



# **End violence: Women's rights and safety online**

## **Internet intermediaries and violence against women online**

### **YouTube: A case study**

*Carly Nyst*

*Association for Progressive Communications (APC)*

*July 2014*



Ministry of Foreign Affairs

*This research is part of the APC "End violence: Women's rights and online safety" project funded by the [Dutch Ministry of Foreign Affairs \(DGIS\)](#) and is based on a strong alliance with partners in seven countries: Bosnia and Herzegovina, Colombia, Democratic Republic of Congo, Kenya, Mexico, Pakistan and the Philippines. For more information visit [GenderIT.org](#) and [Take Back the Tech!](#)*

**Table of contents**

- 1. Analysis of main trends.....3
- 2. Compliance with the Guiding Principles on Business and Human Rights.....4
  - 2.1. Recommendations.....9
- 3. Google/YouTube’s user policies.....10
  - 3.1. Which rights violations are explicitly recognised and provided for in corporate policies?.....10
  - 3.2. What is the process for reporting violations?.....13
  - 3.3. What are the support mechanisms in place for victims/survivors?.....21
  - 3.4. At what point does the intermediary collaborate with others to facilitate access to justice?. 22
  - 3.5. Evolution of Google/YouTube’s policies related to technology-related VAW, 2009 to 2014...22
- 4. Interview with corporate representative.....23

## **Introduction to the research**

This profile is part of a short study of the policies of three major internet intermediaries, Facebook, YouTube and Twitter, with respect to violence against women online. The study aims to map the corporate policies of these intermediaries that allow for the identification, reporting and rectification of incidents of harassment or violence against women via the service that the intermediary provides. In addition to providing a detailed summary of the user policies relevant to this issue, the study also compares the impact and effectiveness of those policies against the framework of the UN Guiding Principles on Business and Human Rights. The study was conducted on the basis of desk research and an analysis of corporate policies and terms of service, and interviews with representatives of the intermediaries.

## **About YouTube**

YouTube is a video-sharing website, established in 2005 and bought by Google in late 2006. There is no publicly available information about its number of employees. In 2007, YouTube introduced a localisation feature, whereby YouTube is localised in 19 countries. In March 2013, the number of unique users visiting YouTube every month reached one billion.

## **Report format**

This report is broken down into four main sections:

1. Critical analysis of main trends.
2. Charting the impact and effectiveness of YouTube's policies and procedures with respect to violence against women, using the framework of the Guiding Principles on Business and Human Rights.
3. A detailed breakdown of YouTube's user policies, redress mechanisms and the evolution of its approach to violence against women.
4. Transcript of an interview with a YouTube representative on the company's approach to violence against women.

## **1. Analysis of main trends**

YouTube is a unique internet intermediary. It provides a platform for the expression of thoughts, ideas and artistic endeavours, and reaches hundreds of millions of people daily. Yet, just as its varied and multicultural content mirrors the societies which it reaches, so too does the spectrum of offensive, hateful, violent and illegal content which can be found online on YouTube mirror the occurrence of such conduct in the offline world.

YouTube has had to walk a fine line between providing an un-moderated and uncensored platform for free expression and exchange, and controlling the proliferation of violent and threatening

content and copyright infringement. With respect to controlling and remedying human rights violations online, it has developed a reporting process both with regard to content and to privacy rights infringements. It has publicly declared that hate speech, harassment and threats will not be tolerated. Yet the effectiveness of its response to such violations is hard to determine.

There is no publicly available information about how or under what circumstances YouTube cooperates with law enforcement or assists users to access justice mechanisms, and this is a clear failing on the part of the company. Lack of transparency across the reporting process is generally a problem for YouTube – there is no information about the complaints handlers, what type of training they receive, how they deal with complaints, what levels of YouTube management are involved in complaints, etc. Staff receive training on, inter alia, violence against women issues, but there is no information about the content of the training or who provides it. YouTube could be far more transparent about its processes in order to give its users more certainty about how their complaints are dealt with.

## **2. Compliance with the Guiding Principles on Business and Human Rights**

Human rights obligations do not only relate to the actions or omissions of states. Companies are also required under international law to respect human rights, to avoid infringing human rights, and to address adverse human rights impacts with which they are involved. This means not only do they have to take action when they play a role in human rights violations, but they also have to take positive steps to prevent, mitigate and remedy human rights violations.

The steps required by companies to fulfil human rights obligations were analysed at length by the UN Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises. The SRSG compiled a set of Guiding Principles on Business and Human Rights, which was endorsed by the UN Human Rights Council in 2011. The Guiding Principles enshrine a framework of obligations, entitled “Protect, Respect and Remedy”, which tells both states and companies what steps they should take to promote human rights.

When it comes to addressing technology-related violence against women, the second pillar of the Framework – Respect – provides guidance for internet intermediaries as to the actions they should take to ensure that women’s rights online are promoted and respected. The Respect pillar sets a number of benchmarks that companies must reach in order to be in compliance with human rights obligations.

The third pillar of the Framework – Access to Remedy – establishes that states must take steps to ensure judicial, administrative or other remedies to ensure that victims of human rights abuses can obtain redress. While the pillar primarily addresses the role of states in this regard, it also provides that business enterprises should establish or participate in effective operational-level grievance mechanisms for adverse human rights abuses.

The Guiding Principles may provide an effective and useful structure within which to engage internet intermediaries on technology-related violence against women concerns. The Framework prescribes a number of actions that can be adapted to analyse the actions of internet intermediaries in this regard.

We have developed a list of questions,<sup>1</sup> to correspond with the Principles, that organisations, advocates and activists can use to analyse the compliance of internet intermediaries with the Guiding Principles. Below, we use the questions to address the question of YouTube’s compliance with human rights obligations in its approach to issues of violence against women.

|                          |  |
|--------------------------|--|
| <b>Policy commitment</b> |  |
| 1.                       | Does the intermediary have a publicly available statement of policy that stipulates the organisation’s policy with respect to violence against women (in all of its forms)?  |
|                          | <p>YouTube does have a publicly available statement of policy (the Community Guidelines) that emphasise that it does not permit hate speech, graphic or gratuitous violence, threats, predatory behaviour, harassment or the invasion of privacy. However, there is no public commitment to human rights standards or to the promotion of rights, other than the encouragement of free speech. YouTube’s available policies do not explicitly address gender-related violence or harassment nor take a strong stance on respect for diversity or for women’s rights.</p> <p>While YouTube’s Community Guidelines are a clear, concise statement that the intermediary will not tolerate hate speech – including that on the basis of gender and gender identity – YouTube could more explicitly commit to a more holistic embrace of human rights standards and a recognition of the range of human rights engaged by the services it provides, beyond solely freedom of expression.</p>   |
| 2.                       | Has the intermediary taken due diligence steps to understand the way in which it may be facilitating violence against women, in order to inform its policies and procedures?   |
|                          | <p>An analysis of YouTube’s publicly available policies and procedures, the evolution of these policies, and discussions with the representative of the company reveal that YouTube has tried to view its role as neutrally as possible since its inception, and has not been enthusiastic about controlling content or intervening in the activities that take place on its platform. To this end, it seems that YouTube has been reluctant to take positive measures to understand its role in facilitating violence, as it pertains to women or to other groups historically subject to violence or discrimination. There are no indications that YouTube has sought the input of stakeholders or community groups, or has commissioned studies or due diligence surveys in this regard. While the company is clearly aware of the potential of the platform to be used to propagate harassment, hate speech or violence, it does not appear to have a comprehensive or evidence-based understanding of the ways in which its processes might contribute to the proliferation of such behaviour.</p> <p>YouTube could take such positive steps to understand the proliferation of violence against women on its platform and to map out the human rights implications of its policies, in order to help it better mitigate and address these harmful effects.</p> |
| <b>Due diligence</b>     |  |
| 3.                       | Has the intermediary engaged in meaningful consultation with women, either by soliciting the input of users or by engaging women’s rights groups and activists, to understand the potential adverse impacts of its services on women’s rights?   |
|                          | <p>YouTube does not appear to have in place a process of consultation with women’s rights groups or activists, nor does it point to any concrete examples of when it has reached out</p>   |

<sup>1</sup> These questions are based on the Access to Justice Framework for Corporate Policies, as detailed in the Ending Violence: Domestic Legal Remedies and Corporate Policies/Redress Mechanisms Research Design.

|   |
|---|
| <p>to such groups for input.</p> <p>However, it is clearly engaged with its community in other areas; for example, it has recently established the YouTube Deputy Program where it is reaching out to active users to help to more actively moderate content. The YouTube representative we spoke to indicated that they would be interested in having more engagement with women’s organisations in this respect, so this might be an interesting angle for organisations to pursue.</p>   |
| <p>4. Is responsibility for addressing issues of violence against women assigned to the appropriate level and function within the intermediary?</p>   |
| <p>With respect to general policy making, it is difficult to see where YouTube has mainstreamed women’s issues into its policies. There is no particular person or division responsible for women’s issues and no working group, stream or committee engaged in these issues. YouTube has not published a policy or approach to violence against women issues, nor have they signed the Women’s Empowerment Principles.</p> <p>With respect to YouTube’s reporting processes, there is greater clarity about how issues of violence against women are prioritised. YouTube’s grievance mechanisms for dealing with reported or “flagged” content are staged, and do in some instances engage people at the higher echelons of YouTube’s management. While initially, standard complaints are dealt with by an employee who is a complaints handler, serious or contentious complaints are escalated to complaints handlers with more experience, and even eventually to policy and legal representatives from other branches of the YouTube structure.</p> <p>However, YouTube neither publishes nor was able to provide disaggregated information about the gender, expertise and training of complaints handlers. It should be encouraged to do so in order to establish that those entrusted with complaints handling are properly qualified and that an effective gender balance is being struck.</p> |
| <p>5. Do internal decision-making processes enable effective responses to issues of violence against women?</p>   |
| <p>Because YouTube’s internal decision-making processes are obscured by a lack of publicly available information about what training is provided on women’s issues, and what percentage of staff members or complaints handlers are women, it is difficult to get a sense of how effective internal decision-making processes are. In addition, there are few concrete public examples of when issues of violence against women have been dealt with at all by YouTube.</p> <p>YouTube should be encouraged to embrace greater transparency of their processes in order to enable greater scrutiny of them.</p>   |
| <p>6. Does the intermediary track how effective its responses to issues of violence against women are, either by tracking indicators or seeking feedback from affected stakeholders?</p>  |
| <p>YouTube does not publish disaggregated numbers about the types of complaints it receives and the outcomes of those complaints, so it is impossible to know how many are resolved to the satisfaction of the reporter. YouTube has made no public effort to solicit feedback from the women’s rights community or from other affected stakeholders as to the effectiveness of its responses to issues of violence against women.</p> <p>As above, we reiterate that YouTube should be encouraged to embrace greater transparency of its processes in order to enable greater scrutiny of them.</p>  |
| <p>7. Does the intermediary publicly communicate both the occurrence of, and its response to, issues of violence against women?</p>   |
| <p>YouTube does not publish information about the actions it takes to mitigate or address instances of violence against women. This is a serious failing which should be rectified by YouTube.</p>  |

| <b>Remediation</b> |  |
|--------------------|--|
| 8.                 | Is there a grievance mechanism in place for individuals or communities who are adversely affected by violence against women?   |
|                    | <p>YouTube adopts a self-reporting model, where users are invited to “flag” content or comments for review by YouTube. Two general types of complaints are possible: a content-related complaint, or a privacy-related complaint. Both complaints are made by an online reporting mechanism, and require the reporter to identify themselves, either by having a Google account (for the content-related complaint) or by providing identifying information. There is some provision for third-party reporting where the complainant does not have access to a computer.</p> <p>The requirement that individuals have a Google account in order to report violations seriously limits the accessibility of grievance mechanisms, and also has implications for limiting consumer rights. It effectively forces people to sign up to a service, which they may want no part in, in order to report a violation of their rights.</p>   |
| 9.                 | Does the intermediary consult stakeholder groups on the design and performance of its grievance mechanism?   |
|                    | There is no indication that YouTube consulted women’s groups or activists in the design of its grievance mechanism, nor is there any provision for feedback on how the mechanism could be better designed or operated.   |
| 10.                | Does the mechanism meet the following effectiveness criteria?  |
|                    | 10a. Legitimacy – is the mechanism viewed as trustworthy, and is it accountable to those who use it?   |
|                    | <p>There is little publicly available information from those who have used the feedback mechanism and found it trustworthy or not. YouTube endeavours to respond quickly (although provides no information on what “quickly” means) to complaints and an efficient system would be more likely to engender trustworthiness. YouTube follows up with the reporter on the outcome of their complaint in relation to content-related and privacy-related complaints, which is an important element of accountability.</p> <p>There is insufficient anecdotal or other evidence to establish whether the YouTube flagging and reporting mechanisms are sufficient. YouTube should thus publish its own information on the number of reports received, responded to, and acted upon, in order to enable more detailed engagement with this question.</p>  |
|                    | 10b. Accessibility – is the mechanism easily located, used and understood?   |
|                    | <p>The YouTube reporting mechanisms are easily located and understood. They are written in plain language and users are guided through the steps quickly and simply. YouTube users are able to choose from 61 languages on the site. Users do not have to search long and hard to locate the reporting mechanism, which is an important attribute.</p> <p>However, individuals are only able to report violations online, and this may exclude individuals without immediate access to computers or without computer literacy from accessing the grievance mechanism.</p> <p>YouTube’s complaint process requires the disclosure of at least basic identifying information, although YouTube indicates that it will not disclose this information to the person responsible for uploading the offensive or violent content. While identification is a necessary step for limiting spurious privacy complaints from being made, the grievance mechanism should be made accessible to those both with and without computer/email access.</p> |
|                    | 10c. Predictability – is there a clear and open procedure with indicative time frames, clarity of process and means of monitoring implementation?  |
|                    | <p>YouTube points to its processes on its website, but the implementation of the process remains obscured. There is no publicly available information about who will be dealing with the complaint, what standards they will apply, or who they will consult on their decision. There are no examples or case studies available for people to understand how previous</p>  |

|  |
|--|
| <p>complaints have been dealt with. YouTube should introduce greater transparency in this regard.</p> <p>Importantly, there is no publicly available information about how or under what circumstances YouTube cooperates with law enforcement or assists users to access justice mechanisms. This is a serious failing of YouTube’s grievance mechanisms and should be rectified.</p>   |
| <p>10d. Equitable – does the intermediary provide sufficient information and advice to enable individuals to engage with the mechanism on a fair and informed basis?</p>   |
| <p>With respect to content-related complaints, YouTube does provide some extended guidance about what constitutes harassment and cyber bullying, hate speech, threats and violence or graphic content – which may assist users in designing their complaint to ensure that it falls within YouTube’s ambit. However, there is little guidance on the privacy-related complaint and what burden of proof users must provide to establish an intrusion in this regard.</p>   |
| <p>10e. Transparent – are individuals kept informed about the progress of their matter?</p>  |
| <p>It is unclear to what extent YouTube keeps reporters updated about the progress of content-related or privacy-related complaints. The YouTube representative indicated that YouTube will report back to the reporter with the outcome of the matter, but this is not explicitly stated in the policies.</p> <p>With respect to comment-related complaints, YouTube does not appear to inform reporters of the outcome of the matter, but rather indicates that if the report is found to be substantiated YouTube will remove the relevant comment.</p> <p>YouTube should be encouraged to improve these processes, as certainty as to the outcome is an important element of assisting victims/survivors of violence against women.</p>  |
| <p>10f. Rights-compatible – do the outcomes and remedies accord with internationally recognised human rights?</p>  |
| <p>There does not appear to be an established process for appealing an adverse decision about content- or privacy-related complaints. This would not comport with the need for independence, impartiality and accountability in processes of remedy and redress.</p> <p>There is a serious need for greater participation of the victim/survivor of violence against women in the grievance mechanism. Reporters should have a greater opportunity to provide information and explanation as to the reason for their complaint and the impact that the offensive or violent behaviour had on them. In the content-related complaint process, they are only able to provide up to 500 characters in this regard. This would seem to be insufficient to enable women to properly elucidate their problems and experiences.</p> <p>There are insufficient victim support mechanisms available through YouTube, and this should be rectified. YouTube should put in place better procedures to ensure that individuals and communities can give input into the design of the processes themselves, which do not reflect a consideration of the particular difficulties of ensuring justice for victims/survivors of violence against women online.</p> |
| <p>10g. Source of continuous learning – does the intermediary draw on experiences to identify improvements for the mechanism and to prevent future grievances?</p>   |
| <p>As reiterated above, YouTube has taken few visible measures to incorporate the particular challenges of violence against women into their complaints mechanism, and without greater transparency around the process and greater introspection on this particular point it is difficult to see how it can begin to take proactive steps to prevent further violence against women on the platform.</p>   |

## 2.1. Recommendations

In order to ensure that it is meeting its obligations to respect and advance human rights standards, particularly the right of women to be free from harassment, hatred and violence online, YouTube should take the following steps:

1. Make a public commitment to human rights standards, and take a strong stance on respect for diversity and for women's rights.
2. Sign the Women's Empowerment Principles.<sup>2</sup>
3. Take more positive steps to understand the proliferation of violence against women on YouTube, and map out the human rights implications of its policies, in order to help it better mitigate and address these harmful effects.
4. Put in place a concrete process of consultation with women's rights groups and activists on the design, implementation and evaluation of policies and procedures.
5. Establish a point-person responsible for understanding and responding to issues related to violence against women, and for establishing – in consultation with the relevant individuals and communities – a YouTube policy towards issues of violence against women.
6. Publish disaggregated information about the gender, expertise and training of complaints handlers dealing with content- and privacy-related complaints.
7. Provide regular training to staff responsible for moderation on issues related to human rights in general, and to the specific realities of women's rights as they pertain to health, sexuality, violence.
8. Provide greater transparency about complaints processes more generally, what standards are applied and how complaints are dealt with throughout their lifecycle.
9. Publish information about instances of violence against women on YouTube, including information on the number of reports received, responded to, and acted upon, in order to enable more detailed engagement on these issues.
10. Consider providing alternative reporting mechanisms for individuals who are not computer literate or who might be unwilling to disclose identifying information in making a complaint.
11. Provide more guidance on the privacy-related complaint and what burden of proof users must provide to establish an intrusion in this regard.
12. Make victim support mechanisms available through YouTube.
13. Establish mechanisms for greater participation of the victim/survivor of violence against women in the reporting process, giving reporters a greater opportunity to provide

---

<sup>2</sup> The **Women's Empowerment Principles** (WEP) is a global initiative focused specifically on corporate social responsibility and women's human rights, and is a collaboration between UN Women and the UN Global Compact (see <http://weprinciples.org>). Signing in support of the WEP entails recognition of the costs of violence against women to businesses, and a commitment to developing internal and external initiatives to increase women's empowerment within the workplace, marketplace and community.

information and explanation as to the reason for their complaint and the impact that the offensive or violent behaviour had on them.

### **3. Google/YouTube's user policies**

Below, we analyse Google/YouTube's current user policies to ascertain to what extent they address, prohibit and provide redress for technology-related violence against women.

#### **3.1. Which rights violations are explicitly recognised and provided for in corporate policies?**

Section 1.2 of YouTube's Terms of Service<sup>3</sup> provides that users agree to use YouTube in accordance with the Terms of Service, the Privacy Policy,<sup>4</sup> and the Community Guidelines.<sup>5</sup>

The Community Guidelines are plainly worded and non-technical. They set down "some common-sense rules that will help you steer clear of trouble," among which are:

- Graphic or gratuitous violence is not allowed. If your video shows someone getting hurt, attacked, or humiliated, don't post it.
- We encourage free speech and defend everyone's right to express unpopular points of view. But we don't permit hate speech (speech which attacks or demeans a group based on race or ethnic origin, religion, disability, gender, age, veteran status, and sexual orientation/gender identity).
- There is zero tolerance for predatory behavior, stalking, threats, harassment, invading privacy, or the revealing of other members' personal information. Anyone caught doing these things may be permanently banned from YouTube.

Please take these rules seriously and take them to heart. Don't try to look for loopholes or try to lawyer your way around them—just understand them and try to respect the spirit in which they were created. Violations of the Terms of Use may result in a warning notification or termination of your account. If your account is terminated you are prohibited from creating any new accounts.

The Terms of Service are more technical in nature, and have a general catch-all provision that applies the Community Guidelines to the content uploaded to YouTube, at Section 7.5:

You agree that your conduct on the site will comply with (and you agree that the content of all of your Content shall comply with) the YouTube Community Guidelines... as updated from time to time.

---

<sup>3</sup> [www.youtube.com/t/terms](http://www.youtube.com/t/terms)

<sup>4</sup> [www.youtube.co.uk/t/privacy](http://www.youtube.co.uk/t/privacy)

<sup>5</sup> [www.youtube.com/t/community\\_guidelines](http://www.youtube.com/t/community_guidelines)

YouTube has a [Policy and Safety Hub](#) site, which goes into more detail about the types of behaviour prohibited by the Community Guidelines.<sup>6</sup> In the [Policy Center](#), users can “Learn More About YouTube’s Policies” and find out how to report behaviour that contravenes the Community Guidelines.<sup>7</sup>

On the page dedicated to **harassment and cyber bullying**, YouTube says:

YouTube users share their opinions on all sorts of topics - from politics to religion, social media to celebrities, and much more. These conversations can sometimes become passionate, so it's important to be tolerant and allow for debate to flourish while avoiding malicious personal attacks.

It comes down to respect. If you're not sure whether a video or comment you've made crosses the line, follow a simple rule of thumb: if you wouldn't say it to someone's face, don't say it on YouTube.

We take this issue seriously and will remove comments, videos or posts where the main aim is to maliciously harass or attack another user. However, at YouTube we understand the value of free expression, so please understand that not all negative or mean videos and comments will be removed.

When content violates our harassment policy, the person who posted that content will receive a strike against their YouTube account. Accounts that are dedicated to harassing a particular user or the community at large will be terminated.

With respect to **hate speech**:

We encourage free speech and try to defend your right to express unpopular points of view, but we don't permit hate speech.

Hate speech refers to content that promotes violence or hatred against individuals or groups based on certain attributes, such as:

- race or ethnic origin
- religion
- disability
- gender

---

<sup>6</sup> [www.youtube.com/yt/policyandsafety](http://www.youtube.com/yt/policyandsafety)

<sup>7</sup> [www.youtube.com/yt/policyandsafety/policy.html](http://www.youtube.com/yt/policyandsafety/policy.html)

- age
- veteran status
- sexual orientation/gender identity

There is a fine line between what is and what is not considered to be hate speech. For instance, it is generally okay to criticize a nation-state, but not okay to post malicious hateful comments about a group of people solely based on their race.

With respect to **threats**:

The YouTube community is important to us and we want to see it continue to flourish. To ensure that this is possible, content that makes threats of serious physical harm against a specific individual or defined group of individuals will be removed.

People who threaten others may receive a strike on their account and their account may be terminated.

With respect to **violent or graphic content**:

Real depictions of graphic or violent content

Increasingly, YouTube is becoming an outlet for citizen journalists, documentarians and other users to publish accounts of what is happening in their daily lives. It is inevitable that some of these videos will contain content that is violent or graphic in nature.

If the violence shown in your video is particularly graphic, please make sure to post as much information as possible in the title and metadata to help viewers understand what they are seeing. Providing documentary or educational context can help the viewer, and our reviewers, understand why they may be seeing the disturbing content.

It's not okay to post violent or gory content that's primarily intended to be shocking, sensational or disrespectful. If a video is particularly graphic or disturbing, it should be balanced with additional context and information. For instance, a citizen journalist who captures footage of protesters being beaten and uploads it with relevant information (date, location, context, etc) would likely be allowed. However, posting the same footage without contextual or educational information may be considered gratuitous and may be removed from the site.

Dramatized depictions of graphic or violent content

Some people post videos that contain dramatized depictions of violence. Much like movies and TV, graphic or disturbing content that contains a certain level of violence or gore is not suitable for minors and will be age-restricted.

Age-restricted content

In some cases, real, dramatized or fake violence may not be suitable for all ages. Similar to movie or television ratings, our age-restrictions help viewers avoid watching content that they may not feel is acceptable for themselves or for their children.

### 3.2. What is the process for reporting violations?

YouTube adopts a self-reporting model, where users are invited to “flag” content or comments for review by YouTube. Two general types of complaints are possible: a content-related complaint, or a privacy-related complaint.

#### **Content-related complaint**

Content is reported and reviewed on the basis of a flagging process, outlined in brief in the Community Guidelines:<sup>8</sup>

Okay, this one is more about us than you. When a video gets flagged as inappropriate, we review the video to determine whether it violates our Terms of Use—flagged videos are not automatically taken down by the system. If we remove your video after reviewing it, you can assume that we removed it purposefully, and you should take our warning notification seriously. Take a deep breath, read our Terms of Use and try to see it from our perspective. If you find other videos on YouTube with the same violations, please flag them so we can review them as well!

It is important to note that in order to flag and report content, an individual must have an account with Google. To open an account, an individual needs to provide the following information:

- Name
- Birthday
- Gender
- Mobile phone number
- Country of residence

**To flag a video**, users click on the small flag embedded underneath the relevant video, and are offered a list of issues to choose from. Hovering over the small question mark to the right of each issue gives a brief description of its applicability.

<sup>8</sup> [www.youtube.com/t/community\\_guidelines](http://www.youtube.com/t/community_guidelines)

Like    About    Share    Add to

### Report this video

**What is the issue?\***

- Sexual content ?
- Violent or repulsive content ?
- Hateful or abusive content ?
- Harmful dangerous acts ?
- Child abuse ?
- Spam or misleading ?
- Infringes my rights ?
- Captions report (CVAA)

Once users click on an issue, they are asked to select from a number of sub-issues that further describe the reason for their flag.

Like    About    Share    Add to

### Report this video

**What is the issue?\***

- Sexual content ?
- Violent or repulsive content ?
- Hateful or abusive content ?
  - ✓ Select one
  - Promotes hatred or violence
  - Abusing vulnerable individuals
  - Bullying
- Child abuse ?
- Spam or misleading ?
- Infringes my rights ?
- Captions report (CVAA)

Flagged videos and users are reviewed by YouTube staff 24 hours a day, seven days a week to determine whether they violate Community Guideline. Accounts are penalized for Community Guidelines violations and serious or repeated violations can lead to account termination. If you would like to report a channel, please [click here](#)

\* Required    **Submit**

Once a user selects a sub-issue, they are invited to provide more information about their complaint before it is submitted.

Like    About    Share    Add to

### Report this video

**What is the issue?\***

- Sexual content ?
- Violent or repulsive content ?
- Hateful or abusive content ?**
- Harmful dangerous acts ?
- Child abuse ?
- Spam or misleading ?
- Infringes my rights ?
- Captions report (CVAA)

**Timestamp selected:**  
2 : 20

**Please provide additional details about:**  
 Hateful or abusive content > Promotes hatred or violence  
  
 500 characters remaining

Flagged videos and users are reviewed by YouTube staff 24 hours a day, seven days a week to determine whether they violate Community Guideline. Accounts are penalized for Community Guidelines violations and serious or repeated violations can lead to account termination. If you would like to report a channel, please click here

\* Required    **Submit**

**To flag a comment**, users click on a drop-down menu to the right of the relevant comment, and are able to select "Report Abuse". They are then offered a list of reasons for the complaint.

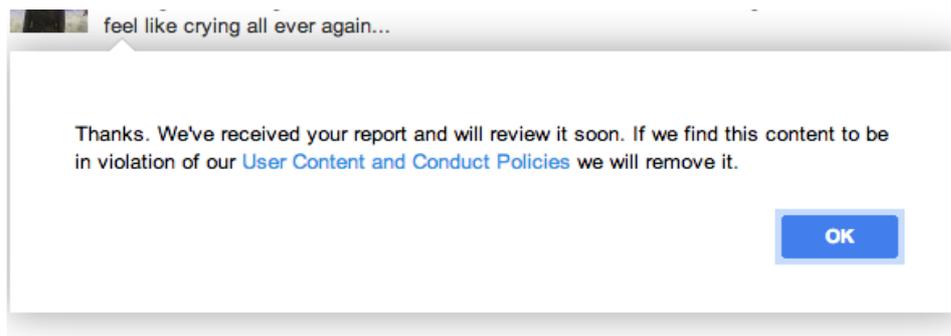
view all 0 replies

**What kind of abuse are you reporting?**

- Unwanted commercial content or spam
- Pornography or sexually explicit material
- Hate speech or graphic violence
- Harassment or bullying
- Copyrighted material
- This account might be compromised or hacked

ArkVie Llorca 18 minutes ago

The reporting process is much simpler for comments; once the user selects the type of abuse they are reporting, the report is immediately submitted to YouTube.



There is also a Reporting and Enforcement Center<sup>9</sup> where users can make a more generalised report or report about specific content or a specific user. This is a different means of getting at the reporting process but is essentially the same thing as flagging a video or comment. An important difference is that this allows an individual to report a user, rather than a specific video.

The person making the complaint is asked to choose from one of five issues. Should they choose hate speech, they are invited to specify what category of hate speech they are alleging. Gender is among the relevant categories.

**Safety and Abuse Reporting**

**What is the issue?**

- Harassment and Cyberbullying
- Impersonation
- Violent Threats
- Child Endangerment
- Hate Speech Against a Protected Group
- None of these are my issue

---

**Hate Speech**

We encourage free speech and defend everyone's right to express unpopular points of view. But we don't permit hate speech (speech which attacks or demeans a group based on race or ethnic origin, religion, disability, gender, age, veteran status, and sexual orientation/gender identity).

Please select the protected group:

- Race or ethnic origin (examples: African American, Pacific Islander)
- Nationality
- Religious affiliation (examples: Muslims, Christians)
- Sexual orientation/gender identity (examples: straight, gay, bisexual, transgender)
- Veteran status
- Disability
- Gender
- Age

### Privacy-related complaint

This process is for users who believe that their personal information is being revealed by another YouTube user. Users can report the infringement to YouTube, who will mediate between the uploader and the complainant to try to resolve the issue.

Unlike for flagging or reporting content, an individual does not need to have a Google account to make a privacy complaint. They will require an email address, although YouTube's Privacy

<sup>9</sup> [www.youtube.com/yt/policyandsafety/reporting.html](http://www.youtube.com/yt/policyandsafety/reporting.html)

Guidelines provide for third-party reporting where individuals do not have access to a computer. The Guidelines note:

### **First-party claims required**

We do not accept claims on behalf of third parties except in the following situations:

- the individual whose privacy is being violated does not have access to a computer
- the individual whose privacy is being violated is a vulnerable individual
- you are the parent or legal guardian of the individual whose privacy is being violated
- you act as a legal representative for the individual whose privacy is being violated

We will not accept privacy complaints filed on behalf of:

- other family members (e.g., husband, wife, cousin, brother, sister)
- co-workers or employees (individuals must report themselves)
- a company

With respect to privacy complaints, YouTube first suggests that a reporter contact the user who is perceived as infringing on their rights, and ask them to remove the video directly. Documenting this attempt is important in case the user refuses and further action is required. If a reporter is uncomfortable contacting the other user, then they can make a privacy-related complaint via the Reporting and Enforcement Center.<sup>10</sup>

In making a privacy-related complaint, YouTube's Privacy Guidelines specify:<sup>11</sup>

- For content to be considered for removal, an individual must be uniquely identifiable by image, voice, full name, Social Security number, bank account number or contact information (e.g. home address, email address). Examples that would not violate our privacy guidelines include gamer tags, avatar names, and address information in which the individual is not named.
- To be considered uniquely identifiable, there must be enough information in the video that allows others to recognize you. Please note that just because you can identify yourself within the video, it does not mean you are uniquely identifiable to others. A first name without additional context or a fleeting image, for example, would not likely qualify as uniquely identifiable.

---

<sup>10</sup> [www.youtube.com/yt/policyandsafety/reporting.html](http://www.youtube.com/yt/policyandsafety/reporting.html)

<sup>11</sup> [www.youtube.com/t/privacy\\_guidelines](http://www.youtube.com/t/privacy_guidelines)

Before someone can proceed with the complaint, they are asked to read the Community Guidelines, and consider whether they should instead flag the video for harassment, and are warned that abuse of the privacy complaint process may result in account suspension. The Community Guidelines, however, do not provide any in-depth information as to the difference between harassment and a privacy complaint.

Should a user persist with a privacy complaint, they go through a six-step process. After confirming that they would like to commence a privacy complaint, they are first asked whether they are being harassed, whether they would instead like to contact the uploader of the content, whether they would instead like to flag the video, and whether they have reviewed the Community Guidelines, and then are warned that abuse of the privacy process may result in suspension.

## Privacy Complaint Process: 2 of 6

### Are you being harassed?

Having your privacy violated can often feel harassing by nature.

If someone is maliciously posting a video of you or directing insults at you, this may be considered harassment.

[I would like to learn more about Harassment](#)

[I still wish to submit a privacy complaint](#)

## Privacy Complaint Process: 3 of 6

### Contact the Uploader

If you think that a video posted to YouTube violates your privacy, begin by contacting the uploader to let them know. You can use our [private messaging feature](#) to do this. It's possible that the uploader isn't aware that someone feels uncomfortable with a video or comment that they've posted.

**Please contact the uploader before you decide to file a complaint.**

[Continue](#)

# Privacy Complaint Process: 4 of 6

## Flag the video

Does this video or comment violate your privacy or is there something else about it that you find objectionable? If you've found material that violates YouTube's [Community Guidelines](#), please use the flagging feature beneath the video to bring it to our attention.

I have reviewed the Community Guidelines

I have not reviewed the Community Guidelines

# Privacy Complaint Process: 5 of 6

## Abusing the privacy process may result in account suspension

If you're using this contact form to harass other people on the site or to create false privacy reports, your YouTube account may be suspended.

I understand that I may lose my account if I submit a false privacy claim.

Continue

If they persist beyond these steps, they are then entitled to make a complaint about the revelation of either their image or name, or their personal data.

# Privacy Complaint Process: 6 of 6

## What personal information is being revealed?

Your image or full name

(Image or full name refers to still images, audio and video footage or text that uniquely identifies you.)

Your personal data

(Personal data refers to an individual's Social Security number, contact information (e.g., home address, email address), credit card number or other financial information that can be used to uniquely identify or contact you).

Users are then required to fill in an online form with information about their complaint. They are informed of the process for dealing with their complaint.

# Submit a Privacy Complaint

To notify the uploader of the privacy complaint and provide them an opportunity to remove or edit their video, we ask that you fill in the information below. Your privacy is respected in this process. At no point will your name or contact information be released to the uploader without your consent.

The uploader has 48 hours to take action on the complaint. You will be notified if the uploader removes the video or if YouTube has taken action on the video. Our communications to you about this process will be sent to your registered email address. Add support@youtube.com to your spam filter to ensure you receive these messages.

Please complete the following required information to the best of your ability

Your legal first name: \*

Your legal last name: \*

Country \*

## Review of complaints

There is no publicly available information about how content- or privacy-related complaints are dealt with internally by YouTube. Our interview with YouTube revealed that there are a number of different stages in the complaint process, depending on whether the complaint relates to content or to privacy:

| Content-related complaint   | Privacy-related complaint   |
|---|---|
| <ol style="list-style-type: none"> <li>1. Complaint is considered by a reviewer, who assesses the complaint against the Community Guidelines.</li> <li>2. If the complaint is a standard one, a decision is made by the reviewer either to               <ol style="list-style-type: none"> <li>a. Accept the complaint, remove the content and issue a “strike” against the user;</li> <li>b. Reject the complaint, and leave the content online; or</li> <li>c. Acknowledge the concern and restrict the access of the content to over-18 only.</li> </ol> </li> <li>3. If the complaint is high-profile, or particularly contentious, the complaint will be escalated, either to               <ol style="list-style-type: none"> <li>a. A second opinion;</li> <li>b. A more experienced reviewer; or</li> <li>c. A group of experts drawn from the policy, legal and PR teams.</li> </ol> </li> <li>4. If the complaint necessitates it –</li> </ol> | <ol style="list-style-type: none"> <li>1. Complaint is notified to the uploader, who is provided with 48 hours to remove or edit their video. At no point is the name or contact details of the user released to the uploader without the user’s consent.</li> <li>2. If no action is taken by the uploader, the complaint is assessed by a review against the Community Guidelines.</li> <li>3. If the complaint is a standard one, a decision is made by the reviewer either to               <ol style="list-style-type: none"> <li>a. Accept the complaint, remove the content and issue a “strike” against the user;</li> <li>b. Reject the complaint, and leave the content online; or</li> <li>c. Acknowledge the concern and restrict the access of the content to over-18 only.</li> </ol> </li> <li>4. If the uploader tries to upload the same video again after it has been removed, their account may be suspended.</li> </ol> |

|  |  |
|--|--|
| particularly when the complaint relates to threats, harm to others or harm to oneself – YouTube will notify law enforcement. However, there are no available guidelines about how and when they do so. |  |
|--|--|

### 3.3. What are the support mechanisms in place for victims/survivors?

The YouTube Safety Center<sup>12</sup> provides a number of tips and resources for users who may be subject to violations online. On harassment and cyber bullying, the site provides the following advice:

We want you to use YouTube without fear of being subjected to malicious harassment. In cases where harassment crosses the line into a malicious attack it can be reported and will be removed. In other cases, users may be mildly annoying or petty and should simply be ignored.

Cyberbullying may include:

- Abusive videos, comments, messages
- Revealing someone's personal information
- Maliciously recording someone without their consent
- Deliberately posting content in order to humiliate someone
- Making hurtful and negative comments/videos about another person

Tips and advice:

- Pause before you post: Think seriously about how you may be perceived online and do not post anything that may compromise your reputation or safety.
- Speak up: Tell friends to stop cyberbullying and voice your opinion against cyberbullying when you see it on the site.
- Tell an adult if you remain concerned about another's actions towards you online.
- Try deleting comments or blocking the user if another user is bothering you, so they can't view your videos or leave more comments. You can also turn

---

<sup>12</sup> [www.youtube.com/yt/policyandsafety/safety.html](http://www.youtube.com/yt/policyandsafety/safety.html)

comments off for any video or manage comments by requiring pre-approval before they get posted.

- Respect people’s opinions online but know when it crosses the line. We want YouTube to be a dynamic and expressive platform but we do not want users to feel intimidated or threatened.
- Report harassing users and content via YouTube’s reporting tool

To learn about tools available to you to help you manage your experience on the site, especially abusive users, please visit our Privacy and safety settings page.

Sometimes criticism and insults can escalate into more serious forms of harassment and cyberbullying. If specific threats are made against you and you feel unsafe, tell a trusted adult and report it to your local law enforcement agency.

There do not appear to be any other victim-support mechanisms available through YouTube.

### **3.4. At what point does the intermediary collaborate with others to facilitate access to justice?**

There is no publicly available information about how or under what circumstances YouTube cooperates with law enforcement or assists users to access justice mechanisms.

### **3.5. Evolution of Google/YouTube’s policies related to technology-related VAW, 2009 to 2014**

YouTube’s policies relevant to technology-related violence against women have not changed significantly in the past five years. The Community Guidelines have remained almost exactly the same during this time, with the exception of the provision relevant to threats and harassment. The wording of that section in the Community Guidelines has evolved to incorporate more stringent language. As at 30 January 2009, the provision was worded:

“Things like predatory behavior, stalking, threats, harassment, intimidation, invading privacy, revealing other people’s personal information, and inciting others to commit violent acts or to violate the Terms of Use are taken very seriously. Anyone caught doing these things may be permanently banned from YouTube.”

In late 2010, the language was changed to the current language, which is more strongly worded:

“There is zero tolerance for predatory behavior, stalking, threats, harassment, invading privacy, or the revealing of other members' personal information. Anyone caught doing these things may be permanently banned from YouTube.”

In December 2008, the Abuse and Safety Center<sup>13</sup> was launched to educate users about online safety threats related to YouTube. The Center was eventually replaced by the current Policy and Safety Hub.<sup>14</sup>

The other major change relevant to violence against women online was the introduction, in late 2013, of a requirement that YouTube users connect with Google+, which requires the registration of a user's real name. The move is designed to reduce, in part, the anonymity previously afforded to YouTube users and commentators, in order to minimise the amount of harassment and abuse, particularly in the comments sections. However, YouTube users expressed considerable outrage at the policy change.<sup>15</sup> Real name registration is generally seen by the internet freedom community as undermining online privacy and hampering free expression, as it may lead to self-censorship and aid in state surveillance. On the other hand, in some sectors it is seen as enhancing internet safety, both by deterring inappropriate behaviour that might proliferate under the veil of anonymity, and by providing more traceable means of detecting online crime.

The change also coincided with a change in YouTube's comments policies and moderation settings, which afford a user more control over the comments posted on their content.<sup>16</sup> This was designed to limit the use of the comments section for "trolling" and abusive or offensive behaviour, which previously proliferated in YouTube comments sections. The motivation was to give users greater ability to limit the offensive behaviour by enabling them to forbid, limit or moderate comments posted in response to their content.

#### **4. Interview with corporate representative**

Respondent name: Abbi Tatton

Position in organisation: Manager of Global Communications and Public Affairs

Contact phone: +1 (202) 487 7920

Contact email: abbi@google.com

I interviewed Abbi Tatton on Thursday, 30 June 2014 for 40 minutes. Unfortunately, the phone call was unable to be recorded, but I took comprehensive notes. Abbi was unable to answer some questions and noted that it was not YouTube policy to give information on others.

#### **Responding to reports of rights violations**

When a complaint is filed by users in relation to a perceived violation of their rights, what is the internal process for responding and what departments are involved?

---

<sup>13</sup> [techcrunch.com/2008/12/11/youtube-opens-abuse-and-safety-center](http://techcrunch.com/2008/12/11/youtube-opens-abuse-and-safety-center)

<sup>14</sup> [www.youtube.com/yt/policyandsafety/?rd=1](http://www.youtube.com/yt/policyandsafety/?rd=1)

<sup>15</sup> [gigaom.com/2013/11/11/users-outraged-over-youtubes-switch-to-google-real-names-policy](http://gigaom.com/2013/11/11/users-outraged-over-youtubes-switch-to-google-real-names-policy)

<sup>16</sup> [gigaom.com/2013/09/24/youtube-adds-advanced-moderation-google-to-comments-to-get-rid-of-the-trolls](http://gigaom.com/2013/09/24/youtube-adds-advanced-moderation-google-to-comments-to-get-rid-of-the-trolls)

*This depends on the nature of the violation. With violations in the form of videos (as opposed to in the content sections), if someone sees a video they feel violates our policies – could be talking about threats, harassment, stalking, violence – they can report it via the Flag button; underneath every video there is a Flag button. This brings the user to a drop-down menu, where they can identify why it makes them uncomfortable. Flagging something doesn't remove a video, it sends the video to a review team, but at the very first stage, it gets prioritised algorithmically, based on whether the video has been previously flagged (if it has, and the flag has been addressed, the video might get pushed lower down the queue). This helps YouTube get to the right videos quickly.*

*The review teams work in several time zones and are staffed 24/7. They review against the community policies/guidelines, and can make one of three decisions: 1) leave the video up; 2) take it down, because it violates the policies; or 3) it doesn't violate the policies but YouTube feels it needs to protect users under 18 because of the content, so they modify the access requirements for the video to make it age-restricted.*

What is the usual response time? If it depends on the case, can you provide a few scenarios which would have different response times, and why?

*We don't provide estimation of turnaround times, and it isn't strict. Everything is seen to quickly. We can't say everything will be seen within an hour, particularly because they are sorted via an algorithm, but they are seen quickly.*

*Overwhelmingly, it is relatively easy to make a determination on a video straight away. Of course, some decisions are more difficult, and in such circumstances there is an escalation process – either the decision will go to a more experienced reviewer, or to a second opinion. High-profile and extremely sensitive cases will go to a group of people taken from across policy, legal and PR teams.*

How many staff are responsible for responding to complaints?

*That is not a number we give out, but there are staff across multiple time zones, and they are working 24/7 to respond to flags.*

How many staff are women?

*Not a number we give out, nor one I have available to me.*

What training do staff receive to fulfil their responsibilities? Does this include gender-sensitivity training? (What is the rationale for this?)

*Staff receive training across the board on all community policies, which include gender, so yes, they absolutely do. We also have people within our policy teams at a higher level with specialised expertise in gender, and they would contribute to high profile and sensitive decisions.*

## **Monitoring reports of rights violations**

Does your organisation keep records of complaints filed by users on rights violations?

*We keep records on the reporter, i.e. the person who uploaded the video or made the comment that was flagged, because we have a multiple strike policy – if someone has complained about a video and it's come down, it's considered a strike. If a user gets three strikes, we terminate an account.*

*In terms of records of the person making the complaint, there is no collection done. We don't require the person to provide any information about themselves beyond being logged into YouTube, which they can do in an anonymous sense.*

Please describe the record-keeping system, how long records are kept for, which departments are responsible for this data. Please provide links to/soft copies of any policies governing data retention of complaints. How many complaints were received in 2013? 2012? 2011?

*Unable to provide information.*

Are the rights violations organised into specific categories (i.e. privacy violation, human rights abuse, sexual harassment, etc.)? What are the categories?

*When users report a video, they're asked to nominate an initial category – sexual content, hateful or abusive content – after which they'll see a drop-down list of more specific complaints; one says "promotes hatred or violence". However, this is just a guide for users. Mostly, users sense something is up, but they're not always sure how to label it, so they'll just choose anything. We don't penalise users if they don't use the right label; all flags go in a prioritisation queue.*

*There is a different category of complaint, which is a privacy complaint. This is relevant when there is a video posted of you, and you're not a public figure and you don't want to be on there, you don't want to be identified. If this video is flagged, our staff will mediate between you and the uploader – we'll contact the uploader and say, you have 48 hours to edit the video to address this, then we'll review it again before we take any further action.*

Is the data disaggregated by gender? (i.e. does the organisation record the gender of the user filing the complaint as part of its records?)

*Users have to be logged in to make a complaint, but there is no reason we would necessarily know their gender, and we're not keeping information about gender.*

Please provide gender-disaggregated data on the number of rights violations by category in 2013, 2012, 2011? (If the representative doesn't have access, who in the organisation can provide this data? If the information cannot be shared, why not? Does this violate a policy?)

*Unable to provide information.*

What is the most common type of complaint received? (If they have no data, ask in the representative's opinion.)

*Unable to provide information.*

Does the organisation have a specific reporting system for violence against women?

No.

Is the possibility of developing partnerships with local women's non-governmental organisations (NGOs), grassroots organisations, research centres and so on for research and/or prevention efforts being explored?

*Yes. Firstly, Google has many partnerships across a range of sectors. Also, YouTube is trialling a new system called the YouTube Deputy Program. At this stage, it is invite only, but we would think about opening up for NGOs or organisations to apply for. Basically, YouTube started seeing that overwhelmingly, people don't flag videos; the majority of flags are made by a small group of users with remarkable accuracy. We invited a small group to have more robust flagging tools, for example they could flag in bulk. Their flags still went through the same review process however. What we found was that their flagging accuracy hasn't diminished at all, and their flagging volume has gone up five-fold. We've now started to invite some organisations we work with to trial the system, with the thinking that some groups may have specialised knowledge or expertise that may help them flag. This is possibly something that we could open up to local women's organisations or the like, if this conversation continues.*

### **Policy design**

What mechanisms are in place for community stakeholders to provide input into corporate policies and practices?

*There is no user policy forum or other permanent mechanism to input into corporate policies. But we have developed policies very broadly on the basis of what we think the community wants.*

There are an increasing number of global initiatives at the international policy level which encourage and highlight the positive strides corporations are taking to uphold and promote citizens' rights. Does your organisation participate in any of these initiatives? Which ones? How often?

*YouTube is obviously part of Google, which participates in a variety of forums. While we have different policies and policy teams, and our policies vary slightly, we are a part of Google and inform their work, so any international initiatives that Google is part of, we are a part of.*

Is your organisation a member of the UN Global Compact? Is your organisation a signatory to the Women's Empowerment Principles? Why/why not?

No.

Are there other declarations/agreements related to respecting users' human rights that your organisation supports or is a signatory to?

*Can't speak to that.*

## **Responding to offline violence facilitated by technology**

What protocols or policies are in place to deal with user complaints/reports of unwanted behaviour that may escalate to physical violence?

*If there is a specific threat – a suicidal threat or a threat against the person – we would report it to local law enforcement. However, we usually don't engage with the complainant, because they are able to remain more or less anonymous, unless they're making a privacy-related complaint.*

Under what circumstances does the organisation collaborate with state agencies (such as law enforcement, social service providers, etc.) when dealing with reports of violence? Can you describe instances where the organisation has collaborated with (a) police, (b) courts and law enforcement, (c) social services (shelters, clinics, therapy, etc.) (d) advocacy groups in relation to rights violations? Has the organisation dealt with any incidents of technology-related VAW? How many in 2013?

*Unable to provide this information.*

Are there any particular cases that your organisation sees as an important success in your efforts to protect your users' rights? Is there a particular case in which the organisation's intervention helped a woman receive access to justice after she filed a complaint? (i.e. where the intervention helped stop abuse from happening, or helped in the prosecution of perpetrators of violence, or ...)

*Unable to provide this information.*

ISBN: 978-92-95102-20-0

APC-201407-WRP-R-EN-DIGITAL-211

Creative Commons Licence: Attribution-NonCommercial-NoDerivs 3.0

<http://creativecommons.org/licenses/by-nc-nd/3.0>