Shachaf, P., & Hara, N. (2010). Beyond vandalism: Wikipedia trolls. *Journal of Information Science, 36*(3), 357-370.

Beyond vandalism: Wikipedia trolls

Pnina Shachaf and Noriko Hara

School of Library and Information Science, Indiana University, Bloomington, IN, USA

Abstract.
Research on trolls is scarce, but their activities challenge online communities; one of the main challenges of the Wikipedia community is to fight against vandalism and trolls. This study identifies Wikipedia trolls' behaviours and motivations, and compares and contrasts hackers with trolls; it extends our knowledge about this type of vandalism and concludes that Wikipedia trolls are one type of hacker. This study reports that boredom, attention seeking, and revenge motivate trolls; they regard Wikipedia as an entertainment venue, and find pleasure from causing damage to the community and other people. Findings also suggest that trolls' behaviours are characterized as repetitive, intentional, and harmful actions that are undertaken in isolation and under hidden virtual identities, involving violations of Wikipedia policies, and consisting of destructive participation in the community.

Keywords: community; hackers; trolls; vandalism; wiki; Wikipedia

1. Introduction

Wikipedia is an ambitious attempt to create an online encyclopaedia that exemplifies the wisdom of the crowd [1]. Started in 2001, it now has more than 13,000,000 articles in over 260 languages. The largest Wikipedia is the English version, which contains 2,976,299 articles as of 3 August 2009 [2]. As of August 2009, there were 10,203,869 users and at least 146,000 active users who contributed to the English Wikipedia a minimum of five times during one month; there were 1674 sysops (system administrators) in the English Wikipedia alone [2]. Because Wikipedia attracts vandals, one of the major tasks of sysops is to fight vandalism [3].

Many acts of vandalism towards Wikipedia have been attributed to trolls. Wikipedia trolls, like other online trolls, are harmful to online communities. Schwartz [4, p. 3] argued that trolls are part of 'a growing Internet subculture with a fluid morality and a disdain for pretty much everyone else online' and claims that a troll is 'a normal person who does insane things on the internet.' Herring et al. [5, p. 372] referred to trolling as a behaviour that 'entails luring others into pointless and time-consuming discussions'. Donath [6] stated that 'trolling is a game about identity deception, albeit one that is played without the consent of most players' and adds that it can be costly in several ways. In a community with a high deception rate, a new user can be bombarded with angry accusations, and some innocent users may never participate again in such a community.

Despite the fact that trolling is a common online phenomenon, it has rarely been the focus of previous research, with only one exception [5]. There is clearly a need to expand our knowledge and understanding of internet trolls in general and of Wikipedia trolls in particular. Therefore, this study aims to identify the range of troll behaviours and motivations and to examine the similarities and differences between Wikipedia trolls and hackers.

2. Background

Wikipedia's success has attracted considerable media and research attention; it has generated debates among educators and researchers. One of the concerns about Wikipedia is whether this type of free-form encyclopaedia can be considered a reliable source and whether the information it publishes is accurate. A stream of research has focused attention on this question. For example, Lih [7] used the number of edits per article, the number of editors, and press citations as an evaluation matrix to confirm the famous mantra for the open source movement, 'given enough eyeballs, all bugs are shallow' [8]. Giles [9] who tested the accuracy of 42 entries in both the Britannica Online Encyclopaedia and Wikipedia reported similar quality levels. Rosenzweig [10] employed the metaphor of open source to evaluate Wikipedia entries. He compared 25 entries of biographies in Wikipedia with two respected online sources: the online *Encarta* encyclopaedia and the carefully edited, high-quality *American National Biography Online* (ANBO). He found only four factual errors in Wikipedia and concluded that Wikipedia was as accurate as *Encarta* but not as good as ANBO, and covered more biographies than *Encarta* but not as much as ANBO. These findings are consistent with the notion that Wikipedia entries are as accurate as those in traditional encyclopaedias.

Researchers have examined the processes of information quality and article evolution over time. Stvilia et al. [11, 12] used Wikipedia to demonstrate the process of evaluating the quality of information. Viégas, Wattenberg and Dave [13] found that about half of all entries having a mass deletion (i.e. more than 90% of the content was deleted) were repaired within three minutes, while the overall average repair time was 7.7 days.

Scholars have also made efforts to develop tools to expand the functionality of Wikipedia. For example, Viégas and colleagues [13] developed a tool that visualizes data available through the history pages of entries. Through the use of this tool, patterns of how people cooperated and disagreed on specific entries emerged. Holloway et al.[14] also used a visualization tool to map the relationship between categories (co-occurrence of categories) and to identify authors who contributed to the top 10 most active categories.

Others have focused attention on the Wikipedia community and have tried to identify, for example, why people contribute anonymously in cyberspace to a free encyclopaedia, despite the lack of common organizational incentives for knowledge sharing, like monetary reward or name recognition [15]. Bryant et al. [16] reported that newcomers tend to be motivated by correcting entries, thus focusing on editing, while seasoned users are motivated to contribute to the Wikipedia project as a whole, thus their attention is also on the well being of the Wikipedia community. Forte and Bruckman [17], in a follow-up study that included interviews with 22 Wikipedians, revealed that the primary reason that these users contribute to the site is their desire to identify and publish facts about the world; that is similar to the motivations of the scientific community. Poe [18] described how, during the early stages of development, the Wikipedia community experienced 'edit wars' in which authors who had different points of view were engaged in endlessly reverting entries.

2.1. Trolls and hackers

The only empirical study about trolls [5] analyzed their activities in an online discussion forum and described troll behaviours under three definitional criteria: 1) outward manifestations of sincerity, 2) flame bait, and 3) attempts to provoke futile arguments. The troll was successful in manipulating members' ideologies into an intense conflict, which was one of the reasons that the members could not effectively ban the troll. Schwartz [4, p. 3] adds 'trolling will stop only when its audience stops taking trolls seriously'.

While research on trolls is limited, research on hackers and other rough users has attracted significantly more attention. Since these terms have been at times used almost as synonymic [i.e. 4], and because the communities of hackers and trolls may share similar behaviours and motivations, reviewing this literature can be useful here. The meaning of the term 'hacker' has changed over time since the late 1950s [19]. In the 1980s, with the expansion of the hacker community, a 'status hierarchy' among hackers was developed as a result of the use of bulletin board system software [20]. Later, in the 1990s, criminalization of hackers evolved [19, 20], and consequently, Falk [21] introduced three categorizations of hackers: 1) organized crime for financial gain, 2) script kiddie for

self-gratification, and 3) hacktivism for political messages and social justice.

Prior research about hackers' motivations used, for example, the Flow Theory, which explicates intrinsic motivations with clear goals and adequate levels of challenge, to elucidate hackers' motivations [e.g. 22–25]. Rennie and Shore [23] contended that Flow Theory could explain the transformation of a lower level hacking activity to a higher level, while empirical data in Voiskounsky and Smyslova's [25] study shows that Flow Theory does not apply to mediocre hackers. Mulhall [22] compiled a list of hackers' motivations based on his review of the literature and suggested methods for preventing hacking activities. Turgeman-Goldschmidt [24] identified hackers' accounts through interviews with 54 Israeli hackers. She reported that their accounts included: 1) fun and exciting activities, 2) personal curiosity, 3) display of computer talent, 4) economic gains, 5) intangible offense, 6) revenge and 7) ease of execution. Wikipedia trolls may share similar motivations with hackers or behave in similar ways. Other online users who exhibit deviant behaviours in online communities, such as question and answer communities, have been discussed in prior research. For example, Gazan [26] identified the behaviours of the rough user in Answerbag, an online question-answering community. The behaviours he listed included abusive language, excessive contact with administrators, creating sock puppets, and requesting special privileges to form exceptions for the rules and policies. He further attributed the behaviour of these users to narcissistic personality disorder (NPD).

The lack of research on internet trolls or on Wikipedia trolls makes it also difficult to compare them to hackers; their motivations and behaviours may resemble those common among the hackers' communities. Our study addresses this lacuna, identifies the factors that motivate Wikipedia trolls, explores the range of trolls' behaviours, and examines the differences and similarities between trolls and hackers. Specifically the study addresses the following three research questions:

1. What are the behaviours of Wikipedia trolls?
2. What are the trolls' ideologies and motivations?[1]
3. How do trolls differ from hackers? (Q3a – Do Wikipedia trolls represent one (distinct) type of hacker?; and Q3b – Are there any trolls with ideologies (hacktivists)?)

3. Methods

The Wikipedia project involves articles in more than 260 languages; the three largest are the English, German, and Japanese Wikipedias [3].[2] The enormous size of the English Wikipedia community, its organizational structure, extensive policies, and activities, all made it extremely difficult to identify active trolls. Trolling behaviours were not easily tracked on the Japanese Wikipedia either. It was then determined that a small, yet active, Wikipedia community may facilitate the identification of active trolls and may enable familiarity with the most active members of the community of Wikipedians; thus, the Hebrew Wikipedia was chosen.

3.1. The Hebrew Wikipedia

Ordered according to the number of articles, the Hebrew Wikipedia is ranked 28th with 94,975 articles. It is ranked second among 253 languages with 63.22 edits per page, preceded only by the English Wikipedia (80.70 edits per page) [2]. As of 3 August 2009, it had 88,480 users, 7,750,464 edits and 54 sysops. The English Wikipedia, for comparison, had more than 2,976,299 articles, 10,203,869 users, 323,386,901 edits and 1674 sysops. There are 54.6 Hebrew speakers per one Wikipedia article in Hebrew, compared to 206.7 English speakers per one article in English. Finally, the ratio between sysops and users in the Hebrew Wikipedia is 1638.8 users to one sysop and 6,095.5 users to one sysop in the English Wikipedia.

3.2. Data collection

Data collection involved: 1) email interviews with eight sysops; 2) online ethnographic observation

of Wikipedians and trolls on various Wikipedia pages (talk pages of articles, users pages and the other community's talk pages); and 3) interpretive content analysis of Wikipedia pages (user pages of trolls and sysops; Wikipedia policy pages about trolls and vandals; and community discussion pages about trolls and vandals) and their respective talk pages, archives and history pages.

3.2.1. Interviewees sample

The original intent of the study was to collect data through interviews with Wikipedians: both sysops and trolls. Major efforts, early on, were put into identification and recruitment of trolls. However, troll recruitment was a challenging task because they hide their identity, operate anonymously, use pseudonyms or sock puppets, and do not provide contact information on their Wikipedia user pages. Furthermore, trolls are blocked soon after they are identified, and their user pages are deleted or modified. Unsuccessful efforts to recruits trolls through sysops were made, but some sysops argued that a study of trolls should not be conducted because public interest in trolls' activities would only encourage their activities. This approach, to ignore trolls as much as possible in order to limit their activities, is perceived to be the best practice in online communities; it is assumed that if ignored, the troll will leave [5, 6]. For these reasons the study includes data that mainly reflect the sysops' point of view.

Out of 21 sysops that were sent a recruitment letter, 15 replied and agreed to participate by email, and eight provided their replies after only one round of requests (represented as interviewees 1–8) at a response rate of 44%. All the participants in the study were male, and this reflects the fact that most of the sysops on the Hebrew Wikipedia were male (92%). Follow-up questions were sent to some participants during data analysis. Theoretical saturation was achieved at early stages, and the data from later interviews supported and reinforced the early findings.

The email interview included open questions that addressed the three research questions; the interview protocol is presented in Appendix 1. The questions were sent in English and the responders replied either in English (3) or in Hebrew (5), based on their individual preferences. The Hebrew interviews were translated into English by one of the authors before data analysis began. In an effort to maintain the confidentiality of the interviewees, the findings are reported in aggregated terms; we do not report individual names, Wikipedia user names, educational background, user's age, their contribution type or topics, or the history of their individual contributions on Wikipedia.

3.2.2. Selection of troll cases

The study examined four cases of trolls. The interviewees named 11 trolls; Table 1 provides the type and frequencies of troll identification. The authors, for further analysis, selected trolls that were identified by at least two sysops during the interviews and also have been named trolls on the Wikipedia pages. By triangulating these sources, four trolls were chosen to be the focus of individual case analyses. All of the trolls were males, many of whom were also active in other Wikipedia languages (e.g. English, German, Yiddish or Russian).

Table 1
Trolls' identifications

Troll number	Frequency of identification by interviewees	Triangulation with Wikipedia pages	Possessed sysop privileges in the past
1	8	Yes	
2	5	Yes	
3	4	Yes	
4 [a]	2	No	Yes
5	2	Yes	
6	2	No	
7 [a]	2	No	
8 [a]	1	No	
9	1	No	Yes
10	1	No	
11	1	No	

3.3. Data analysis

Data analysis involved interpretive content analyses and case synopses, as well as pattern coding of the interview data. Following Miles and Huberman [27], an individual case synopsis was written for each troll, which included a description of each troll's behaviours and interactions with other Wikipedia users, as well as the period in which the troll was active. The case synopses were written based on the ethnographic observations and on the traces the trolls left on Wikipedia pages. Using pattern-coding techniques, these sections of the data were combined with an analysis of the interview data [27]. Pattern codes were created in a tentative form as the themes were created, and were then tested on the next case synopsis as these themes were added. A comparative analysis among the trolls' case synopses was conducted with the addition of each case synopsis. These analyses served to identify similarities and differences among the individual trolls. Simultaneously, the interview data was analyzed to identify trolls' behaviours and motivators. Through content analyses of the eight interviews, concepts were identified and grouped into categories. The categories from the interview data were then triangulated with the data from the four cases. Similarities and differences between trolls and hackers were identified from the interview data; these and the behaviours and motivators were triangulated with the four troll cases and with existing literature.

3.4. Limitations

Due to the small sample size (four troll cases and eight interviewees) and the fact that data collection was based on the Hebrew Wikipedia, transferability of findings is limited. It is possible that trolling behaviours will be culturally dependent like other user behaviours on Wikipedia [28]. Another limitation is that the study used a restrictive definition of trolls. Many would argue for a wider definition that would consider any act of vandalism on Wikipedia as an act of trolling. In such an event, the range of behaviours and motivations may expand as well. Finally, the interview data only reflect the sysops' points of view, and it is likely that the trolls have different perspectives.

4. Findings and discussion

Before answering the three research questions, the four trolls' cases are briefly described (Table 2).

Table 2
Description of the four trolls

Troll number	**1**	**2**	**3**	**5**
Number of sysops who identified the troll	8	5	4	2
Was blocked in	2004	2006	2006	2006
Areas of contributions	Pornography, porn stars, and Hitler	Math	Harry Potter	Wrestling
Gender (and occupation if known)	Male	Male; College student	Male; College student	Male
Status of the current user page	The entire user page was written by sysops, and the user is blocked	The entire user page was written by sysops, and the user is blocked	The user page was modified by sysops, and the user is blocked	The user page was modified by sysops, and the user is blocked
Deviant behaviour	His behaviour is characterized by argumentative writing style. He was obsessed with porn stars, tricked Google, inserted pornographic photos on main page, and damaged wiki-stress image	His behaviour is characterized by argumentative writing style. He cursed and made aggressive threats towards sysops. He used an offensive signature, and used a sock puppet	His behaviour involved using multiple user names. Each name was created and used for one login and was indicative of the damage that he made while he contributed under a particular user name	His behaviour involved argumentative style and very poor writing skills. He cursed sysops often and initiated personal attacks towards other users. He wrote very aggressive threats against sysops and their family members

Troll 1 created articles and uploaded photos of porn stars against the wishes of the Wikipedia

community. Although the sysops perceived his actions as damaging and requested that he cease this behaviour, he persisted. As a result, most of his actions and contributions were reverted by other users. At this point, when he could no longer contribute biographies and photos of porn stars to Wikipedia articles, he posted the 'censored' information on his Wikipedia user page. He then tried to manipulate Google into retrieving his Wikipedia user page page by inserting a long list of search terms at the top of his page. When he was asked by the sysops to remove this information from his user page, he refused by arguing that he had the right to decide the content of his user page. Most of his interactions with the sysops involved poorly written personal attacks. In fact, troll 1 was an active user on the English Wikipedia where he had created a porn portal and contributed to biographical articles (largely about porn stars). At one point he made an effort to become an administrator on the English Wikipedia but with the intervention of sysops from the Hebrew Wikipedia, he was denied these privileges.

Troll 2 contributed little to Wikipedia articles and was identified as a troll due to his activities at the community level. He had an argument with one of the sysops about the moral legitimacy and subjectivity of blocking users on Wikipedia. The conflict between the two intensified and caused chaos among the Wikipedia community as users took sides. Certain members of Wikipedia believed that troll 2's views were justified, while others supported the sysop's decision to block troll 2. The supporters of the sysop argued that this sysop was a more important contributor to Wikipedia than troll 2 (although the whole basis of troll 2's argument was that it should not matter how much one contributes but rather whether one abides by the rules of the community). Later on, troll 2 was accused of having a dual identity on Wikipedia, which he furiously denied. This conflict again intensified and sides formed for and against the identification of the two users as one troll; they also argued about the consequences of the identification of the two users as one troll. While troll 2's contributions to Wikipedia articles were minimal, his major distraction was creating chaos and disorder among the Wikipedia community.

Troll 3 had many identities on Wikipedia, which garnered exceptionally different attitudes from the community (compared to their attitudes toward other trolls). On two of his accounts, he was widely regarded as a highly contributive member to Wikipedia articles, especially those dealing with subjects like Harry Potter and Pokémon. Despite his contributions through his two main accounts, troll 3 also created additional accounts with usernames that implied what his next vandalistic intentions were. For example, he would create a user name 'deleting leftist user pages' and, logging in with this user name, he deleted user pages he believed to have leftist political views. IP addresses for these accounts were banned (the same IP addresses as those of troll 3's two main accounts). Eventually, as suspicion grew of troll 3's multiple identities, he was confronted in real life by a Wikipedia sysop at a local university campus.

Troll 5 presented the most extreme deviant behaviour out of all four trolls. Most of the articles he edited were related to wrestling. They were often poorly written and included numerous grammatical errors. When one of the sysops asked troll 5 to improve his editing, troll 5 responded with strong anger; then, troll 5's account was blocked, which he aggressively protested. He used vulgar language in all of his comments and physically threatened the sysop who blocked him, and this sysop's family, more than once, with physical violence.

4.1. Trolls' behaviour (what do they do?)

A variety of deviant trolling behaviours have been identified from the data. Behaviours that have been mentioned by at least three interviewees were included, following Miles and Huberman's [27] rule of thumb. Each of the following behaviours was also observed in at least two troll cases:

1. Trolls are engaged in intentional, repetitive, and harmful actions;
2. Their activities largely violate Wikipedia policies;
3. They are active not only on the encyclopaedic part of Wikipedia (e.g. writing and editing articles), but have high interest and destructive involvement within the Wikipedia community (e.g. discussion about policies, and intensive interactions with other users and sysops);

4. They work in isolation under hidden virtual identities.

4.1.1. Repetitive, intentional and harmful actions

The most common behaviour of trolls involves harmful actions to Wikipedia (both at the community and at the encyclopaedic levels). The interviewees repeatedly referred to the notion of damage made to Wikipedia (content or community). They claimed that trolls simply 'cause damage' (interviewee 8) in Wikipedia; they are 'embarked on a quest of damaging…Wikipedia' (interviewee 3). The most pertinent characteristic of this destructive behaviour is its repetitiveness. For example, interviewee 5 contended that a troll is 'a user who repeatedly damages articles,' and interviewee 3 concurred and claimed that trolls are engaged in 'all sorts of repetitive harmful actions.' Furthermore, interviewees repeatedly emphasized the concept of intentionality in trolling behaviours; they are characterized by 'ill intentions' (interviewee 6). The sysops made a clear distinction between trolls who meant to cause damage and users who unintentionally caused damage. And interviewee 8 added: 'Any user can make 'mistakes…[but it will not be characterized as trolling behaviour, because of the] hope for useful contributions'. Accordingly, if someone has 'good intentions' (interviewee 6) behind his actions, the user is not a troll. For example, multiple political opinions, even controversial ones, are expressed on Wikipedia pages, yet these acts will not be perceived as trolling. These acts are identified as being against the neutral point of view (NPOV) policies.

The repetitive, harmful and intentional behaviours characterize all four trolls, the most illustrative cases being trolls 1 and 3. Specifically, this is evident in troll 1's insistence on sharing information in Hebrew about porn stars, and his actions against the Wikipedia community on this matter; and troll 3's multiple user names, including creating a user name that declared the damage that he planned to pursue and then executing his plan by deleting entire articles, specific letters and certain user pages.

4.1.2. Policies violation

Trolls behave in ways that intentionally ignore and violate Wikipedia policies. These violations involve a wide range of activities, some of which are relatively easy for sysops to track and repair, but all of which are unnecessary and time consuming. Violations of policies that characterize trolling behaviours include 'repeatedly inserting text that violates policies, registering derogatory usernames, blanking (removing all content of) articles, renaming articles to random names, and inserting intentionally misleading, wrong, or irrelevant information' (interviewee 3). These actions reduce the accuracy and reliability of the Wikipedia project. The difficulties that a troll creates by 'ignoring policy… and [for example] adding "illegal" porn-related articles' (interviewee 8) are serious. Likewise, the troll 'often creates silly articles (e.g. "Yosi is a very cute kid!"), articles without content or full of nonsensical speech (e.g. "hdgfsakjhgfd")…or add[s] curse words' (interviewee 2). Ignoring Wikipedia policies alone does not make one a troll, but combined with intentionality and repetition, these harmful behaviours constitute trolling. Trolls 1 and 3 engaged in actions that violate Wikipedia policies: troll 3 edited politically controversial topics and deleted articles, letters from words and user pages; and troll 1 insisted on including information against the community's wishes.

4.1.3. Destructive community involvement

Trolls' actions negatively affect the Wikipedia community's well being. 'Trolls operate at the community level of Wikipedia in order to destroy it. They do not necessarily cause damage to the articles' (interviewee 4), but instead harm the community. Interviewee 4 suggested that trolls' actions are aimed at 'flaming arguments and conflict among community members'. Trolls' activity at the community level is conducted primarily on community pages, such as the Village Pump. 'Trolls usually enter [the Village Pump] much earlier than other users do. These trolls are engaged in policy discussions and debates' (interviewee 4). Trolls' contributions to the discussions are uncooperative, unfriendly or unhelpful. Trolls simply 'attack users…[and] make disturbing comments from time to time' (interviewee 4). Interviewee 6 argued that most of the time a troll is:

> a user who repeatedly picks on other Wikipedia users and bothers them, [by] cursing, [evoking] personal and irrelevant attacks ("you're blocking me because you have a small penis")…exposing user's private information ("His real name is John Doe and he lives on 221 Baker Street"), and threatening physical vio-

lence…("I'll break every bone in your body").

All four trolls had a destructive community involvement. Troll 5's poorly written contributions and argumentative style generated significant tension among the Wikipedia community members. Troll 2 voted on removing sysop privileges as soon as his second day on Wikipedia; a privilege restricted to sysops only. This troll tried to gain power and privilege on Wikipedia before becoming a respectable contributing member of the community. Furthermore, his contribution to articles was almost non-existent, and most of his activities were at the community level, interacting with other users and mainly with sysops. Troll 2 caused a vast conflict among sysops when they were trying to decide whether his sock puppet was the same user as he was, and if he was a troll. It was suggested that his damages to the community were the worst of all the trolls (excluding only those of troll 1).

4.1.4. Hidden virtual identity and isolation

Wikipedia is based on collaborative efforts, and users contribute according to a set of policies, authority structures and coordination mechanisms. Wikipedia trolls, on the other hand, work alone in a non-cooperative manner; 'Each troll operates alone, and I can't recall an example of trolls working together' (interviewee 1). It was clearly emphasized that 'we did not observe attempts to coordinate efforts to attack Wikipedia; I am not aware that such coalitions were ever formed' (interviewee 3). In addition to working in isolation, a troll does not provide information on his real identity and uses multiple virtual identities (some of which act as sock puppets). 'Registered trolls are usually tricky and change usernames quickly' (interviewee 8). Among the four trolls, troll 2 had one male and one female user name, and troll 3 used ad hoc names for each of his destructive behaviours. Donath [6] argued that hidden identity is a common deceptive online behaviour. Various types of deception, besides trolling, have been identified [6]; some types reflect the true self, or are caused by psychiatric illness. Some deceptive acts are motivated by non-malicious intent like play or privacy concerns, yet others are motivated by malicious intent [29].

The behaviours that have been identified above extend and support prior research on the behaviours of internet trolls [5] and rough users [26]. Wikipedia troll behaviours resemble the three types of troll behaviours that have been discussed by Herring et al. [5]: 1) outward manifestations of sincerity, 2) flame bait, and 3) attempts to provoke futile arguments. Similar to Herring et al.'s troll, most of the Wikipedia trolls initially made some valid contributions to the encyclopaedia. They wrote and edited articles (manifesting their sincerity) before they started ignoring Wikipedia policies (flame bait) and began to argue with other users (attempts to provoke). However, our study, that emphasizes that trolls' behaviours are repetitive, is an extension beyond Herring et al.'s study [5]. A possible reason that the repetitive behaviour of internet trolls was not delineated in Herring et al.'s [5] study was the fact that their analysis is based on interactions over eight days, while the present study documents longer periods of interactions and analyzes multiple trolls.

Furthermore, Wikipedia trolls' behaviours involve abusive language (e.g. cursing and personal attacks on users) and excessive contact with administrators (e.g. flaming conflict, unwillingness to adhere to Wikipedia policies, and harmful activities at the community level), just like Gazan's [26] rough users. The present study not only supports Gazan's [26] descriptions but also extends it by suggesting that these behaviours are repetitive, ill intentioned and are conducted by each troll alone in a non-collaborative manner. It is possible that some of the Wikipedia trolls possess the characteristics that Gazan [26] attributed to NPD or other personality disorders.

4.2. Trolls' motivations

Trolls' motivations (i.e. the factors that direct and energize behaviours) were identified from the interview data and the case analyses. Three motivators of the trolls' attacks are described next. Again, following the same approach as in the case of trolls' behaviours, only motivations that have been observed in two troll cases, and which were specified by at least three interviewees, are included.

4.2.1. Boredom, attention seeking and revenge

Boredom was the most common motivator proposed by the interviewees. Interviewee 2, for instance, said: 'trolls are motivated by boredom…[they seek]…revenge on the community or on some members of the community'. Another interviewee claimed that: 'trolls are either there to have fun, or [for] revenge. In the case of [troll's name], he is furious about how he was treated in the past, and he is still haunting the site from time to time, having some very bizarre notions in mind' (interviewee 6). Troll 5 specifically made threats for revenge; he argued that sysops blocked him out of revenge. Besides boredom and revenge, several interviewees suggested that trolls simply 'seek attention' (interviewee 2) or 'desire attention' (interviewee 3). Trolling satisfies needs for attention or achievement, and even recognition, by affecting the actions of sysops and other users. Trolls 2 and 3 are good examples of the way trolls affect sysops' actions, making the sysops spend time fighting vandalism and arguing against the trolls.

4.2.2. Fun and entertainment

Interviewees suggested that trolls simply wish to have fun with Wikipedia while interacting with other users. One interviewee said: 'the joy they get from vandalizing' (interviewee 5) is a motivating factor to engage in destructive actions on Wikipedia. Another interviewee commented that '[troll's name]… greatly enjoyed what he did, and he was named "the creative troll" for a reason. When he was caught performing one of his silly tricks, [he was found]…barely holding himself from laughing loudly' (interviewee 6). The case of troll 3 probably best exemplifies this motivator with his creation of multiple user names and multiple online personae in order to have fun with the Wikipedia community. **4.2.3. Damage to the community and other people**

The interviewees argued that the trolls' agenda is simply to damage the community. One interviewee said that a troll: '[will] try to flame conflict among the community members until it will destroy the community' (interviewee 4). Trolls are motivated to create conflict, and consequently they are trying to irritate and upset community members. One of the interviewees indicated: 'they want to hurt other people' (interviewee 5). Unlike other Wikipedia vandals that damage Wikipedia articles and do not cause conflict among users, the trolls act at the community level, inflame conflict and attack users. All four Wikipedia trolls are clearly motivated to cause harm. The best examples are those of troll 5's attacks on sysops and death threats to the sysops' families, and troll 2's inflammatory conflict among the sysops.

Wikipedia trolls' motivations resemble those of online hackers. Mulhall [22] listed a series of hacker motivations, which some of the Wikipedia trolls share. Among them are: 1) intellectual curiosity (troll 3), 2) excitement (troll 3), 3) revenge (troll 5), 4) greed/wealth, 5) challenge, 6) access to information (trolls 1 and 5), 7) power (trolls 1 and 5), and 8) prestige (trolls 1 and 5). This similarity is echoed by the interviewees' comments, which provide additional support for Mulhall [22] finding of excitement and revenge motivators.

Yar [30] suggested that hackers' motivations can be viewed from two different perspectives, internal (hackers' own perspective on their motivations) and external (non-hackers' perspective on hackers' motivations). Yar [30] pointed out that under the internal perspectives, hackers tend to justify their activities and declare that they are motivated by reasons such as intellectual curiosity, free access to information, opposition to authority and boredom. On the contrary, the external perspective argues that hackers' motivations are drawn from internet addiction disorder [see for example, 26]. The Wikipedia trolls' motivators listed above reflect mostly the interviewees' point of view, and are therefore mostly external. Yet, the interviewees also mentioned a few internal motivators and these include, for example, boredom and opposition to authority (and policies). These, along with the external motivators, were clearly presented on the trolls' user pages and were apparent through the case analyses.

Most studies on hackers' motivations have focused attention on the internal perspective. For example, Turgeman-Goldschmidt [24] interviewed hackers in an effort to identify their accounts of hacking. The reasons hackers listed for their hacking behaviors varied but include: 1) fun and exciting activities 2) personal curiosity, 3) display of computer talent, 4) economic gains, 5) intangible offence, 6) revenge, and 7) ease of execution. Of these seven motivations that Turgeman-Goldschmidt [24] identified, four are shared with Wikipedia trolls and supported by the four trolls' cases: the

desire to have fun, thrill, and excitement; ease of execution; intangible offences; and their interest in revenge. Trolls 1 and 3 were motivated by the desire to have fun and to be engaged in exciting activities, and some trolling actions were taken also to satisfy these trolls' personal curiosity for its own sake. Finally, troll 5 was motivated by revenge.

However, unlike Turgeman-Goldschmidt's [24] results, the present study does not find that Wikipedia trolls are motivated by the deterrent factor, nosy curiosity, computer virtuosity or economic accounts. Possible explanations for the lack of similarity between the two studies in identifying these motivating factors include: 1) some accounts are irrelevant in the case of Wikipedia – a free encyclopaedia. For example, the economic account motivator is not relevant in the context of Wikipedia trolls, since neither Wikipedia nor its founders charge users for services; 2) it is possible that the difference is due to the diverse sources of data, which represent different points of view on the same phenomenon. Turgeman-Goldschmidt [24] interviewed hackers, while trolls were not interviewed in this study. Instead, the findings are mostly based on sysops' interviews, case synopses and online ethnographic observations (which are biased and partial). Thus, the disparity may be attributed to the different methods for data collection, rather than a reflection of a real difference between trolls and hackers. As indicated by Yar [30], inside perspectives (i.e. interviews with hackers) are inclined to present justification of their behaviours. Consequently, by interviewing trolls, researchers are more likely to gain a favourable justification of their behaviours and ideologies. *4.3. Trolls and hackers*

One of the goals of our study was to identify the similarities and differences between trolls and hackers, and to examine whether they are part of the same community. Two specific questions were addressed as part of the intent to understand the relationship between trolls and hackers: Q3a – Do Wikipedia trolls represent one (distinct) type of hacker?; and Q3b – Are there any trolls with ideology (hacktivists)? Our findings support the idea that Wikipedia trolls are part of the hackers' community. A few Wikipedia trolls share behavioural and motivational characteristics with the script kiddie group and perhaps even with the hacktivists group.

4.3.1. Wikipedia trolls as one type of hacker

Most Wikipedia trolls share common attributes with the second type of Falk's hackers, script kiddie; the most compelling example in this study is troll 3. A script kiddie is someone who downloads existing code and executes commands to cause harm. The troll cases support the idea that Wikipedia trolls share behavioural and motivational characteristics with script kiddie hackers, who tend to undertake hacking activities for personal enjoyment. Wikipedia trolls, like this type of hacker, operate at a lower level of the hackers' community [21].

The interviewees argued that trolls and hackers are comparable in that 'both find a leak in the system and [take] advantage of this knowledge to cause damage to the system' (interviewee 5). Another interviewee claimed that trolling 'is easier [than hacking]…but the idea is the same' (interviewee 8). Yet, hackers and trolls are perceived differently by the sysops. For example, interviewee 2, who argued on one hand that trolls are one type of hacker, also suggested on the other hand that hackers are geniuses and trolls are not, and claimed:

> A troll is a particular case of a 'hacker'; someone who takes advantage of a system's weakness (and it doesn't matter at all if the weakness is on a technological level, like a loophole in the security or in the freedom level – when anyone is allowed to edit an article) in order to cause damage to the computer system. The hacker usually has some stereotype of a genius (usually uncalled for) that doesn't exist in the case of a Wikipedia troll.

Interviewee 6 echoed this idea and claimed that hackers' actions are argued to be more intelligent than those of trolls: 'I met only one user who can be considered a "Wikipedia hacker" (in that his actions were more clever and amusing than the usual), but he was generally known as [a] troll'.

The interviewees contended that trolls and hackers also differ in their motivations and claimed that: 'hackers' attempts to damage computer systems are driven by a desire to achieve…a destruction of perceived evil of some sort…I am not sure what the trolls perceive as evil' (interviewee 3). Interviewee 4 added: 'hackers damage systems and trolls attack people'. Despite these differences,

interviewee 3 argued: 'I think the differences between hackers and trolls are subtle, especially in Wikipedia'. Thus, we conclude that trolls are part of the hackers' community and share some characteristics with other hackers.

4.3.2. Wikipedia trolls as hacktivists

The lack of attribution of ideology to Wikipedia trolls by the interviewees was contrasted by their perceptions of hackers; they perceived hackers as having ideologies, good cause and talent. This perception indicates that the interviewees generally considered hacktivists when asked about hackers.

After an effort to reveal the trolls' ideologies was pursued, it became apparent that a coherent set of shared ideologies could not be delineated from the data. At first glance it seems as if the trolls adhere to two ideas: the first is freedom of expression and the second is against (Wikipedia) bureaucracy. A few interviewees argued that at times, trolls are motivated by 'freedom of expression as the highest value…[and pursue] provocative examination of the freedom of expression' (interviewee 4) for achieving the same goal. They do so by 'taking advantage of the freedom on Wikipedia and damage articles' (interviewee 2). Yet, these arguments by trolls are not made in any organized manner, even though most of the trolls agree with them. It is possible that the lack of shared ideology is due to the fact that trolls largely operate individually; they do not form any type of coalition and, therefore, are not able to converge on a set of ideas.

Furthermore, all the interviewees felt that the concept of ideology is not relevant in the discussion of trolls. One interviewee said: 'I don't think trolls have a set ideology' (interviewee 5). Another interviewee echoed: 'I assume the troll thinks more of himself, not really meaning to actively deceive readers (except some types of political trolls…). I have no idea what the troll's actual ideology is, and I don't know if there is one' (interviewee 8). However, at the same time, a few of the interviewees identified and named trolls that held an ideology; among these trolls are trolls 4 and 9, who were sysops in the past, as well as troll 11 (see Table 1). Yet, because the identification of these users as trolls was not triangulated with other interviewees or with analyzed Wikipedia pages, these users were not included in this study as trolls.

The present study can, therefore, provide only limited support for the idea that trolls are motivated by ideology and conclude that Wikipedia trolls do not share similar characteristics with Falk's [21] hacktivists. None of the analyzed trolls belong to the hacktivists group, nor was this idea supported by the interview data. Yet, the fact that a few interviewees mentioned 'political trolls' (trolls driven by political ideologies) and other users with an ideology as trolls, may indicate that these hackers (hacktivist trolls) are active on Wikipedia. Future research may expand the definition of trolls and may find support for behaviours and motivations similar to the third type of Falk's [21] hackers, the hacktivists.

5. Conclusion

Given the lacuna in research on internet trolls (or Wikipedia trolls), this study is the first to analyze multiple trolls that are active on Wikipedia. This paper extends prior research and clarifies the relationship between trolls and hackers. This study identifies Wikipedia trolls' behaviours and motivations and compares and contrasts hackers with trolls. Despite the fact that Wikipedia sysops perceive major differences between hackers and trolls, partially because trolls do not appear to act on any ideologies, our study concludes that Wikipedia trolls are one type of hacker and that their behaviours and motivations resemble those of hackers. This paper extends and supports prior research on the behaviours of internet trolls [5] and rough users [26], and argues that Wikipedia trolls' motivations resemble those of online hackers [22, 24, 30]. The first step for security policy formulation and implementation for an online and a virtual environment is to understand motivations and behaviours of various types of vandals, hackers and trolls. Future research on Wikipedia may focus attention on the trolls' perspective, rather than the point of view of the sysops. Research into the activities of trolls on Wikipedias in other languages or various wiki-projects may elucidate the interactions between Wikipedians and trolls in different communities and examine how different settings may affect trolls' behaviours.

Endnotes

1 Motivation refers to the factors that direct and energize behaviour. Motives are the desired goals that underlie behaviour. Ideology refers to an organized collection of ideas.
2 As of 3 August 2009, the five largest Wikipedia languages were English, German, French, Polish, and Japanese (ordered according to number of articles).

References

[1] J. Surowiecki, *The Wisdom of Crowds* (New York, NY, Doubleday, 2004).
[2] Wikipedia, *Wikipedia: List of Wikipedias (2009).* Available at: http://meta.wikimedia.org/wiki/List_of_ Wikipedias (accessed 3 August 2009). [3] D. Riehle, How and why Wikipedia works: an interview with Angela Beesley, Elisabeth Bauer, and Kizu Naoko. In *Proceedings of the 2006 International Symposium on Wikis*, Available at: http://doi.acm. org/10.1145/1149453.1149456 (accessed 29 September 2006).
[4] M. Schwartz, The trolls among us, *The Times Magazine* August (2008). Available at: www.nytimes. com/2008/08/03/magazine/03trolls-t.html?pagewanted=1&_r=2&th&emc=th (accessed 3 August 2008).
[5] S.C. Herring, K. Job-Sluder, R. Scheckler and S. Barab, Searching for safety online: managing 'trolling' in a feminist forum, *The Information Society* 18(5) (2002) 371–383.
[6] J.S. Donath, Identity and deception in the virtual community. In: M.A. Smith and P. Kollock (eds), *Communities in Cyberspace* (London, Routledge, 1996) 29–59.
[7] A. Lih, Wikipedia as participatory journalism: reliable sources? metrics for evaluating collaborative media as a news resource, paper presented at *5th International Symposium on Online Journalism*, 16–17 April 2004.
[8] E. Raymond, The cathedral and the bazaar, *First Monday* 3(3) (1998). Available at: www.firstmonday.org/ issues/issue3_3/raymond/ (accessed 20 October 2006).
[9] J. Giles, Internet encyclopedias go head to head, *Nature* (2005). Available at: www.nature.com/news/2005/ 051212/full/438900a.html (accessed October 2006).
[10] R. Rosenzweig, Can history be open source? Wikipedia and the future of the past, *Journal of American History* 93(1) (2006) 117–146.
[11] B. Stvilia, M.B. Twidale, L. Gasser and L.C. Smith, Information quality in a community-based encyclopedia. In: S. Hawamdeh (ed.), *Knowledge Management: Nurturing Culture, Innovation, and Technology – Proceedings of the 2005 International Conference on Knowledge Management* (Charlotte, NC, World Scientific Publishing Company) 101–113.
[12] B. Stvilia, M. Twidale, L.C. Smith and L. Gasser, Information quality work organization in Wikipedia, *Journal of the American Society for Information Science & Technology* 59(6) (2008) 983–1001.
[13] F. Viegas, M. Wattenberg and K. Dave, Studying cooperation and conflict between authors with history flow visualizations. In: *Proceedings of the Computer-Human Interaction 2004* 6(1) (2004) 575–582.
[14] T. Holloway, M. Božicevic and K. Börner, Analyzing and visualizing the semantic coverage of wikipedia and its authors, *Complexity* 12(3) (2007) 30–40.
[15] A. Ardichvili, V. Page and T. Wentling, Motivation and barriers to participation in online knowledge-sharing communities of practice, *Journal of Knowledge Management* 7(1) (2003) 64–77.
[16] S. Bryant, A. Forte and A. Bruckman, Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In: *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work*, Available at: http://doi.acm.org/10.1145/1099203.1099205 (accessed 25 April 2007).
[17] A. Forte and A. Bruckman, From Wikipedia to the classroom: exploring online publication and learning. In *Proceedings of the 7th International Conference on Learning Sciences* (2006) 182–188. Available at: http://portal.acm.org/citation.cfm?id=1150061 (accessed 25 April 2007).
[18] M. Poe, The hive, *The Atlantic Monthly* (2006). Available at: www.theatlantic.com/doc/200609/wikipedia (accessed 25 October 2007).
[19] H. Nissenbaum, Hackers and the contested ontology of cyberspace, *New Media & Society* 6(2) (2004) 195–217.
[20] J. Thomas, The moral ambiguity of social control in cyberspace: a retro-assessment of the 'golden age' of hacking, *New Media & Society* 7(5) (2005) 599–624.
[21] C. Falk, Ethics and hacking: the general and the specific, *Norwich University Journal of Information Assurance* 1(1) (2005). Available at: http://journals.sfu.ca/nujia/index.php/nujia/article/view/3/7 (accessed 15 June 2008).
[22] T. Mulhall, Where have all the hackers gone? Part 3 – motivation and deterrence, *Computers & Security* 16 (1997) 291–297.

[23] L. Rennie and M. Shore, An advanced model of hacking, *Security Journal* 20 (2007) 236–251.
[24] O. Turgeman-Goldschmidt, Hackers' accounts: hacking as a social entertainment, *Social Science Computer Review* 23(1) (2005) 8–23.
[25] A.E. Voiskounsky and O.V. Smyslova, Flow-based model of computer hackers' motivation, *CyberPsychology & Behavior* 6(2) (2003) 171–180.
[26] R. Gazan, Understanding the rough user. In: Diane Nahl and Dania Bilal (eds), *Information and Emotion: The Emergent Affective Paradigm in Information Behavior Research and Theory* (Medford, NJ, Information Today, 2007) 177–185.
[27] M.B. Miles and M. Huberman, *Qualitative Data Analysis: An Expanded Sourcebook* (2nd edition) (Thousand Oaks, CA: Sage Publications, 1994). [28] U. Pfeil, P. Zaphiris and C.S. Ang, Cultural differences in collaborative authoring of Wikipedia, *Journal of Computer-Mediated Communication* 12(1) (2006) article 5. Available at: http://jcmc.indiana.edu/ vol12/issue1/pfeil.html (accessed 21 January 2007).
[29] S. Utz, Type of deception and underlying motivations: what people think, *Social Science Computer Review* 23(1) (2005) 49–56.
[30] M. Yar, Computer hacking: just another case of juvenile delinquency?, *The Howard Journal of Criminal Justice* 44(4) (2005) 387–399.

Appendix 1

Interview protocol

1. Please describe your activities with Wikipedia? (e.g. how long have you been active with Wikipedia? How frequently do you contribute? What roles do or did you play within the Wikipedia community? What kind of Wikipedia activities do you do?)

2. How do you contribute to Wikipedia? (e.g. what are your beliefs about your contributions? What kind of contributions do you make? What are your favourite topics (subject domain) and/or issues (policies)?)

3. What does the term 'troll' mean to you? What kinds of activities are typical of Wikipedia trolls? Do you see a difference between the damage that is caused unintentionally and those that are caused by vandals? What are the similarities and differences between them? In your opinion do the activities of Wikipedia trolls and hackers resemble each other? What are the similarities and differences between them?

4. How do you track trolls, and what do you do once you have identified a potential troll? With how many trolls have you interacted with in the past? Do you currently interact with any trolls? If so, do you interact with them on Wikipedia space or via email? What are their user names on Wikipedia? Can I contact any of these trolls? Is there anyone else that you know that may be able to help me make contact with trolls?

5. Would you consider '[xxx]' a troll? Could you please explain why? Is '[yyy]' a troll? Why?

6. Do trolls form coalitions or operate individually?
7. Could you please describe your ideology (regarding Wikipedia) and compare and contrast it with the trolls' ideology?