistic, which the participants construct as being appropriate to the ongoing social action" (Watts 2003: 21).

4. Strategies of antagonistic interaction – baiting, biting, flaming

Having discussed the characteristics of trolling behaviour on a general level, Sections 4 and 5 now narrow the scope of the study, presenting an analysis of one particular discussion that was attacked by trolls. The following analysis is based on the second of the data sets described above (i.e. a single discussion containing 150 messages and totalling approximately 20,000 words). In this section, I begin by examining how trolls announce their presence on the forum and attempt to provoke a reaction from core community members (to use the fishing metaphor introduced above, this initial move is 'baiting', while the reaction is 'biting'). I then move on to discuss some salient aspects of the antagonistic facework used during the ensuing interaction (the 'flaming' stage).

At this point, a brief outline of the wider socio-political context of interaction is necessary in order to clarify the attitudes held by the participants in the discussions. The discussion analyzed here, entitled "Debate: Have police been too soft on the rioters?", is a response to the riots that affected London and several other British cities in August 2011. The disorder began when a peaceful protest against the shooting of a man by police in London developed into a riot, with missiles thrown at police officers. Rioting, accompanied by arson and widespread looting, quickly spread to other areas of London and beyond. The most serious disorder lasted from Saturday 6 August until Thursday 11 August, by which time it had claimed 5 lives and left many people seriously injured, as well as costing an estimated £200 million in damage. The discussion analyzed here was hosted on the website of the *Daily Express*, a middle-market national tabloid newspaper with a robustly right-of-centre political orientation. The core community of the discussion forum – the in-group – fits this political profile. The primary out-group against which the in-group directs its anger consists of the direct perpetrators (i.e. the rioters and looters), who are repeatedly described by in-group members as "vermin", "scum", and the like. However, in addition to this primary out-group, community members also seek to blame other groups for the events. These groups, though not directly involved in the rioting and looting on the streets, are viewed as being responsible for the situation which led to the rioting; they can be described as the secondary out-group. The secondary out-group is constructed along ideological lines. In-group members frequently assign blame for the riots to left-wingers; they see the establishment (the media and political elites) as being dominated by left-liberal ideology, political correctness, a decline in moral standards, welfare dependency, immigration and multiculturalism. This viewpoint is expressed by one poster as follows:

(2) THE REAL CULPRITS ARE CHILDREN'S RIGHTS PANELS, HUMAN RIGHTS ACTIVISTS, SOFT JUDGES, SOFTER PAROLE BOARDS, pc looneys AND A SUCESSION OF WEAK-KNEED GOVERNMENTS.
OH I FORGOT TO MENTION THE LEFT WING LOONEY BBC

Crucially, this in-group/out-group axis also underpins the antagonism among the discussion participants. Dissenting voices which contradict the prevailing in-group ideology are automatically assigned by the core community to the secondary out-group. This secondary out-group thus provides a ready-made, pre-constructed category for the conceptualization of antagonists' social identity.

Baiting and biting

The opening move in a trolling attack is the initial provocation. The troll makes his/her first move by posting a provocative message, which acts as 'bait'. In the discussion analyzed here, the first appearance of a troll on the scene comes in the fifth message (out of 150), which is posted by a user calling him/herself 'SBWhy':

(3) These are mainly disenfranchised young people who need care and attention. Sending them to prison will do no good as they will be the same when they come out, and heavy-handed policing tactics will only encourage further disobediance [sic]. I say, let them wear themselves out and then get the main perpetrators in for some counselling in order to understand their feelings and motives.

This post, with its emphasis on compromise and trying to understand the rioters, displays a viewpoint that is diametrically opposed to the prevailing worldview expressed by core community members (not only in this particular discussion, but in numerous other threads on similar topics hosted on the *Daily Express* website). Whereas the prevailing opinion of the core community is that the police were far too soft on the rioters and should have used extreme force, the poster SBWhy expresses a negative evaluation of such an approach in the phrase "heavy-handed policing". Additionally, several lexical items implying a degree of sympathy for the rioters ("disenfranchised", "need care and attention", "counselling", "feelings and motives") provocatively activate a long-running script which holds that criminals are treated too leniently by a politically correct justice system which prioritizes the human rights of perpetrators over those of victims. This script is frequently present in various discussions hosted on the *Daily Express* website, and it appears to be an important element in the in-group's worldview. SBWhy's post thus immediately marks its author out as a member of the secondary out-group. As has been noted above, such a dissenting message is likely to be perceived by core community members as an intrusion, and it is likely to invite an antagonistic response.

A second case of provocative 'baiting' can be seen in the following message, posted by another troll, who joins the discussion writing under the user name 'FoxtrotDelta':

(4) Ultimately, I think there's only one way to get out of this and that is to have an amnesty. Tell these kids that if they bring everything back by the end of the week, they won't be arrested. Everyone can get a bit carried away after a shandy and sherbert fountain combo – you wake up in the morning regretting what went on the night before – give them a chance to undo that hurt, that's what I say.

The message appears to parody the core community's expectations of out-group members' worldview by presenting an exaggeratedly naive and lenient view of the perpetrators. The rioters are characterized as mere harmless children who became hyperactive after consuming sugary drinks and confectionery (though an alternative reading is also invited here, as "shandy" and "sherbet" are slang terms for beer and cocaine respectively), and their violent and destructive actions are trivialized by the downtoner in "a bit carried away".

My approach to Ex. 3 and 4 is based on a trio of crucial assumptions: that both posts are deliberately calculated to provoke in-group members, that they involve a degree of parody or exaggeration, and that they therefore do not (necessarily) represent an entirely sincere expression of the posters' own opinions. Clearly, the problem for the analyst in such cases is that it is often not possible reliably and unambiguously to assign intent to a speaker's utterance, so any judgement on the sincerity or insincerity of an utterance is potentially problematic. Nevertheless, this notion of sincerity lies at the very core of trolling (cf. Hardaker's contrast between sincerity and deception, discussed above), and it cannot be easily sidestepped. Moreover, I would argue that it is often possible to make a reasonable, context-informed judgement on the likely intent of an utterance, even in the absence of incontrovertible linguistic evidence. The resulting judgement is interpretative rather than analytical, but it does not necessarily lack validity. In the case of Ex. 3 and 4, the leniency of the views expressed by the two trolls appears to be extreme, going some way beyond the views expressed by the large majority of posters on the left-leaning *Guardian* website (a natural home for those who are viewed as secondary out-group members by the *Daily Express* core community). On this basis, it seems a plausible assumption that, although Ex. 3 and 4 may to some degree reflect their posters' actual views, they are in fact parodying in-group members' expectations of out-group views, offering a 'heightened' version of those views in order to encourage in-group members to 'bite'.

Indeed, the message posted by SBWhy (Ex. 3) soon generates three angry replies from in-group members. The replies consist of brief, high-intensity negative evaluations, typically followed by a second rhetorical move in which the poster expands on the evaluation and attempts to add persuasive force to it. Here, and in subsequent examples, contextual glosses and explanations are added by me in square brackets:

(5) PC RUBBISH. [PC = politically correct]
Its BECAUSE of this "they need understanding" attitude that we are in this position. The answer is to hit them HARD.

RUBBISH, RUBBISH – COMPLETE AND UTTER RUBISH [sic]
… They are nothing but common criminals.

DISENFRANCHISED UTTER B*LLOCKS …
LOCK THESE VERMIN UP FOR GOOD, WORK THEM TOI [sic]
DEATH TEACH THEM TO SPEAK PROPER ENGLISH …

Flaming

Such an opening exchange – the troll's initial post and the in-group member's angry reaction – may develop into a chain of mutually antagonistic responses ('flaming') which frequently escalate in intensity to become a 'flame war'. The dynamics of antagonistic facework are complex, and space constraints do not allow for an in-depth treatment (for a more detailed account of antagonistic facework strategies in online discussions generally, not only in trolling, see Hopkinson 2012). Here I will focus on two salient points: the nature of face attacks carried out by the antagonists, and the adoption by trolls of various personas to achieve a range of face management goals.

Face attacks

When analyzing the nature of face attacks, I draw mainly on Spencer-Oatey's (e.g. 2000, 2002, 2007) model of face as consisting of three components: quality face, social identity face, and relational face. In my data, attacks on opponents' face are targeted primarily against quality face and relational face; social identity face (i.e. a person's membership of a social, ethnic, professional etc. group) was not found to be a significant target, though analysis of a larger corpus (see Hopkinson 2012) reveals examples of this too.

The notion of quality face concerns the individual's self-esteem, arising from his/her claim to be a possessor of positive personal qualities (competence, abilities, appearance, etc.) on whose basis he/she is favourably evaluated by others (Spencer-Oatey 2002: 540). Attacks targeting quality face are typically directed against the opponent's intelligence or mental capacity:

(6) Tell us, when are you going to get out of the trees and stop dragging your knuckles on the floor?

or they may express mock-concern for the opponent's mental health or emotional well-being; the opponent is thus belittled and characterized as deserving of pity:

(7) You twist other peoples words then claim to be upset by them. It seems to me that you really do have a serious problem which you need help with.

However, although attacks on quality face are most prevalent in my data, another layer of antagonism is often added by attacking the opponent's relational face. This aspect of face concerns the individual's status as a participant in the given interaction, including what Spencer-Oatey terms "role rights and obligations" (2007: 647). Attacks on this aspect of face represent a breach of sociality rights, defined as "fundamental personal/social *entitlements* that individuals effectively claim for themselves in their interactions with others […]. Sociality rights […] are concerned with personal/social expectancies, and reflect people's concerns over fairness, consideration, social inclusion/exclusion and so on" (Spencer-Oatey 2000: 14). A denial of an individual's sociality rights thus represents an attack on that person's relational face. This occurs particularly if that individual is ridiculed (for example by means of sarcasm) or if his/her words or views are deliberately misrepresented or distorted. The distortion of an opponent's views is an important strategy in flaming behaviour. Instead of engaging directly with the opponent's actual words, the speaker first constructs a particular opinion, then attributes that opinion to his/her opponent, and finally attacks that opinion. This usually involves a misrepresentation of the opponent's views, which are often exaggerated or simplified in order to make them more vulnerable to negative evaluation. The purpose of this strategy is to construct an easy target: it is easier to attack a 'straw man' (as this type of misrepresentation is commonly known) than to engage with a complex, nuanced view. In my data, the strategy is enacted by up-scaling the force of the utterance (Martin and White 2005: 140 ff.). This is done by adding various means of intensification and quantification (e.g. completely, always/never, everything/nothing, all, huge, vast, tiny, etc.). This up-scaling distorts the opponent's views by making them seem more categorical or simplistic, thus making them easier to attack:

(8) You clearly believe that **anyone** who is a 'real' English person couldn't **possibly** commit these terrible crimes, and therefore **anyone** who does is **clearly** an illegal immigrant. [emphasis added]

As was noted above, such misrepresentations violate the individual's sociality rights – the desire and expectation to be treated with fairness in the interaction. If one posts a comment on a discussion board, one might reasonably expect the views expressed in that comment to be attacked, as that is a fair and legitimate behaviour in this type of discourse. However, to be attacked for words which one never actually said, which were 'put into one's mouth', is likely to be perceived as unfair, and thus as an attack on one's relational face.

Trolling personas

Trolls in my data adopt various personas during the course of their interaction, enabling them to achieve certain interactional goals. The concept of 'persona' recalls the notions of sincerity and insincerity introduced above when discussing the trolls' opening moves. It is possible to identify two distinct modes of verbal behaviour in the data – sincere and insincere. When posters are behaving insincerely, their expressed views and projected identity are in some way inconsistent with their real views or identity; they are playing a role, donning a mask. This inconsistency recalls the notion of 'mismatch' which is central in Culpeper's (2011) approach to impoliteness.

The adoption of alternative personas takes three main forms in my data, each of which serves a range of strategic purposes enabling trolls to achieve certain face management goals.

Firstly, trolls exaggerate for parodic effect, adopting a persona which meets the in-group's expectations of out-group members. This strategy has been discussed above with relation to the trolls' opening moves, in which they express views of excessive leniency in order to provoke an angry reaction. This 'heightened' persona serves to increase the probability that an in-group member will 'bite'.

Secondly, trolls frequently adopt an ironic, sarcastic persona, expressing mock-respect and admiration for their opponents ("I AM IMPRESSED WITH MR BULL…") or behaving with mock-politeness. This persona serves primarily to ridicule opponents, attacking both quality face and relational face.

Thirdly, trolls may adopt the persona of a naïve, guileless innocent. Wearing this mask, the troll deliberately misinterprets the conversational implicatures contained in the opponent's words, pretending not to understand the intended meaning of the utterance and often interpreting it literally. Thus, for example, trolls may interpret aggressive rhetorical questions as if they were genuine questions, and provide a pseudo-genuine answer. In the following example, the troll (9b) responds to a quality face attack from an in-group member (9a), who implies that the troll possesses ape-like qualities:

(9a) Tell us, when are you going to get out of the trees and stop dragging your knuckles on the floor?

(9b) GIVE ME SOME CREDIT...
Come on now. Firstly, you must be impressed that I have rigged, not just a computer, but internet access as well, up into the branches of this tree. Surely, for a tree dwelling ape, this is something of a stunning achievement, is it not? Secondly – how long do you think my arms are? So I am up in the branches of a tree and yet my knuckles are still dragging on the floor? Surely, if you applied even a semblance of logic, my getting out of the trees would result in me dragging my knuckles on the floor, not put an end to it?

This strategy serves as an effective means of face defence. By refusing to acknowledge the conceptual mechanism on which the attack is based, the troll subverts the attack, taking the sting out of it.

In addition to face attack and face defence (outlined above), these alternative personas also perform a third function: they enable trolls to achieve pre-emptive face preservation by helping to minimize the potential impact of any future face attacks against them. The adoption of an ironic voice generally projects a jocular, facetious persona which signals to other participants that the poster remains distanced from the strong emotions that his/her posts may provoke, and that he/she has not invested too much emotional capital in the interaction. The benefit of this emotional detachment is that it reduces the poster's emotional exposure to the pervasive antagonism of the discourse, shielding him/her from the impact of potential face attacks and thus pre-emptively lowering the risk of future face loss. If one demonstratively does not take the interaction seriously, that necessarily implies that any future face attacks by opponents will not be taken seriously either.

In summary, the 'insincere' personas outlined above all involve playing a role, donning a mask, striking a pose for rhetorical effect. (The theatrical metaphors which naturally spring to mind underline the crucially important fact that the genre of online discussions is not a private conversation, but a spectator event.) The adoption of alternative personas also represents one of the main differences between the trolls and the in-group members in my data. The trolls oscillate between sincere and insincere modes of behaviour; their identities within the interaction are fluid, flexible, and can be adapted to serve various strategic purposes. By contrast, the in-group members overwhelmingly maintain a consistent, 'sincere' identity throughout the interaction. Future research on a larger corpus will reveal whether this is a general tendency or a mere anomaly of the data.

5. Trolling as a socially constructive practice: emergent networks and community-building

Above, when discussing means of provocation and antagonistic facework in trolling, I have focused mainly on the dyadic interaction of individuals with other individuals. However, internet discussion is of course a fundamentally social practice. In this final section, I broaden my focus to address this social dimension of trolling, outlining how the practice can have the effect of building new communities and strengthening existing ones.

When examining the effects of trolling behaviour on the discourse community as a whole, I base my account on Watts's notion of emergent networks, which represents an extension of Milroy's (1980) social network theory. Watts states that "socio-communicative verbal interaction entails the establishment, reestablishment and reproduction of social links between the interactants, which emerge during the interaction. It is these networks of social links set up during ongoing verbal interaction that I wish to call *emergent networks*." (Watts 2003: 154) Watts

distinguishes between emergent networks and latent networks; the latter are pre-established networks which are the products of historical practice. If emergent networks are constructed recurrently – for example among regular participants on a discussion forum – they will gradually solidify into latent networks, as the participants become familiar with each other's virtual selves. Within an emergent network, Watts distinguishes between unidirectional links (when one participant addresses another), ambidirectional links (when two participants address each other), and multidirectional links (involving more than two participants).

In this section I will examine two types of emergent networks created during the interaction. The first involves networks among antagonists, while the second involves networks among fellow members of the same community.

The discussion analyzed here consists of 150 separate posts and involves 28 participants, of whom 25 can be identified as core community members and 3 as trolls (out-group members). This level of participation produces a complex set of networks, which can most clearly be illustrated in schematic form. In the two diagrams given below, the interaction in the forum proceeds chronologically along the horizontal axis from left to right, while the individual posters are listed along the vertical axis. Each post which generates a response from another participant is plotted with a dot, and it is linked to the response by a line. Only those posts which generate these ambidirectional links are plotted on the diagram; isolated posts which generate no response are omitted. Each post is assigned a number based on its order of occurrence; this scale is represented on the horizontal axis, along with the time of posting. Only the first 51 posts are represented in the diagrams, covering a period from around 10 a.m. to around 6 p.m. on Wednesday 10 August 2011. The participants (each identified by a letter; see the key to Figure 1) are grouped into two communities, with the three trolls clustered at the top of the diagram and the core community members at the bottom. These two opposed communities are graphically separated by a grey band which is analogous to a 'no man's land' between two opposing armies; any lines crossing this grey band represent antagonistic responses to a previous post by a member of the opposing community. A graphic representation of this type cannot capture the full complexity of the interaction; its aim is merely to illustrate as clearly as possible how particular networks emerge and develop, and how they fit into the context of the interaction as a whole.

The first diagram shows the development of antagonistic exchanges between in-group and out-group members. These generally form ambidirectional networks, with two posters participating in a chain of responses which typically escalate in intensity to create a 'flame war'. The diagram in Figure 1 shows two antagonistic exchanges, highlighted by means of thicker connecting lines.

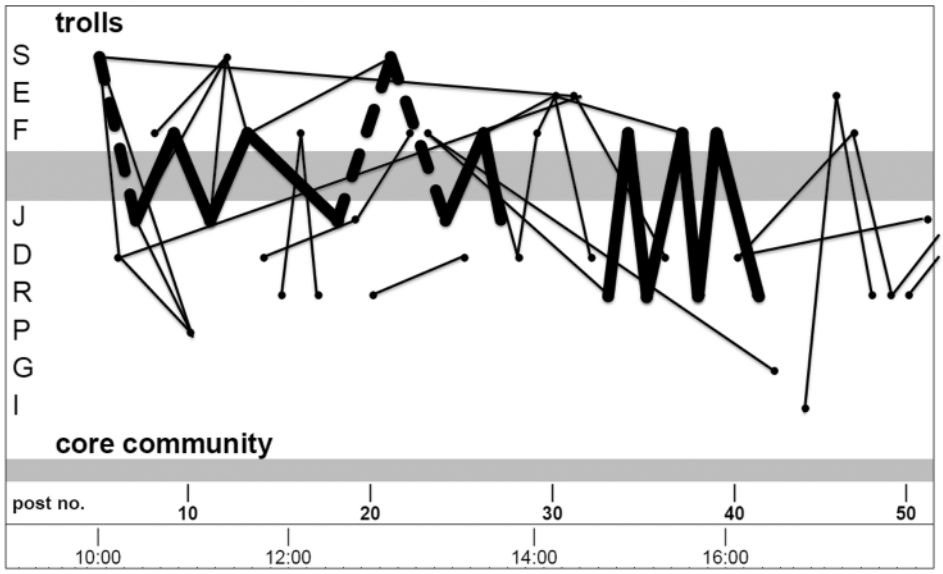

Figure 1. Antagonistic exchanges
(Abbreviations of participants: S = 'SBWhy', E = 'Europhile', F = 'FoxtrotDelta', J = 'John_Bull', D = 'Dragon', R = 'ramblingrose', P = 'REALENGLISHPATRIOT', G = 'GUTLESCHNAPERZAPPER', I = 'incognitono1')

The exchange represented by the thicker lines at the right of the diagram is a flame war between the core community member ramblingrose (abbreviated as 'R') and the troll FoxtrotDelta ('F'), in which ramblingrose has the last word. This is a simple ambidirectional exchange, which proceeds without any contribution from any other participant; although the exchange occurs in a public forum, it resembles a private conversation.

However, the exchange represented at the left of the diagram is somewhat more complex, and involves elements of a multidirectional network. This interaction begins when the troll SBWhy ('S') posts the first provocative message of the discussion (post no. 5, reproduced as Ex. 3 above). This post generates three hostile responses from core community members, one of whom is John_Bull ('J'); his response (post no. 7) is represented by the thick broken line at the far left of the diagram. However, it is not SBWhy who responds to post no. 7; instead it is his/her fellow out-group member FoxtrotDelta (the most active of the trolls in the analyzed discussion). An exchange then develops between John_Bull and FoxtrotDelta. Although this exchange is mainly ambidirectional, at one point (post no. 21) the original troll SBWhy steps back into the discussion to support FoxtrotDelta by responding to John_Bull; this shift in participation is represented by the adjacent pair of broken lines. After receiving this support from SBWhy, FoxtrotDelta then re-assumes his/her original place in the exchange. Drawing

a sporting analogy, the two trolls can be likened to partners in a tennis doubles match, both participating in the same rally. Though this multidirectional network is primarily antagonistic, it is also overlaid with supportive links, as the two out-group members offer each other mutual assistance.

A more extensive picture of supportive behaviour within the interaction is shown in Figure 2, which highlights supportive networks that emerge among members of both communities during the course of the discussion. Participants on both sides explicitly express support for posters whom they consider to be like-minded. It is through this process that two distinct communities crystallize. Although trolling may at first sight appear to be a fundamentally destructive type of behaviour, paradoxically it can also play a constructive role, mobilizing participants to support fellow community members and thus strengthening the community's internal bonds.

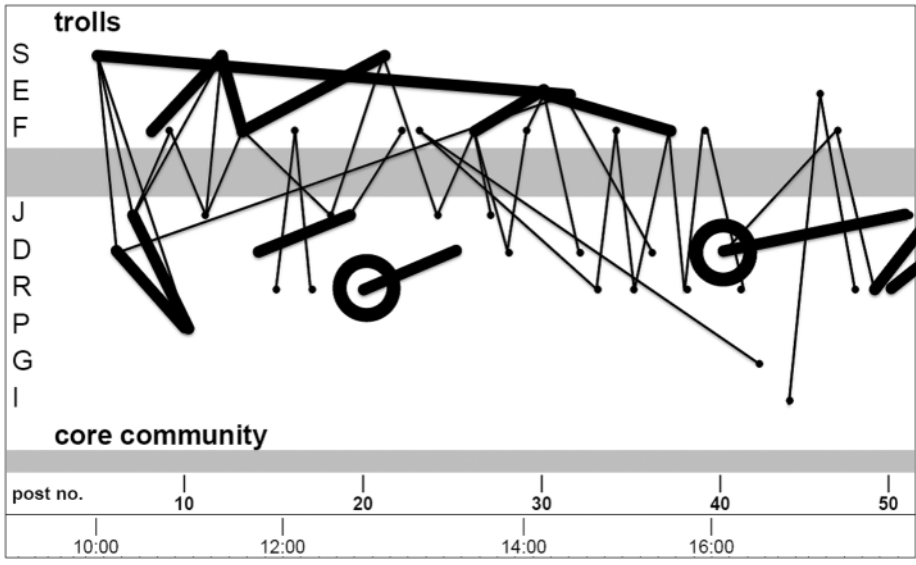

Figure 2. Supportive networks
(The two circles represent appeals to the core community; see the discussion of Ex. 12.)

The thick lines in Figure 2 represent supportive responses to fellow community members. Explicit support within the core community – the in-group – is generally expressed through simple statements of agreement, e.g. "SPOT ON JOHN BULL", "Couldn't agree with John_Bull more!", "I absolutely agree with that comment", "Absolutely right", and so on. Similar statements are also found among out-group members; for example, the troll Europhile ('E', post no. 30) responds to FoxtrotDelta (post no. 26) by writing "You have far too much grey

matter to be on here" (i.e. on this particular discussion board). FoxtrotDelta then reciprocates with the following post:

(10) To Mr Europhile – Everyone needs a summer holiday, I thought I might spend mine here. It's a bit like the Costa Del Sol, only without the volume in the early hours of the morning or the sunburn. It is approximately as likely to give you the sh!ts though....

The jocular tone of this post is typical of the supportive exchanges among the three trolls in the analyzed discussion. It is used by these three participants as a signal that they share the same intention – i.e. to provoke in-group members for purposes of amusement – and that they are therefore 'on the same wavelength'. This jocularity can also be seen, for example, in the response by Europhile (post no. 31) to SBWhy (post no. 5), which involves ironic mock-admonishment:

(11) You should not really be on this forum spouting about disenfranchised kids you know, you will give Mr Bully a heart attack, and that would be a shame as he and his kind are a source of constant entertainment with their, in my opinion, vile, racist, nazi, facist [sic], pig ignorant rhetoric.

Besides expressing support for a like-minded poster, Europhile in Ex. 11 also performs a face attack against the in-group member John_Bull, who is encoded in the third person. The use of third person forms can be viewed as a means of aggravating the face attack on the opponent; although the negative evaluation is ostensibly not directed at the opponent, the opponent can of course read it (and presumably is intended to do so). This strategy is analogous to a type of behaviour that may occur in a face-to-face group conversation, namely turning one's back on one of the group members (either literally, or simply ignoring that person's presence) and then openly directing criticism of that excluded individual to other group members while that individual is still standing there. Culpeper (2011: 136) lists 'turning one's back on someone' as a conventionalized non-verbal impoliteness behaviour, and its potential to damage the opponent's face is clear.

In-group members also sometimes express negative evaluation of trolls via third person encodings. Figure 2 shows two posts which do this (marked with circles); the posts are not directed at specific members of the in-group, but instead are addressed to the community in its entirety. Such a strategy essentially constitutes a generalized appeal for support from one's peers. As Figure 2 shows, in both cases support is forthcoming, as fellow community members step in and perform face-enhancing acts to their colleague's benefit. The strategy is therefore double-edged, serving not only to aggravate a face attack against an opponent, but also helping to cement alliances among group members. The following example shows Dragon ('D', post no. 40) turning to address the in-group as a whole when criticizing an opponent (FoxtrotDelta) who has made pedantic comments about another poster's spelling:

(12a) DON'T YOU JUST HATE .....
..... these anally retentive obsessive types who seem to think that infantile points scoring over grammar or spelling makes them look clever. They just dont see that it actually makes them look like petty minded obsessive idiots.

This generates a supportive response from John_Bull (post no. 51):

(12b) You always get posters like this in the school holidays. They are just bored kiddies with nothing productive to do.

This strategy of address is reminiscent of the theatrical practice of asides – a dramatic technique in which a character in a play temporarily steps out of the action and speaks directly to the audience. The theatrical metaphor is apposite here; as has been noted earlier, one of the most salient characteristics of this genre is its status as a spectator event, involving not only the relatively few individuals who are actively participating in the discussion, but also the (more numerous) mass of 'lurkers' who are following the discussion yet do not feel the need to become involved in it directly.

In summary, the practice of trolling helps to create both antagonistic and supportive networks. Far from being a purely destructive type of behaviour, it plays a significant role in building communities and strengthening bonds among community members. This process is particularly important given that one of the main functions of online discussion fora is to enable their participants to define their own social identity as an in-group. If out-group members were entirely absent from the forum, in-group members would certainly still be able to develop a shared identity by defining themselves against that out-group and by supporting each other's negative evaluations of it. However, if out-group members are actually present in the discussion (in the form of trolls, for example), then the in-group has the opportunity to become actively engaged in verbal conflict. In-group members may thus join forces and form ad hoc alliances, which ultimately help to cement the community's internal bonds. Over time, such ad hoc alliances may develop into latent networks, as participants become familiar with each other. Of course, the same applies not only to in-group members, but also to trolls themselves, who may decide to join forces and 'hunt in packs' when attacking a forum.

6. Conclusions

Although trolling is a somewhat vague and fuzzy concept, at the heart of this type of online behaviour is the notion of deliberate provocation for purposes of personal amusement. My data suggests that users' definitions of trolling may vary depending on the topic of the discussion. In the corpus analyzed for this study – which is taken from intensely polarized discussions on a controversial current

affairs topic – any out-group member daring to intrude on the in-group's discussion may be singled out as a troll.

Trolls' opening moves – through which they first announce their presence to other participants in the discussion – can be seen as dropping 'bait' into the water and waiting for the in-group members to 'bite'. This 'baiting' commonly involves exaggeration; trolls may parody in-group members' expectations of out-group views, offering a heightened version of those views in order to encourage hostile reactions. If this strategy is successful, a 'flame war' may ensue, involving attacks not only against opponents' quality face (by impugning their intelligence and other positive qualities), but also against their relational face (by distorting their views and thus violating their sociality rights, i.e. the expectation that they will be treated fairly). Trolls adopt a range of fluid personas, oscillating between the sincere expression of their views and various forms of role-playing, particularly involving strategies based on irony. These 'insincere' strategies perform multiple face management functions, enabling trolls to attack opponents' face, defend their own face, and pre-emptively preserve their face.

Given that internet discussion is a social practice, trolling naturally has consequences for the discourse community as a whole. Both antagonistic and supportive networks can be created as a result of trolling behaviour; ad hoc alliances emerge, which may eventually develop into latent social networks. Trolling is thus – paradoxically – not only a destructive form of behaviour; it also has the potential to be profoundly constructive, stimulating community-building and strengthening group identities.

Notes

1 Source: <http://www.express.co.uk/posts/view/264151/DEBATE-Have-police-been-too-soft-on-the-rioters-> (10 August 2011). Retrieved 30.05.2012.

Acknowledgements

This article is an output of the ESF-funded project "Posílení rozvoje Centra výzkumu odborného jazyka angličtiny a němčiny na Filozofické fakultě Ostravské univerzity" (OP Vzdělávání pro konkurenceschopnost, reg. no. CZ.1.07/2.3.00/20.0222).

References

Angouri, Jo and Tseliga, Theodora (2010) "you HAVE NO IDEA WHAT YOU ARE TALKING ABOUT!" From e-disagreement to e-impoliteness in two online fora. *Journal of Politeness Research, 6*, 57–82.

Chovanec, Jan (2006) "Competitive verbal interaction in online minute-by-minute match reports". *Brno Studies in English, 32*, 23–35.
Culpeper, Jonathan (2011) *Impoliteness: Using Language to Cause Offence.* Cambridge: Cambridge University Press.
Hardaker, Claire (2010) "Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions". *Journal of Politeness Research, 6*, 215–242.
Hopkinson, Christopher (2012) "Antagonistic facework in online discussion fora". In: Hopkinson, Christopher, Renáta Tomášková and Gabriela Zapletalová (eds.) *The Interpersonal Function of Language Across Genres and Discourse Domains.* Ostrava: University of Ostrava, 113–153.
Kleinke, Sonja (2010) "Interactive aspects of computer-mediated communication". In: Tanskanen, Sanna-Kaisa, Marja-Liisa Helasvuo, Marjut Johansson and Mia Raitaniemi (eds.) *Discourses in Interaction.* Amsterdam: John Benjamins, 195–222.
Labov, William (1972) *Language in the Inner City: Studies in the Black English Vernacular.* Philadelphia: University of Pennsylvania Press.
Lewis, Diana (2005) "Arguing in English and French asynchronous online discussion". *Journal of Pragmatics 37*, 1801–1818.
Marcoccia, Michel (2004) "On-line polylogues: Conversation structure and participation framework in internet newsgroups". *Journal of Pragmatics 36*, 115–145.
Martin, James R. and Peter R. R. White (2005) *The Language of Evaluation: Appraisal in English.* Palgrave Macmillan: London and New York.
Milroy, Lesley (1980) *Language and Social Networks.* Oxford: Blackwell.
Neurauter-Kessels, Manuela (2011) "Im/polite reader responses on British online news sites". *Journal of Politeness Research 7*, 187–214.
Santini, Marina (2007) "Characterizing genres of web pages: Genre hybridism and individualization". In: *Proceedings of the 40th Hawaii International Conference on System Sciences*. <http://www.nltg.brighton.ac.uk/home/Marina.Santini/HICSS_07.pdf>. Retrieved 17.09.2012.
Spencer-Oatey, Helen (2000) "Rapport management: A framework for analysis". In: Spencer-Oatey, Helen (ed.) *Culturally Speaking: Managing Rapport Through Talk Across Cultures*. London and New York: Continuum, 11–46.
Spencer-Oatey, Helen (2002) "Managing interpersonal rapport: Using rapport sensitive incidents to explore the motivational concerns underlying the management of relations". *Journal of Pragmatics 34* (5), 529–545.
Spencer-Oatey, Helen (2007) "Theories of identity and the analysis of face". *Journal of Pragmatics 39* (4), 639–656.
Upadhyay, Shiv R. (2010) "Identity and impoliteness in computer-mediated reader responses". *Journal of Politeness Research 6*, 105–127.
Wanner, Anja (2008) "Creating comfort zones of orality in online discussion forums". In: Magnan, Sally Sieloff (ed.) *Mediating Discourse Online*. Amsterdam: John Benjamins, 125–149.
Watts, Richard J. (2003) *Politeness.* Cambridge: Cambridge University Press.

Christopher Hopkinson is an assistant professor at the Department of English and American studies, University of Ostrava, Czech Republic. His research interests range from translation studies to discourse analysis, with a particular emphasis on computer-mediated communication. His current research focuses on the genre of online discussions and facework strategies in CMC. He has also published studies of the discourse of advertising and commerce. He is the author of the monograph *Shifts of Explicitness in Translation* (2008).

Address: Christopher Hopkinson, Ph.D., Department of English and American Studies, Faculty of Arts, University of Ostrava, Czech Republic. [email: christopher.hopkinson@osu.cz]