

Facing Defacement

Combating harassment with user page protection

by Pax Ahimsa Gethen, August 2017



Background

Have my full name, gender identity, photo, and link to website on my user page

Attacked and stalked by anonymous editor in 2016

- Deadnaming, misgendering, and graphic sexual harassment
- Vandalized my user pages on en:wp, Commons, and Meta
- Stalked all of my edits for months after my user pages were protected
- Got Support and Safety involved, on-wiki harassment stopped eventually
- But continued to stalk my blog until I shut off comments

Inspire Campaign to address harassment

June 2016 Inspire Campaign solicited ideas for addressing and preventing harassment

I proposed applying semi-protection to user pages by default

- Semi-protection: No edits from anonymous or new (unconfirmed) users
- On English Wikipedia, users are autoconfirmed after 4 days and 10 edits

Originally proposed semi-protecting both the user page and user talk page

Idea tied for 5th place on campaign leaderboard, with 28 endorsements

Request for Comment

Created RfC in August 2016 based on positive response from Idea Campaign

Got help from Idea Campaign manager Chris “Jethro” Schilling (in a volunteer capacity)

Proposed separate options from semi-protection and extended-confirmed protection

- On English Wikipedia, extended confirmed requires 30 days and 500 edits

Also proposed toggle switch in user interface so user could unprotect page themselves

Excluded talk pages from default protection based on feedback

Sub-pages in user space also excluded from default protection

Rationale

Harassment on Wikipedia is a significant problem, especially for marginalized people

- Emotional distress
- Decreased participation

Harassment on the Internet in general is a serious problem

Addressing abuse is necessary for healthy functioning of online communities

- Wikipedia *is* a community; every editor is a living person with real feelings
- Communities should be obligated to use tools to reduce and prevent harassment

Objections and responses

Preventing anonymous edits to pages is against the open nature and spirit of Wikipedia.

- While users don't "own" their user pages, these pages are *not* encyclopedia articles. Template available:
 - **This is a Wikipedia [user page](#).** This is not an encyclopedia article. If you find this page on any site other than Wikipedia, you are viewing a [mirror site](#). Be aware that the page may be outdated and that the user to whom this page belongs may have no personal affiliation with any site other than Wikipedia itself. The original page is located at <https://en.wikipedia.org/wiki/User:Funcrunch>.

Vandals will just turn to harassing users on talk pages and stalking their edits on articles.

- Vandalizing a user's page is like spray-painting graffiti on their front door
- Not being able to prevent *all* vandalism and harassment is no excuse for inaction

Unconfirmed users can make useful contributions, such as tagging user pages for spam and other unpermitted uses.

- Harm caused by harassment outweighs convenience and immediacy of anon tagging; can still report pages
- No demonstrated, compelling need for unconfirmed editors to edit other users' pages

RfC: Results and implementation

RfC closed October 2016 with a consensus for semi-protection

- Some editors contested the close based on !votes in each section
- Closing administrator defended decision; not just a matter of counting !votes

Phabricator task filed; decided on edit filter rather than default protection

- Easier to program, and keeps log of edit attempts as a bonus

Edit filter implemented on English Wikipedia in November 2016

User or admin can now insert “unlocked userpage” template to allow unconfirmed edits

Instructions available on Meta for implementing on other language Wikipedias

Examples and themes

Not a scientific analysis; pointing out edits reflecting:

- Racism
- Sexism
- Homophobia and transphobia
- Threats of bodily harm

Examples gathered January—February 2017 and July 2017

Examples are uncensored (except to remove names) and may be triggering

Specific examples of prevented vandalism

Racism, sexism, transphobia

- SCREW YOU BLACK NI...G..G...ER NI..G..GE..RS ARE DUMB AND POOOR!
- Fuck all u niggas
- Ugly ass nigga
- Cunt
- Bitch
- I'm a cunt
- this man is a cunt
- I'm a coc sucking cunt who loves to kill innocent people
- Don't touch the chapter 8 page anymore you whore
- [name redacted] is a transsexual who likes to have sexual intercourse with all wikipedia admins and likes to have her boobs sucked by the admins.
- [Replaced "semi-retired" and "feminismsupporter" with "transsexual"]

Specific examples, continued

Homophobia (plus more racism)

- I am a gay man who likes taking it in the bum from well endowed African gentlemen.
- I am gay: I really like sucking big black cocks. I am a 40 year old gay man with no life. I suck dick as my job.
- I'm also gay asf and love fat black dicks in my ass
- Also love gangbangs with big black men
- I ALSO WANT TO SAY THAT I LOVE BIG DICKS IN MY ASS
- HEY YOU IS GAY.
- I'm very ugly and gay but I need a boyfriend
- [name redacted] enjoys the taste of cock
- [name redacted] is a homosexual
- This user is strictly, unquestionably, undeniably and exclusively gay beyond belief. He has dick sucking contests all the time.
- be a Godless faggot like this skeptical POS

Impact of hate speech

Hate speech has consequences regardless of the actual race, gender, or sexual orientation of the targeted user

Homophobic language conveys that being gay and participating in homosexual acts is a bad thing

Hate speech creates an unwelcoming or threatening environment for people from marginalized communities

Attacks might seem merely juvenile to some, but can be devastating to others

Specific examples, continued

Threats

- harassment will continue till your death [name redacted]
- im coming for you / lock your doors / close your windows / i know were you live
- kill yourself you fat bastard
- Eat s*it and die you a-hole.
- Shoot me
- [name redacted] says that I am going to commit suicide today
- please hang yourself and die
- I am a big fat cunt and should be hanged until dead

Taking threats seriously

You never know for sure who is joking and who actually intends to act out violently

Regardless, threats of death and bodily harm are nothing to kid about

After reporting one attack to I was told the report was “stale” as it was more than a couple of hours old; IP address not blocked

But I have been generally successful at getting serious vandals blocked

Responding to vandalism attempts

I monitor log for edit filter 803 periodically for attempted vandalism and harassment

For significant attempts I use warning template (usually starting with level 2):

- Please refrain from attempting to make unconstructive edits to Wikipedia. Your edits appear to constitute [vandalism](#) and have been [disallowed by an edit filter](#). If you would like to experiment, please use the [sandbox](#).

As this warning confused some targeted users, I added notice to my talk page:

- **Please read:** If you're coming here wondering why I pinged you about vandalism to your user page when there is no obvious recent vandalism in that page's history, it was probably *attempted* vandalism by an unconfirmed or anonymous user which was caught by [this edit filter](#). I've been monitoring that [filter log](#) because I started the [RfC](#) that resulted in its creation.

I report serious/repeated vandalism attempts to AIV messageboard

Next steps?

Comprehensive analysis of abuse log over the first year of filter implementation

New survey of users, asking about harassment and safety

Continue efforts to address and prevent harassment everywhere on Wikipedia

Raise awareness of harassment everywhere on and off of the Internet

Contact info

Pax Ahimsa Gethen

User:Funcrunch (English Wikipedia, Commons, Meta)

funcrunch.org

funcrunch-files@funcrunch.org