Braving the Backlash

A guide to handling hate on
your brand's social channels

we
are.
social

Contents

Contributors

Research team:
Mobbie Nazir, Chief Strategy Officer, We Are Social, **Charlie Cottrell**, Head of Editorial, We Are Social, **Ray Murphy**, Senior Editor, We Are Social, **Cristina Sarraille**, Senior Strategist, We Are Social, **Andre van Loon**, Research & Insight Director, We Are Social, **Nick Bennett**, Senior Project Manager, We Are Social, **Helena McAleer**, Events & Innovation Manager, We Are Social

Research participants:
Fadzai Madzingira, Policy Associate, Facebook **Chris Pitt**, Head of Marketing, HSBC UK **Jamie Buxton**, Head of Social, HSBC UK **Richard Burgess**, Social Media Manager, Audi UK **Stephen Mai**, Chief Content Officer, Boiler Room **Matt Scarff**, Director of ITV Creative and ITV Experiences, Co-Chair of ITV Pride, ITV **Iain Walters**, Director of Marketing, Pride in London **Cai Wilshaw**, Head of Corporate Partnerships, PinkNews **Aimee Challenor**, Equalities (LGBTIQA+) Spokesperson, Green Party of England and Wales **Kate Dale**, Head of Campaign Strategy, Sport England **Dr Jillian Ney**, Digital Behavioural Scientist, DRJN **Susie Hanlon**, COO, The Mandy Network **Roberta Lulli**, Task Force Against Hate Speech, Amnesty International **Shabana Kausar**, Violence against Women and Girls Tri-Borough Strategic Lead, Royal Borough of Kensington and Chelsea **Ray Rahmati**, Senior Market Director, Head of Agency Partnerships, Spredfast Inc. **Anurag Abinashi**, Director, Market Strategy & Agency Partnerships Lead (EMEA), Spredfast Inc.

Introduction

01

Hate on social media is a growing problem, and one that's being largely ignored by the advertising industry. The global rise of populist politics and Britain's vote to leave the European Union have contributed to the normalisation of vitriol, both online and offline. As the UK's Crown Prosecution Service said in a recent announcement during its crackdown on digital hate crime: "the internet and social media in particular have provided new platforms for offending behaviour."

Approximately 40% of Americans have personally experienced online harassment, according to the Pew Research Centre, while 62% consider it a major problem. The Anti-Defamation League, an international Jewish non-governmental organisation, found that in the run up to the 2016 US presidential election, more than 2.6 million tweets were sent containing 'language frequently found in anti-Semitic speech', which generated 10 billion impressions; in the month following Donald Trump's victory, the Southern Poverty Law Centre reported a dramatic spike in Islamophobic pictures and memes circulating on social media.

When you start to examine the problem of online hate more closely, you soon see that it affects all marginalised groups across racial, religious, gender, physical ability and sexual divides. But for many social media users, acts of hate aren't always easy to identify, especially when they themselves are not the target of the aggressor. These malicious acts can be carried out on the threads of popular internet forums, in the online comments section of our favourite newspapers or on the social pages of some of the world's biggest consumer brands. Not every act of online hate is visible, but each one is an assault on the very values of inclusion and equality.

As advertisers' comms become more diverse, and as more brands choose to stand behind issues like LGBTQ+ pride, disability and women's rights, they are, unfortunately, exposing themselves to a greater risk of triggering hateful comments from their followers and users they're targeting.

So why, then, are brands that claim to support diversity so quiet on this issue of online hate? We believe it's because they fear the possibility of backlash when trying to challenge it. As a socially-led creative agency that sees hate across our clients' channels every day, we think this needs to change—so we've written this white paper to arm brands with the tools they need to do a better job of handling hate on their social channels. We want to help brands publicly stand up against hate instead of simply deleting or hiding it, because sweeping the problem under the carpet does not make the problem go away, especially for the communities that are being attacked.

Brands are in a powerful and influential position, and can, if they choose to, turn the tide on hate, which can change culture for the better. If you want to say you were part of it, read on.

> " **If you have created a space, and you've created a conversation, you have a responsibility to the people who come to that platform.** "

Kate Dale
Head of Campaign Strategy, Sport England

Our Research

To gain a deeper understanding of the problem of hate on social media, we spoke to a variety of parties who are on the receiving end of it.

We interviewed major brands that are already doing an excellent job of braving the backlash on their channels; we surveyed community managers and content moderators at some of the world's biggest brands to hear how they currently handle hate on their social channels; we spoke to policy exerts at social platforms and community management software companies; and, perhaps most importantly, we ran roundtable discussions with community leaders and influential commentators from the LGBTQ+, women's equality and people of colour movements.

We wanted to include the viewpoints of representatives of some marginalised communities, because we believe the impact of hate and the response to it on social cannot be fully understood without considering these perspectives.

"Platforms, such as Facebook, Twitter and YouTube have brought huge benefits to society. But... there are growing concerns about some of the ways [they] have impacted our lives and the overall wellbeing of our societies. In some cases, these new platforms have been used to exacerbate, fuel, and deepen the divisions within our communities. "

Sadiq Khan, SXSW Festival, 2018

02

Community Guidelines

03

Uniting against hate

Social platforms are ramping up their efforts to tackle hateful comments, through policymaking and increased investment in both technology and human moderators. This comes with challenges.

The platforms' challenges

Cultural: Platforms' community guidelines can't defend against all forms of hate because they have to be broad enough to have global application; while certain language and behaviour may be acceptable in some cultures, it could be considered offensive in others. Drilling down into local context requires a serious investment in technology and human moderators with local expertise.

Technology: Artificial intelligence plays a key role in spotting hateful content, but platforms don't publicly explain the weighting of technology versus human involvement in the content moderation process. The biggest challenge for AI, currently, is understanding cultural cues and context in speech.

Human: All social platforms rely on their community to flag hateful content, but Facebook, Twitter, and YouTube employ human moderators to identify and deal with it. High reliance on human moderators means harmful content stays on the platforms for longer, doing more damage to those who view or interact with it.

Political: In May 2016, Facebook, Microsoft, Twitter and YouTube signed up to the European Commission's 'Code of Conduct on countering illegal hate speech online' which saw them agree to do 'further work on improving the feedback to users and being more transparent towards the general society'. More specifically, the agreement included commitments to review content that had been flagged as hateful within 24 hours and, where appropriate, remove it if it breaches EU or national law. Penalties for not meeting this target include hefty fines, although the amount varies from one country to another. Increased pressure from governments and regulators means platforms tackle hate Wild West style: 'shoot first' (i.e. remove the reported content) then ask questions later. It is important to note that platforms are only taking this cautious approach to more extreme types of hateful content wich breaks their community guidelines—they are not doing this for more subtle forms of hate, hence the reason for this white paper.

Code of Conduct in numbers

81	**70**	**42**	**1/2**	**24**	**26**
per cent of reported content reviewed within 24-hours	per cent of the content flagged to the platforms in 2018 was removed.	per cent increase of hateful content being removed since 2016	of hateful content flagged by users was hosted on Faecbook	per cent of hateful content flagged by users was found on YouTube	per cent of hateful content flagged by users was found onTwitter

*Data based on Facebook, Twitter, YouTube and Microsoft activities during a six-week period in November and December 2017. Source: Code of Conduct on countering illegal hate speech online:, EU Commission

Platform	Policy summary	Background	Content Flagging process	Distinguishing factors
Facebook	Facebook defines hate speech as a direct attack on people based on 'protected characteristics'—race, ethnicity, national origin, religious affiliation, sexual orientation, sex, gender, gender identity and serious disability or disease. Content that expresses support for hate groups, terrorists or criminals will also be removed. It won't remove content that violates their Community Standards if it's considered newsworthy, significant, or if letting it remain on the platform is in the public interest.	The platform has the longest and most comprehensive definition of what harmful content is, how it gets flagged and why it's removed. To create these Community Guidelines, Facebook worked with academics, legal experts and community groups to understand how they could protect users from harm while maintaining the core values that attract people to the platform in the first place. Facebook continues to work with these groups to regularly update these policies.	Combined with AI, Facebook's 8000 human moderators review content in more than 100 languages, on a 24/7 basis. As well as moderators, it relies on users to flag harmful content that breaches its Community Standards	In May 2018, the world's biggest social network launched 'Facebook Forums: Community Standards': a series of public events held in countries including Germany, France, the UK, India, Singapore, and the US. It invites communities and policy specialists to come and feed back directly on Facebook's community guidelines, which informs regular updates to the policies.
Instagram	Instagram's policy states that it will remove content that contains credible threats or hate speech, content that targets private individuals to degrade or shame them, personal information meant to blackmail or harass someone, and repeated unwanted messages. The platform might allow hateful speech to be shared when it challenges it or raises awareness but in those instances they demand the users to express their intent clearly.	Without mentioning specific moderation tools, the photo and video-sharing platform suggests it has a global team that reviews reports or flagged posts from the community, working as quickly as possible to remove content that doesn't meet its community guidelines.	On Instagram, a large part of the responsibility for reporting hateful content falls to the user who can report speficic posts, block users or accounts, and delete comments that are hateful. However, these options are available only on the mobile app, not on the desktop version.	Instagram's community guidelines were updated in May 2018 at the same time as Facebook's, one of the differences being users are not 'forced' to use their real names. Recent failures such as removing a gay couple kissing photo show how easy it is to report and take down content for the wrong reasons.
Twitter	Twitter doesn't allow content that promotes violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease. It also does not allow accounts whose primary purpose is to incite harm towards others on the basis of these categories.	Twitter has been under a lot of pressure from media and NGOs in the past year to combat the trolls on its platform. As a direct result, its community standards were updated at the end of 2017.	Although it doesn't publicly discuss the use of AI or human involvement in tackling hateful content, the media reported on a proposed Twitter policy that would reduce the reach for accounts posting harmful content (aka The Twitter 'Jail').	For high-profile users who breach the community guidelines, Twitter retains the power to remove the 'verified account' label (see the case of Richard Spencer in the US), remove followers or even delete the accounts (especially Alt-right, white supremacist and nazi propaganda accounts).
YouTube	For YouTube, hateful content promotes violence against or incites hatred against individuals or groups based on attributes, such as: race or ethnic origin, religion, disability, gender, age, veteran status, sexual orientation, gender identity. Criticism of a nation-state is generally ok, but if the primary purpose is to incite hatred against a group of people based on their ethnicity, or if it promotes violence based on any of these core attributes, like religion, it violates YouTube's policy and will be removed.	YouTube has been shaken by a few scandals due to its young Creators pushing the boundaries of what's acceptable, with brands asking for YouTube to do more to tackle hateful speech.	Staffers review hateful content that's been flagged by users 24/7. The video platform has also developed a 'Trusted Flagger' program in collaboration with several NGOs. Content flagged by Trusted Flaggers is not automatically removed but, because of their high degree of accuracy, content reporte by Trusted Flaggers is prioritised for review.	YouTube is the only platform punishing users who abuse their reporting tool; if they consider it was used in an abusive matter, they give that user strikes or warrant termination.

" We take great pride in spending a lot of time writing out those policies and testing them with our community moderators before launching them… I'm also a big believer in encouraging groups and pages [who] would like stricter guidelines, to create those. "

Fadzai Madzingira
Policy Associate, Facebook

Why All Brands Need A Policy

04

Overview

The hate brands receive on their social pages generally falls into one of three catergories. First, the kind of hateful comments or content that breaches platforms' community guidelines and, in the most extreme cases, breaks local laws. As discussed in the previous section, this will usually be dealt with by the platforms and, where necessary, escalated to law enforcement agencies.

Secondly, you have users voicing unfavourable opinions which are critical of the brand or its rationale for supporting said cause, without using language that's discriminatory, violent or threatening. If your brand encourage freedom of expression, you will likely let this be.

Finally, there's a third form of hate that can be the most challenging for brands to navigate: the grey area between hate speech and free speech. This covers hateful comments on a brand's social pages that offend an individual or a community based on their race, gender, sexuality, religion or physical ability and could damage and offend marginalised groups who see it, but might not breach platforms' community guidelines.

We believe more needs to be done to address this grey area. In fact, we believe that all brands with a presence on social media—not just those brands that get behind a purpose or 'divisive', progressive causes—should have a publicly-available anti-hate policy that goes further than the platforms' own guidelines. As publishers of content and leaders of communities, brands have a responsibility to protect those who might see comments left on their page, from harm.

Silence is complicity

Our community manager survey suggests that the majority of brands (89%) are silencing hateful comments by deleting or hiding them, instead of publicly standing against them. Removing hateful comments without any providing any explanation to the offender or, more importantly, to the offended, is problematic for several reasons.

First of all, if users aren't aware of the rules that govern your social media pages, they may not be aware why their comment has been removed—it may also risk looking like you're censoring their freedom of expression.

Second, when you sweep hateful under the carpet, you are ignoring the harm caused to those communities targeted by the hate. If you don't tolerate hate on your pages, show it. Finally, research shows that when trolls are censored, they simply migrate to other pages. Therefore if you are silent in your response to hate, you are complicit in its continuation.

Your social pages are your shop floor

We believe that your brand's social media pages are representative of the business and its values. Most brands would actively stand against and denounce hate elsewhere in their business—their warehouses, their call centres, their shop floors — so why would you silence it on your social media pages?

Brands Are Silencing Hate On Social Media

Ignore 88.7%   **11.3% Respond**
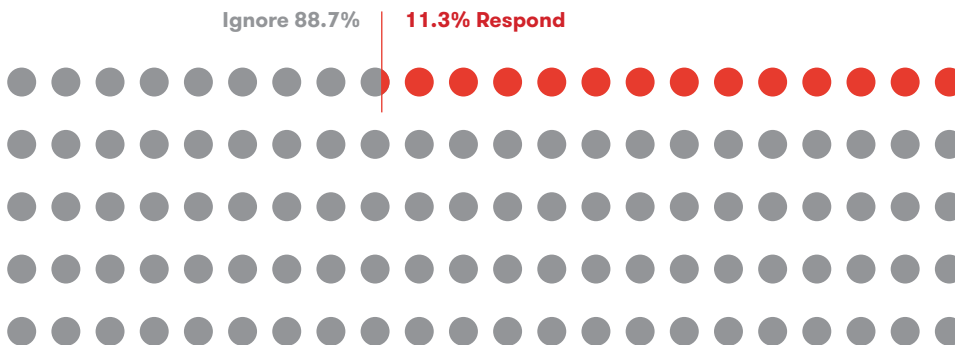
In Brief

- The majority of brands delete or hide hateful comments
- Quietly deleting hateful comments isn't enough
- Brands should publicly stand against hate on their pages

" What's really helpful, from a brand perspective, is to have a really clear direction on what you stand for and what your purpose and role is within the communities that you serve. **"**

Jamie Buxton
Head of Social Media, HSBC UK

How To Create
Your Anti-hate Policy

05

Building your policy

Creating your own anti-hate policy is perhaps the most straightforward way of standing against the problem of online malice. It will not only show the public and marginlised communities where your brand stands on the issue, but it will also guide your page moderators in their response to hateful comments, and discourage hateful sentiment overall.

As a starting point, we recommend using social media platforms' community guidelines as a basis for your anti-hate policy. These vary in size and scope—Reddit's is 36 words, while Facebook's stretches to 25 pages. All contain useful phrasing on the type of language and behaviour that won't be tolerated on their platforms. But in order to create rules for your brand's social channels that go further than the platforms' guidelines, you can add in additional wording from your brand's existing internal HR or operational policies. This could be an employee code of conduct, rules for your retail spaces, or customer care policies for your call centres. Anyone reading your policy should be clear on the kind of enviornment you want to create for users.

Make it public

Publish your policy on your social media pages so everyone is clear on the rules:

- You could post the full policy, a condensed version or a URL to the full version
- Facebook advises communities on its network to highlight their page rules by creating a post and pinning it to the top of their feed. You can do this on Twitter too
- On YouTube, you can publish your anti-hate rules as a comment on your brand's post and pin the comment to the top
- On LinkedIn, you could post your house rules on the 'About us' section on your brand page
- When running a campaign on your social channels that's likely to trigger backlash, we recommend posting a reminder of your anti-hate policies as a comment (pinned, where possible)

Case Study: Audi Uk — Our House Rules

Our Social Media House Rules: concerning all Audi UK social channels.

Audi experiences are meant to be shared. Our social channels exist for that very purpose. We want to see your photos, watch your videos and hear your opinions. We'll keep you up to date with to-the-minute Audi UK news and post content for you to enjoy. In return, your thoughts and constructive feedback are always encouraged.

From 9am to 11pm (GMT), seven days a week, our social media team are available to answer your questions and help with anything Audi UK-related. We'd rather not delete anything, but will remove posts that are irrelevant to our community. Disrespecting other Audi UK followers won't be tolerated. And we ask that you always comply with each respective channel's terms of use.

We want everyone to enjoy the discussion, but we will, in extreme cases, block users and delete comments of the following nature:

- Abusive, threatening, indecent, explicit or discriminatory language and content
- Intentional interruption or disruption of discussion
- Links that direct people to irrelevant sites or adverts
- Spam or content that has no relevance to Audi fans
- Repetitive content
- Corrupt or illegal content
- Commercial solicitation or requests for donations

And remember, our social media pages are public. Anything you post on Audi UK channels is visible to anyone who visits them. Photos of your vehicle are uploaded at your own risk.

We know that for most of you, these things go without saying. Such behaviour is rare; the above rules help to keep it that way. Carry on sharing your passion for Audi. We look forward to hearing from you soon.

“ I do not think the responses are for the troll, I think the responses are for the people that are being attacked. **”**

Stephen Mai

Chief Content Officer, Boiler Room

Responding To Hate: The Three Rs Model

06

Overview

Without having a proper system in place, it's difficult to know how to respond to hate when it occurs. So we've devised a model called 'Report-Reply-Remain' to help you handle hate with confidence.

Categorising hate

We recommend recording all negative and hateful comments in a spreadsheet and scoring them out of ten—one being the least hateful and offensive to marginalised communities, and ten being the most hateful and offensive. The comments can then be put into the following three buckets, depending where your brand decides to set its threshold:

1
Remain

Comments of this nature are allowed to remain on your page. These might be critical of your brand, communities or individuals, but are not discriminatory, violent, sexual or threatrening.

2
Reply

Hateful comments that should be publicly challenged. The type of reply your brand uses should be guided by your brand's anti-hate policy and could be one or more of the following:

Creative reply: Occasionally, there might be an opportunity to turn hate into a creative opportunity. US snack brand Honey Maid received hateful comments when it launched 'This is Wholesome'— a campaign celebrating all families. In response, it commissioned two artists to print off the hateful sentiment and turn them into an artwork, before releasing it as a follow-up online video.

Humorous reply: A funny and well-timed reply can outshine a hateful one. But this will only work where it fits with a brand's tone of voice. When promoting its popular series 'RuPaul's Drag Race', Netflix used a humourous to reply to shut down discriminatory comments, which in turn gained many more likes and shares from its users.

Collective reply: Some brands don't have the resources to respond to all of the hate it receives. Where this happens, you could respond to a group of hateful comments with a single reply. South London's Merton Police did this when it received backlash after publishing a Facebook post celebrating the Islamic holiday Eid. The police force's reply said: "This is our community and we respect EVERYONE in it", which in turn received more likes than all other comments on the post.

Template reply: Using pre-preapared template responses can help your community managers respond quickly to large volumes of hateful comments—particularly around the launch of a large social campaign. These could include a reference to your anti-hate policy or, as HSBC UK did around their 2018 Pride content, it might mention your brand's commitment to equality. However, if the templated response is repeated too often, and without variation, it can start to look insincere.

3
Report

Platforms might miss the odd extremely hateful comment—we advise reporting this. The comment should be screenshotted, as well as the user's profile where possible, in case it needs to be used as evidence at a later stage.

One More Thing

- The more diverse your community management team, the better positioned they will be create bespoke responses to hate, which could gain real support from the community that's being attacked.

Response times

Responding to every hateful comment, the instant it has been left on your page won't always be possible. When launching a campaign that contains an increased risk of triggering hateful backlash, we recommend monitoring comments closely for the first few days, and then on a less frequent basis as time goes on. For example, you may decided to review comments every 15-30 minutes on the first day of the campaign, then every two hours for a week after launch, and so on.

Outside of campaigns, we advise that you aim to respond to hateful comments within 24 hours, as per the European Commission's code of conduct. It's worth ensuring this target timeframe is clearly stated in your anti-hate policy, so your audience is aware how long it may take you to respond.

Case Study: Spredfast Inc— Handling Hate At Scale

"A lot of the time, brands don't think about the potential backlash until it happens. So instead of being proactive about it, they're being much more reactive, which, unfortunately, is not the right way to go about it [...]" — Ray Rahmati, Spredfast Inc.

Brands with small community management teams or those that receive a large amount of mentions or comments every day, should consider using social monitoring tools to help automate some of the process. One such tool, Spredfast, allows page administrators to monitor all social media channels in one place, flag comments for response and prioritise negative or positive comments made by users with a big following. The platform contains a 'Spredfast Intelligence' tool that searches a brand's channels and surfaces comments that include words from a predetermined list.

In January 2018, Spredfast helped launch 'We Counter Hate'. Using machine learning and Spredfast's existing platform, the project tracked when hate speech was shared on Twitter, based on a list of over 100 pre-agreed trigger words.

When a hateful comment was detected, an automated system would send out a tweet explaining that the user's hateful comment had been flagged and that the brand would make a charitable donation every time the comment got retweeted, to the cause that the attacker was trying to harm. This response reduced hateful retweets by 50–60 per cent.

Here, Ray Rahmati, Senior Market Director, Head of Agency Partnerships at Spredfast Inc., explains how to handle hate at scale.

"Where I always start [...] is having the right plan in place. Really thinking proactively about what things we need to be considering before we go live... so when an issue does happen, your response is much quicker, you have the right people to get involved, you have the messaging already developed. We guide our customers to set up listening monitors early on. Any time there is a mention of those terms [on a brand's banned list], we have the ability to quickly send an alert. A lot of our customers have thresholds, so if this thing gets retweeted more than a dozen times, an alert can go to the customer care team to start engaging with the users."

"

There's something problematic about individual actions being the focal point; it's about that much broader response [...] Even if it doesn't shut things down straight away, it's about changing culture "

Shabana Kausar
Violence against Women and Girls Tri-Borough Strategic Lead, Royal Borough of Kensington and Chelsea

Authenticity And Working With Communities

07

Our roundtable groups said that many brands are guilty of jumping on the bandwagon when it comes to supporting minority causes—only talking about LGBTQ+ issues around Pride, or including token people of colour and women in their videos, without having any real diversity in their staff, agencies or production teams. This often leads to inauthentic creative campaigns that can be easily spotted by underrepresented communities and can further alienate these groups.

Another reason that so many brands get diversity and representation wrong in their comms, we were told, is that they don't do enough to involve marginalised communities in their creative, their policies or their overall marketing. However, the brands that are making a real effort to build long-term partnerships with those groups that are more likely to create content that really resonates with them.

Case Study: Absolut — Community Matters

"Working with artists in the LGBTQ+ community ensures that our advertising and creative is an authentic representation of and celebrates the community in an engaging way." – Adam Boita, Absolut

Absolut vodka has a long history of working with the LGBTQ+ community. It was the first spirits company to support LGBTQ+ issues in the US, placing ads in 1981 in 'The Advocate' and 'After Dark'—two prominent LGBTQ+ magazines of the era. It has continued this relationship through philanthropic activity, sponsorship of LGBTQ+ cultural events and brand partnerships with iconic LGBTQ+ creatives including David Spada and Tom Ford.

Here, Adam Boita, head of marketing for light spirits at Pernod Ricard, explains the brand's connection with the LGBTQ+ community:

"Inclusivity is the starting point for all our creative as it underpins our brand beliefs. We also sense-check creative with thought leaders, LGBTQ+ activists, like Stonewall, whom we work in partnership—not just on Absolut activations but with Pernod Ricard UK as a business. We choose [to collaborate with creatives] who have a natural interest and affinity to the work we are producing.

"[In 2017] we photographed people from the 72 plus countries where it is still illegal to be of LGBTQ+ orientation, to raise awareness that there is still much work to do for the community internationally. Through our efforts we managed to raise £60,000 for Stonewall to support the great work they do with activists all over the world.

"This summer we released our third consecutive bottle to support the LGBTQ community [...] 50 pence from each bottle sold will be donated to Stonewall and we are proud to be continuing our support of the vital work they do to promote equality.

"Our vision of a better tomorrow includes acceptance, not only for the LGBTQ community, but support for everyone.

"We support a healthy discourse within the community [on our social pages] that hopefully instigates positive change. We have been encouraged to see that users self-regulate when discussions arise on our platforms that cause offence. That said, Absolut have community guidelines in place and there is no tolerance for hate speech— we would flag those comments in line with platforms' standards."

"Shallow diversity is the worst kind— make sure your policies and actions support your message, or you can be found out. "

Iain Walters
Director of Marketing, Pride London, Kensington and Chelsea

Conclusion

08

So, you've created anti-hate guidelines for your social pages and published them. You've put a system in place to guide your community managers' response to hate so you're prepared when it comes. You've sense-checked your policies with some of the communities most affected by discrimination and malice online, and begun to form authentic, long-term partnerships with them. You've even gone a step further and hired a diverse team of community managers to help take different viewpoints into account. What effect with this have?

In the short term, you should see a reduction of hate on your channels (see our Spredfast case study on page 16) and you should see support from the community when you decided to take a stand against hate. In the long-term, once your followers start to become aware of your support for equality and your refusal to tolerate hate, your community should feel empowered to start self-policing hate on your page.

Over time, this is something that brands like Netflix, Absolut and Boiler Room have all managed to do on their social channels—to build a community of followers that understand that these pages are no place for hate. The advantage of this is you can spend less time responding to individual comments and your pages can become a safer space for marginalised communities to voice their opinions.

If you're a brand that geniuinely wants to stand up for equality, then this needs to go beyond representation and diversity in your adverts. You need to be prepared to follow through, from your creative to your community management. As Cai Wilshaw of PinkNews told us: "There are plenty of people who would give their time freely to tell you [...] how to be ready for any backlash. And you should be ready for it. If you're not, you shouldn't do it."
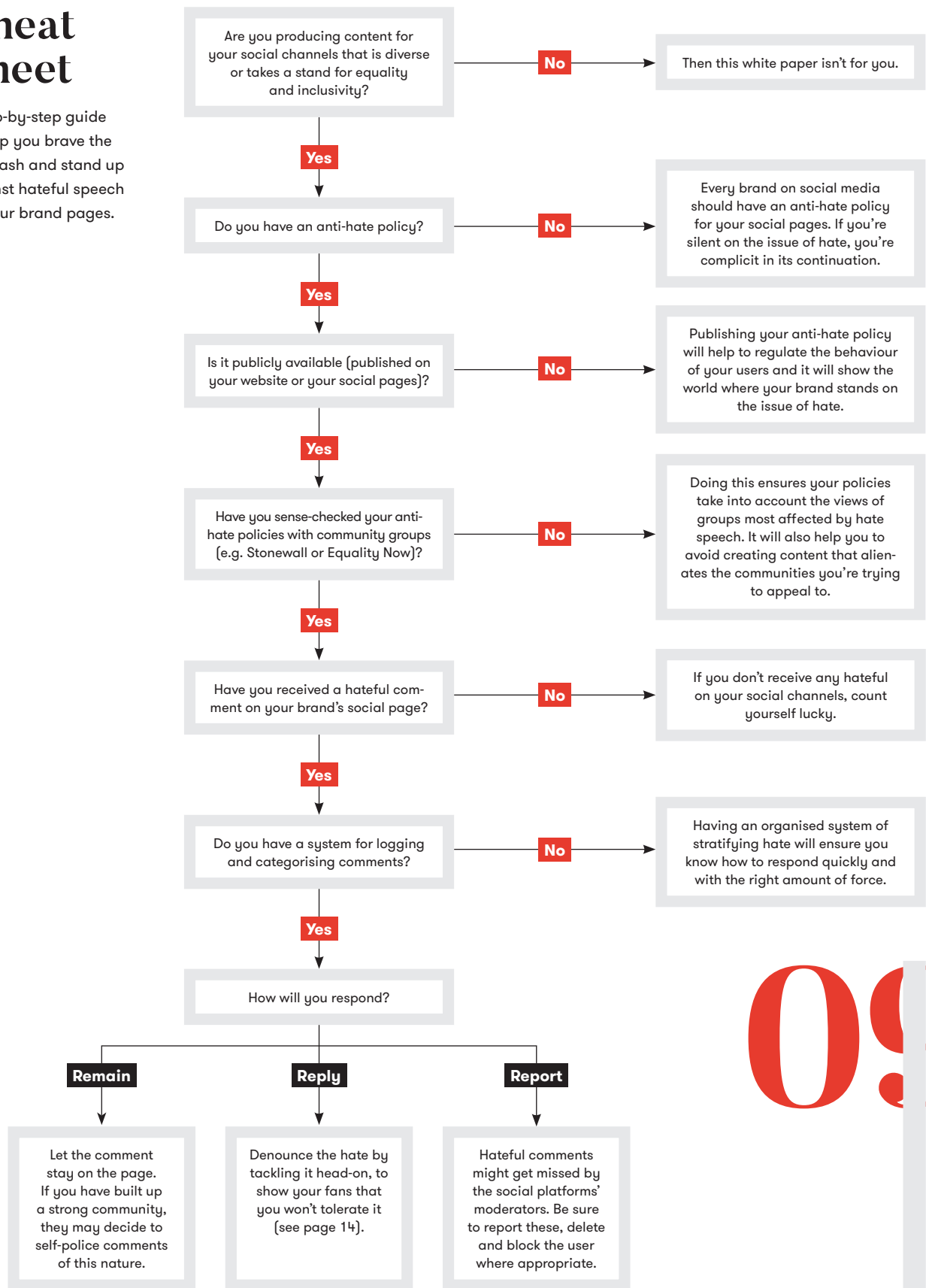
If a small handful on the world's most influential brands come together to take a stand against the trolls, we can turn the tide on vitriol and make the world, online at least, a less hateful place.

Cheat Sheet

A step-by-step guide to help you brave the backlash and stand up against hateful speech on your brand pages.

Are you producing content for your social channels that is diverse or takes a stand for equality and inclusivity?

No → Then this white paper isn't for you.

Yes ↓

Do you have an anti-hate policy?

No → Every brand on social media should have an anti-hate policy for your social pages. If you're silent on the issue of hate, you're complicit in its continuation.

Yes ↓

Is it publicly available (published on your website or your social pages)?

No → Publishing your anti-hate policy will help to regulate the behaviour of your users and it will show the world where your brand stands on the issue of hate.

Yes ↓

Have you sense-checked your anti-hate policies with community groups (e.g. Stonewall or Equality Now)?

No → Doing this ensures your policies take into account the views of groups most affected by hate speech. It will also help you to avoid creating content that alienates the communities you're trying to appeal to.

Yes ↓

Have you received a hateful comment on your brand's social page?

No → If you don't receive any hateful on your social channels, count yourself lucky.

Yes ↓

Do you have a system for logging and categorising comments?

No → Having an organised system of stratifying hate will ensure you know how to respond quickly and with the right amount of force.

Yes ↓

How will you respond?

Remain

Let the comment stay on the page. If you have built up a strong community, they may decide to self-police comments of this nature.

Reply

Denounce the hate by tackling it head-on, to show your fans that you won't tolerate it (see page 14).

Report

Hateful comments might get missed by the social platforms' moderators. Be sure to report these, delete and block the user where appropriate.

09

Contacts

Got a question about this whitepaper?

Get in touch with us via bravingthebacklash@wearesocial.net

or visit our social pages @wearesocial

we
are.
social