

DEMOS

**SILENCE,
WOMAN**

AN INVESTIGATION
INTO GENDERED
ATTACKS ONLINE

ELLEN JUDSON
AMELIA STEWART
JOSH SMITH
ALEX KRASODOMSKI-JONES

OCTOBER 2021

Open Access. Some rights reserved.

Open Access. Some rights reserved. As the publisher of this work, Demos wants to encourage the circulation of our work as widely as possible while retaining the copyright. We therefore have an open access policy which enables anyone to access our content online without charge. Anyone can download, save, perform or distribute this work in any format, including translation, without written permission. This is subject to the terms of the Creative Commons By Share Alike licence. The main conditions are:

- Demos and the author(s) are credited including our web address **www.demos.co.uk**
- If you use our work, you share the results under a similar licence

A full copy of the licence can be found at **<https://creativecommons.org/licenses/by-sa/3.0/legalcode>**

You are welcome to ask for permission to use this work for purposes other than those covered by the licence. Demos gratefully acknowledges the work of Creative Commons in inspiring our approach to copyright. To find out more go to **www.creativecommons.org**



A report in partnership with BBC Panorama.

CONTENT WARNING: This report contains discussion and examples of hateful speech, including explicit language and sexualised and gender-based attacks.

Published by Demos October 2021
© Demos. Some rights reserved.
15 Whitehall, London, SW1A 2DD
T: 020 3878 3955
hello@demos.co.uk
www.demos.co.uk

EXECUTIVE SUMMARY

Online abuse and online violence against women is a topic that has been often in the headlines, but so far, with few real solutions: platforms have introduced measures to 'empower users', to little effect; governments are arguing over the definitions of harm before enacting any regulations; civil society are doing their best to support those targeted by online abuse, but struggling to keep up.

Part of the problem is that 'online abuse' is often talked about in generalities, as a homogenous entity which we can simply decide to stamp out. The reality, however, is much more complex.

In partnership with BBC Panorama, we investigated the gendered abuse that contestants on reality shows *Love Island* and *Married At First Sight UK* faced online; we also investigated how women who were not in the public eye were being impacted by online gendered abuse through their reporting online of their own experiences. Investigating thousands of social media posts across multiple platforms, we here present our findings: the key commonalities; crucial challenges; and possible ways forward, for tackling online gendered abuse.

KEY FINDINGS:

- Women online are being driven out of communities they value and subject to daily abuse: there is an overarching sense that they are not seen as deserving to exist safely in these spaces
- Defining gendered abuse as a tight category of illegitimate speech is unlikely to be a successful mitigation strategy: particularly as definitions are weaponised as excuses for perpetrating abuse
- Gendered abuse disproportionately targets women: while gendered abuse of both men and women draws on stereotypes that are rooted in misogyny
- Gendered abuse online is an issue that pervades the whole online ecosystem: it is not confined to one topic or one platform
- There exists a vicious cycle: abuse is commonplace, so it is predictable. This means it is seen as avoidable, and therefore the target's responsibility to manage; so it continues to be commonplace

Our findings are reported in full below.

PART I: GENDERED ABUSE OF CONTESTANTS ON REALITY TV

KEY FINDINGS

- Gendered abuse online is a spectrum rather than a tightly defined category
- Gendered abuse and attacks online disproportionately target women in reality television shows; and many of the attacks which target men are themselves rooted in misogyny
- Women face not only attack for their perceived character, but are the subject of extreme misogynistic sexualisation and objectification
- People in these spaces acknowledge that abuse is a problem, and often want to fight against it: but they disagree about what qualifies, and face divisive fights over the issue
- This is not a problem confined to one forum or space online: it's a symptom of a wider toxic ecosystem

INTRODUCTION

Social media is an essential part of reality television - audience engagement, interactions, and memes are as much part of the culture of the series as the show itself. But it has become an unfortunate pattern that with the advent of a new reality show season, that inevitably follows an onslaught of harassment and abuse levelled against those on the show, from contestants to presenters to bystanders.

This is not just online discussions getting heated: this abuse has escalated to participants in these shows, disproportionately women, especially women of colour, being on the receiving end of extremist threats and online violence. Stars from the shows have recently [given evidence to Parliament](#) about the abuse they have faced. The mental health of people who participate in reality television has become an even more serious concern in recent years, following deaths by suicide of some of those who have taken part.

Ending this kind of abuse is one of the key priorities of the new UK proposals for digital regulation, the Online Safety Bill: however, there is division over how effective the Bill will be in leading to meaningful action from social media platforms that will reduce the effect of these harms.

We are delighted to have been able to partner with Panorama to explore in more detail how online abuse manifests against participants in reality shows, the gendered dimensions of this, and the challenges of responding to it effectively.

For this project, we looked at the online discussion around participants in reality shows in the UK in 2021, focusing on *Married At First Sight UK* and *Love Island*. These shows follow participants and their romantic relationships with others on the show.

METHODS

Married at First Sight UK is a Channel 4 programme, [described as](#) a 'bold social experiment where single people, matched by experts, marry total strangers, who they meet for the very first time on their wedding day'. Series 6 of the show aired in September 2021, with the premiere having [820,000 viewers](#) - an E4 record for an original series launch. Eight couples participated in the series, including the show's first gay couple, as well as three experts.

We collected 48,364 tweets through Twitter's public API which used the hashtag #MAFSUK, from September 28th to October 5th. This period covered the airing of the last 4 episodes, during which time couples who had remained in the 'experiment' had their final dates and decided whether to continue their relationships outside of the show or not, concluding with a reunion episode which brought back all those who had participated in the series to discuss their time on the show.

This period also saw Channel 4 experiencing [technical difficulties](#), which was the subject of much discussion in our dataset, including abuse and attacks levied against the team responsible at E4. On Monday 5th October, the night of the reunion episode, Facebook services also went down temporarily, meaning there was likely more discussion on Twitter than otherwise there might have been given the lack of alternative platforms.

Love Island is an ITV programme, a competition reality show which sees contestants aim to win the show by coupling up and winning public votes. The 7th season aired across July and August 2021, with an [original 37 contestants](#).

We collected posts and comments through the Reddit API from the two active subreddits (subforums within Reddit focused on specific topics) relating to the UK series of *Love Island* with the most members: r/LoveIslandTV (66k members) and r/LoveIsland_girls (916 members).

In total, we collected 46,588 documents (original posts and comments on those posts) from Reddit. Due to the length of these texts, we split each document into individual sentences for analysis, for a total of 89,245 documents for analysis.

SUBREDDIT	POSTS COLLECTED	COMMENTS COLLECTED
r/LoveIslandTV	985	45,434
r/LoveIsland_girls	110	59
Total	1,095	45,493

Our data collection and analysis was carried out using Method52, a suite of tools for collecting and analysing large free-text datasets developed by Demos in partnership with the University of Sussex.

Our research questions were as follows:

- How are reality show participants being discussed negatively online?
- Are there differences between how men and women in the public eye (in this case, reality show participants) are discussed online?
- How does gendered abuse online intersect with other forms of abuse?

Our approach to analysis was exploratory and iterative based on what we were observing in the data, as well as our prior knowledge of the contours of gendered abuse online. We used a variety of methods, which changed based on the findings for the dataset in question (set out in more detail below) but included combinations of the following:

- Filtering documents according to whether they contain one of a set of 'naive' keywords likely to be related to gendered abuse, reviewing the matches and then amending the keyword list to more accurately reflect terms being used in the dataset and re-filtering.
- Using unsupervised machine learning models to cluster terms that frequently co-occur in the data to identify patterns of language.
- Training natural language processing classifiers to identify relevant posts at scale (such as: identifying personal attacks against participants)
- Qualitative analysis of posts and Tweets to identify themes within relevant material, such as:
 - documents which received high or low levels of support
 - documents which mentioned a participant
 - documents which included a keyword judged as likely to indicate relevance to gendered abuse

Any posts quoted in this report have been bowdlerised, so that the sense has been preserved but not the exact words used, in order to protect the privacy of the original author. We do not include usernames or identifying references to individuals.

FINDINGS

Gendered abuse online is a spectrum rather than a tightly defined category

In reviewing and classifying posts as ‘gendered abuse’ or not, it was clear that constructing a tight and accurate definition of abuse is a significant challenge. The presence of a gendered slur alone (often used a proxy for abuse) is an inaccurate measure, as people would use these terms (such as ‘bitch’) in affectionate, supportive, or irrelevant ways. The presence of negative criticism of a person also proved too broad a definition, that would end up capturing a huge amount of legitimate discourse about a show which is publicly broadcast.

Example tweets talking about the same situation in different ways that sit on the border between legitimate discourse and an abusive attack:

<i>‘She doesn’t deserve him: he’s a really good guy’</i>	<i>‘She is just... There aren’t words for how she’s behaved’</i>	<i>‘Why’s he talking? He’s very rude’</i>
<i>‘TRASH - just like her...He deserves much better!’</i>	<i>‘I’d forgotten how much of a horrid bastard she is’</i>	<i>‘He’s such a prick: I don’t want to see his face ever again’</i>

Another element that made the definition process more challenging was the legitimate discussion of topics such as emotional and physical abuse which often arose in the context of behaviours which viewers felt could be symptoms of abusive tendencies. However, the spectrum of responses to this behaviour ran from expressions of concern to extrapolation and sensationalization that crossed over with personal attacks against an individual’s character.

‘I’m genuinely very concerned for her...he’s really unsettling’

‘He’s that creep in a club that you try to get your friends away from but he follows you around after buying you a drink’

‘If you said he’d snapped at last, and now he was wearing her skin, I wouldn’t be shocked at all’

We also commonly saw people using harmful gendered tropes even when trying to speak in support of someone - such as in ‘taking their side’ against another participant on the show who then became the subject of an attack.

‘She’s still trying to gaslight him: I hate her, what a bitch’

Examples of extreme or violent abuse were limited: however, this is likely due to moderation and removal of extreme content, rather than its absence altogether, and also to the fact that we were only able to access public posts. We know from participants speaking about their experience that women are frequently subject to extreme violent threats via Direct Messages.

‘She’s disgusting! WTF, man! Electric chair her!’

These factors all mean that definitively identifying a post or comment as ‘gendered abuse’ is a challenge; with limited examples at the extremes, and many which crossed over several categories of legitimate content with personal attacks and insults. This has particularly significant ramifications for how we can respond to gendered abuse: relying purely on removal, for instance, of a tightly defined category of abuse, is likely to leave a great deal of harm online untouched.

Gendered abuse and attacks online disproportionately target women; many of the attacks which target men are themselves rooted in misogyny

Given these challenges, our approach centred on what was extremely common, however, was personal attacks on people’s characters or behaviour, sometimes, but not always, accompanied by the use of gendered slurs.

These attacks were personal, unpleasant and sought to undermine the credibility or character of the target (rather than e.g. simply saying someone was not liked, or that they hoped two people on the show would end their relationship). Undoubtedly some of what we classified as ‘personal attacks’, others would count as ‘legitimate critique’. However, they weaponise gendered stereotypes to personally attack - not only criticise or express a dislike of - participants on the show: describing people in ways that would never be acceptable in person.

‘She’s been here less than a minute, and she’s as foul-mouthed and stupid as always. You can not buy class’

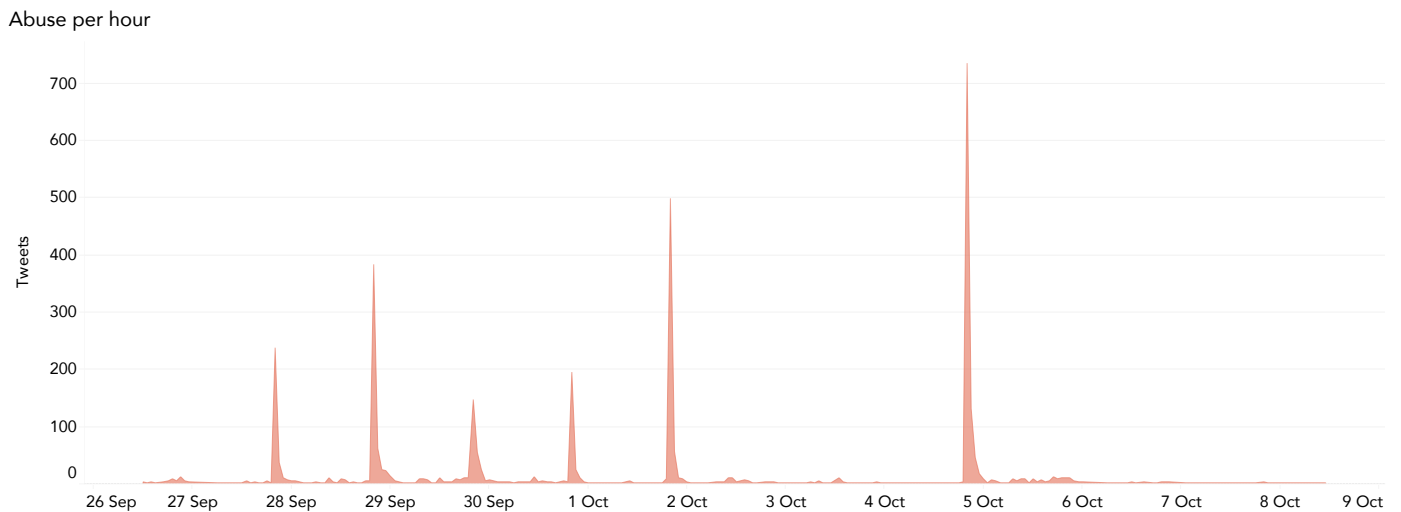
‘Fuck off, no-one is as fake as she is. Hungry for fame, self-obsessed and generally a horrendous human’

‘Here’s the shit house no. 2: she has that crazy look around her....crazy like Fatal Attraction Glenn Close’

Using this definition, we identified a subset of the *Married At First Sight* data as likely to be related to gendered personal attacks, as through a combination of iterative keyword filtering of the data, and training a natural language processing classifier to identify relevant and irrelevant posts.¹

Classified as Irrelevant: criticism of general behaviour or of relationship	<p><i>'She's really not the right woman for him'</i></p> <p><i>'The disrespect was how she treated him the WHOLE time: gtf out!!'</i></p>
Classified as Relevant: attack on personal character, appearance or motivations, leaning on gendered stereotypes	<p><i>'Those two are really toxic: my blood pressure is super high while I'm watching'</i></p> <p><i>'She is so fake'</i></p>

Volume of tweets likely to be personal attacks over time: attacks spike during broadcast but a tail continues afterwards as well

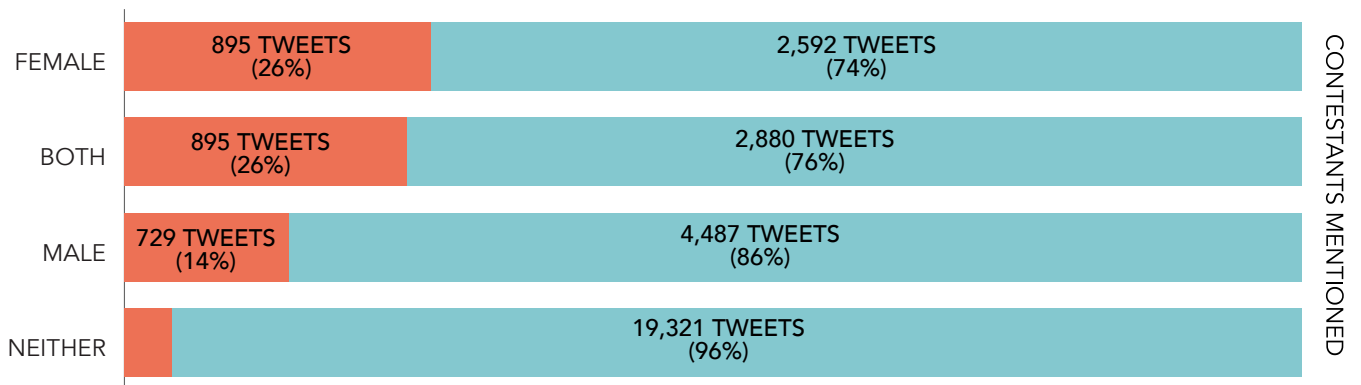


We then investigated within this dataset, and how many times women vs men participants' names were mentioned, and within the tweets that mentioned a name, who the attack was in fact directed at.

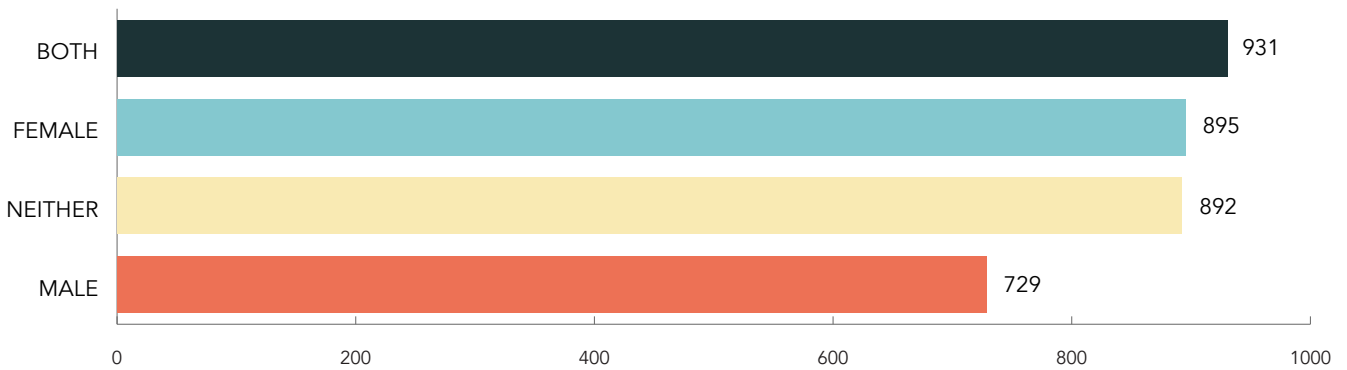
¹ Classifier accuracy: Relevant F-Score 0.714 (186 labelled), Irrelevant F-Score 0.716 (170 labelled), Accuracy 0.715

We found that:

- A greater proportion of tweets which are likely abusive mentioned women participants - even though men were mentioned more often generally



- Of tweets which were likely abusive, women were mentioned more often than men in tweets



However, being named in a tweet that is abusive is not a guarantee the named person is the one being abused. Our data suggests that not only are women participants mentioned in attacking tweets far more often than men participants: unless only a man is mentioned, women are still more likely to be the target of an abusive attack - and are often the subject of attacks even when not mentioned by name at all.

The tweets classified as likely to be a gendered personal attack against a participant were split into the following categories (a random 10% or 5% sample of which was then examined:)

Names a woman participant	39.7% attacking a woman (72/181)	4.4% attacking a man (8/181)
Names a man participant	15.6% attacking a woman (10/64)	42.2% attacking a man (27/64)
Names both a man and woman	49.5% attacking a woman (49/99)	11% attacking a man (11/99)
Names neither ²	40.8% attacking a woman (29/71)	14.1% attacking a man (10/71)

NB: a post not being coded as an 'attack' does not mean it was positive or neutral: posts were often critical or negative but not in a way that qualified as a personal attack

2 (although some names did appear e.g. spelt wrong so not identified by our keyword filter)

We also examined the gendered tropes which were being perpetuated through these attacks.

The tropes we saw repeated time and again included women as devious and violent at worst: annoying and inconvenient at best. Personal attacks levelled against participants ranged from the intended 'comedic' to the downright vitriolic.

TROPE	DESCRIPTION	EXAMPLE KEYWORDS	EXAMPLE COMMENT
Women as crazy	Women seen as unstable	Crazy	<i>'Good of the mental institution to give the girls a pass for the day'</i>
Women as emotionally volatile	Women described as excessively and negatively emotional, displaying or being driven too much by emotion	Needy, moody, desperate	<i>'So he and the bunny boiler aren't together any more - I wonder why that is...[knife emoji]'</i>
Women as devious	Women as lying, manipulative, gaslighting, abusive	Fake, snake, lying	<i>'What a toxic snake'</i>
Women as evil	Women as irredeemably bad: who deserve bad things that happen	Toxic, witch, vile, poison, banshee, demon, devil, karma, just deserts	<i>'She's a fucking cunt, and I'm not sorry'</i> <i>'Get in the bin: such gaslighting, you're a hag'</i>
Women as an impermissible inconvenience	Women as inconvenient or unpleasant to be around, unlikeable, not pleasant company/to watch	Loud, annoying, whining, screeching	<i>'She needs to turn down the volume, and if not, please can someone take out her batteries?'</i> <i>'Feeding a gremlin after 12pm and you end up with her'</i>
Women as entitled	Women as having inflated views of themselves	Self-centered, attention seeking	<i>'Just want to slap her, she's self-centered'</i>

We found similar themes in the *Love Island* discussions, criticising those on the show not only for their behaviour on the show but also afterwards, their social media posts, and so forth - with hostility towards women for their apparent entitlement or distastefulness.

'She's the most Pick Me to ever be on the show.'

'Such a beg'

The tropes used against men were also extremely often misogynistic: including men being attacked for being weak, pathetic, or soft. Even where the author

TROPE	DESCRIPTION	EXAMPLE KEYWORDS	EXAMPLE COMMENT
Men as weak	Men as womanly, weak, not masculine, overly emotional, often controlled by a woman (girlfriend or mother)	Apron strings, pathetic, sap, weak, balls, a pair	'He's a weak-ass guy'
Men as creepy or scary	Men as a threat to women: violent, weird, out of place,	Creepy, strange, weirdo, scary, skin crawl, strange, bad vibes, psycho	'She probably did not want you to be a literal psychopath but yet we are here'

was intending to express support for a participant on the show, these same tropes were being invoked: relating to men who had been viewed as 'weak' or 'emasculated': employing the trope that men should be more 'masculine' and less like a 'woman'. There was also criticism of men for being emotionally immature; creepy, or aggressive (as discussed above).

'Glad to see he's grown a pair! <3'

It also was not only main participants on the show that were in line for attack. In *Married At First Sight*, participants' family and friends also featured in several episodes. Some content was supportive (e.g. where the audience agreed with a parent's assessment of a relationship) but others attacked their character or behaviour - particularly the mothers for being overly attached, manipulative or selfish in their relationship with their child, and the female friends for being annoying or loud.

'Is her [female] friend on drugs? That expression is insane'

'His mum gives off bad vibes'

'His mum never cut the umbilical cord'

Women face not only attack for their perceived character from their time on the show, but are the subject of extreme misogynistic sexualisation and objectification

'Looking like a dirty fucking slut in that picture, what a naughty bitch - betting she would be begging for a rough fuck'

One of the subreddits we examined was primarily a place for trading pictures and screenshots of women who had appeared on *Love Island*. The most common type of post on the forum (68 posts (38%)), contained just a contestant's name and a picture.

The forum as a whole promoted the objectification of women: in particular, we found 21 posts / comments (13%) which were explicitly objectifying: 15 (9%) by mentioning body parts, and 5 (3%) imagining sexual acts performed with a contestant, including group sex with other forum members. 7 posts / comments (4%) mentioned or showed acts of masturbation by forum members; this was an implicit focal point of discussions happening elsewhere; e.g. Discord. The forum also appears to serve as a meeting place for further conversation. 18 posts / comments (11%) were about soliciting conversation with others about the women; 17 mentioned or asked for a Discord server. Many explicitly indicated that this was to 'chat about the girls' in a sexual context, or trade images.

We saw this objectification in the #MAFSUK discussion as well: users commonly criticising women's appearance and making derogatory sexual comments.

*'Get your tits out, and your flaps, they're basically on display *rolls eyes*'*

'Can anyone just pull the lips off her face pls'

'Her other lips are probably nasty too'

Not only are women disproportionately subjected to gendered personal attacks, the tropes being employed all spoke to a similar concept: women behaving in ways they 'had no right to be' - women being too loud, too annoying, undeserving of respect or happiness; too confident; badly motivated. What comes through strongly is the sentiment being communicated that women do not deserve what they think they deserve: to speak, to take up space, to have a relationship, to wear certain clothes, to look a certain way, to express themselves, without thereby incurring hatred and critique. This is an attitude that persists and is weaponised against women, undermining and discrediting them, and driving them off of online spaces where others have determined they do not deserve to belong.

People in these spaces acknowledge that abuse is a problem, and often want to fight against it: but they disagree about what qualifies, and face divisive fights over the issue

In our analysis of the data discussing *Love Island*, we found a different kind of conversation. We observed much more discussion of bullying and harassment than original bullying and abusive content: such as people chastising others for bullying, describing abuse they had seen, or talking about abuse that was occurring on other platforms. Partly this could be due to the more extensive existing discourse about the problem of online abuse relating to *Love Island*.

Our hypothesis is, however, that a significant amount of abuse or bullying comments have been removed by the platform or the moderators of the subreddits. This theory is supported by an investigation of the wider context of a subset of these posts discussing bullying (the comment thread they appeared in), which often contained comments marked as deleted by users, automated or human moderators.

'This user's always saying hate and shit, deleting it, and saying this subreddit should be spreading more kindness etc. etc. - it's bullshit!'

'Can confirm that user commented awful things & now deleted them...'

'I have the screenshots of where he was talking shit: now he's trying to deny it, it's even more ridiculous'

It could also be linked to factors such as the possibility of lengthier posts on Reddit meaning that opinions are explained with more nuance. However, the amount of comments calling out hatred implies that moderation out of abuse is a more likely explanation for its absence than restraint on the part of users.

'It isn't just this subreddit, there have been people saying in public that he should die, he's evil - directly to him in comments'

'She's been called on this subreddit loads - bitch, weak/pathetic'

However, comments were not only raising awareness of abuse: there was significant amounts of debate over what constitutes abuse, what was 'legitimate criticism', and how far anyone in the public eye had a right to complain about it - with this causing heated arguments, and users attacking each other on the forum, as well as the Islanders they were debating about.

'There's a huge difference between someone making fun of you, casually, not even swearing, and being abused to the level where it poses a risk to your mental health'

'They all get this sort of crap. It isn't right, but don't pretend he's a victim, he won £50k...I don't think he's reading this little subreddit and getting annoyed at the legit critique of his douche behaviour'

'Pls, don't play down bullying behaviour. This is gross. No-one deserves that, neither does he - death threats don't count as legitimate criticism'

'I don't see that crap saying he should die in this forum - it'd get downvoted, or removed'

'It's not criticising someone's behaviour to tell people to F off and get lost, calling them twats/pricks: that's hate. Haven't you made judgements about her character too? Or is that just criticism as well...'

In particular, the hashtag #BeKind, created in 2017 as a stand against online trolling, was the subject of much discussion: some imploring people to uphold the principle of #BeKind; others commenting on the hypocrisy they perceived in online forums where people would preach to others to be kind (particularly to their personal favourite Islanders) while engaging in hate and abuse themselves against those they didn't like.

'Literally, people will send h8 in a post/comment, then they delete it, and on the next one they post '#BeKind': the hypocrisy, lmao'

'Be Kind did not last for long, did it?'

'Bullying is a totally different thing to saying bitchy things'

This division, as well as arising between what abuse was legitimate or not, arose in the context of racism particularly, with arguments about racism in relation to *Love Island*. Arguments were common about whether contestants (particularly women) were more or less favoured by the public because of their race (either on grounds of positive or negative discrimination). There was a clear divide between those who felt that there was clearly racism in the production and reception of *Love Island*; and those who claimed that differences in treatment of the Islanders was unrelated to race and who objected to conversations focusing on race.

'It's sad that people voted for her because she is black?' they voted for her bcos they luv her!'

'If all the finalists were black except one white person - that couple would win'

'Thus the conclusion is racial bias...What has he done to deserve so few followers, when others have way higher?'

'Everyone is turning it into a race issue but he behaved the worst...'

'Bringing up institutional racism b/cos your fav doesn't win Love Island dilutes the needed conversations on racism'

'This sub can't not complain about white women for one day - I'm not accepting that I'm the one with a racial issue'

The common thread across these discussions, both of abuse and racism specifically, is a sentiment coming to the fore that those raising issues of discrimination or oppression are overreacting and exaggerating how bad things are: as well as invalidating people's experiences on the basis of what the standard or the expectation is. If something is expected - hate, criticism, prejudice - when you go into the public eye, the narrative appears to be (from one side of the aisle) that you effectively sign up for that, and hence lose your right to complain or demand more equal treatment.

This is not a problem confined to one forum or space online: it's a symptom of a wider toxic ecosystem

As well as discussions about their own subreddit - both positive and negative - there was a great deal of - often hostile - discussion about other online communities, compared to their own.

'When it became so toxic on twitter that it was upsetting me, I'm really glad I found somewhere that #BeKind was taken properly seriously'

We also saw reports of users attempting to counter or help to counter the abuse that Islanders were experiencing: either by counter-balancing it with positivity, or by warning them to stay away from certain online spaces and communities: using in some cases the same mechanism of Direct Messages that are often channels for the worst abuse and threats.

'They deserve some love, Twitter has been horrible, so I sent them a direct message'

'I just messaged her family member in DMs to try to ensure she won't look at Twitter'

This was again seen particularly in the discussion about racism against Love Islanders: there was much discussion in the subreddit about other online

communities which they perceived to be more biased than they were. In particular 'Facebook mums' who were described as voting based on racial bias and engaging in online abuse of Black contestants on other platforms, and critique of 'Fiat500 Twitter', generally younger white women on Twitter. There was more support for 'Black Twitter', including as a perceived voting bloc, and pride in how far the Black contestants on the show had gotten despite the racism that they faced.

'So devastated when I realised that Black Twitter is a tiny percentage of the whole popn'

'I hope they prove the racist Facebook mums wrong'

POLICY RAMIFICATIONS

These findings, in many ways, show that online conversations reflect back policy conversations about online abuse. The absence of a clear definition; arguments over when criticism becomes abuse, when it is justified, and when it is not, echo the debates which have been occurring in Westminster and the media over the last two years since the Government published proposals to tackle online harms - including abuse. But since these discussions are taking place on the very platforms which would come into scope of such a regulation, they have important implications for policy.

We can't fix the problem of bad speech one post at a time: we need to be thinking about one ecosystem at a time

These findings show that drawing tight lines between 'abusive' and 'not abusive' content is incredibly difficult even in individual cases, and becomes progressively more difficult as these judgements scale. Even in discussions trying to stop abuse themselves, there is a lack of consensus as to what counts as 'acceptable' discourse online. That there is overlap within discussions of personal attacks against public figures, for instance, calling them abusive and manipulative, with people sharing their own personal experiences of abuse and being subject to manipulation, goes to show why a 'just ban it' approach will not work if we are aiming to maximise user safety.

Platforms need to get better at content moderation and removal of content that breaches their terms of service. But if we rely only on trying to define specific kinds of content to support rigorous content-moderation-removal systems to tackle online abuse, we are always going to miss abusive speech and censor legitimate speech. Seeking to define our way out of online abuse is a Sisyphean task.

But doing nothing and relying on the existing lines in the sand, the threshold of illegality of speech for intervention, to reduce the harms is not an option: the speech discussed in this paper is very unlikely to be reaching any criminal threshold, but the scale means that the harm it can be causing to its subjects is significant.

So how do we square the circle of tackling online abuse? We need to think beyond content: to systems. The role of regulation should be oversee systems which increase the risk of harms arising from content to users: to require platforms to make changes to their systems that change their space and their communities, not just their content. This may be through empowering online communities to more effectively define their own norms; or changing their systems of algorithmic curation and amplification that encourage an atmosphere where antagonism, controversy or sensationalism are the norm and discussion quickly devolves into attacks and abuse. It might be supporting people to interact and communicate with others in positive ways, and disincentivising the use of those same channels for weaponisation and harassment.

And without greater understanding of the impact and nature of online abuse, attempts to tackle online abuse - whether by individuals, online communities, or regulators - will be met by the same refrains: it's 'not that bad', it's a reasonable trade for the benefits that come with public attention, and 'it doesn't count as abuse'. We need to listen to those affected by online abuse, and what the impact of their experiences has been on their lives (see part II).

And to support this understanding, to look at patterns of online abuse at scale, data access is essential. The reason we can study Twitter and Reddit in such depth compared to Facebook or Instagram, where anecdotally we know much abuse and harassment takes place, is because of the data access they grant. To facilitate this greater understanding, platforms should be compelled to provide data access to the regulator and independent researchers.

Systemic problems need systemic solutions. Addressing the problems of online abuse goes much further than regulation: it means addressing racism, misogyny; supporting digital and media literacy; building a responsible and ethical media; securing digital rights, listening to users and building and developing technology that serves them. But systems-regulation is an important part of the puzzle: and one that could come to fruition very soon.

PART II: THE SILENCING EFFECT

KEY FINDINGS:

- Gendered abuse is driving women out of communities they value
- Online 'humour' is threatening women's safety
- Gendered abuse intersects with many other forms of abuse such as racism, homophobia and transphobia
- Women's privacy is being violated and their personal information used to abuse them
- Women are attacked for speaking out about their experiences - creating a vicious cycle of abuse that women can only break by disengaging from online spaces
- Women are blamed for the abuse that they receive: abuse is portrayed as inevitable, and so it is the responsibility of the target to keep themselves safe
- Discussions about women's safety are seen as unfairly neglecting men's experiences
- Current systems of redress put the burden on women to manage the abuse they receive online and are failing to meet even minimum standards of supporting them in doing that.
- Social media platforms provide a source of solidarity and a space for women to share their experiences

INTRODUCTION

The impact of online abuse and harassment on women is increasingly being acknowledged in public and policy discussions. Yet meaningful action to tackle abuse from social media platforms has been tokenistic at best and non-existent at worst. Currently, the focus of public campaigns has often been the experiences of women in public life, who are disproportionately targeted by widespread abuse and threats. In Part I, we focused on abuse targeted at celebrities featuring on reality television shows.

However, there has been less attention paid to the impact on women not in the public eye. This paper aims to contribute to improving that balance.

Here we present the findings of a qualitative analysis of social media posts from across multiple different forums discussing women's experiences of abuse online. What we see is women speaking about being degraded and insulted; expressions of overwhelming feelings of exhaustion in the face of constant abuse; and women on the verge of cutting themselves off from online communities altogether. There is frustration at the platforms repeatedly failing to take any action against harassers, and the solace that women can find in solidarity from other users is counterbalanced by the further attacks they face for speaking up.

METHOD

We examined social media posts across five different platforms used by different demographics for different purposes: Twitter, Reddit, Youtube, Instagram and TikTok. We also compared these posts with what we found in 'traditional' media, examining the comments section of Daily Mail articles focusing on women experiencing online abuse.

In total we examined 87 posts. This is by no means a representative sample, but presents a cross-section of how women speak about their experiences of abuse online.

We sought to identify posts of predominantly women discussing misogynistic abuse that they had received online and the impact it had had on them, or where third parties were discussing the abuse that other women had received online. We also looked at how other online users of all genders responded directly to women speaking out about the abuse.

We used a variety of search terms to locate posts, including 'online abuse', #sexistabuse, #sexistsocialmedia, #sexisttwitter, 'sexism reddit', 'sexist abuse online', and 'being a woman online'. We identified these terms by beginning with naive searches such as 'online abuse' and then refined our searches based on terms and hashtags we saw being used. We also looked at online communities, such as Two X Chromosomes, which were particularly likely to have discussion of gendered abuse.

As the content we examined was posted with a reasonable expectation of privacy by users, we do not include any genuine names, usernames or identifying information. Any quotes of posts we include have been bowdlerised so that the sense is preserved but the words used are altered to protect against identification of the author.

Gendered abuse is driving women out of communities they value

Evidence that this abuse has a clear and real effect on women's mental health is widespread. Many women spoke of wanting to stop their social media presence and activism, something that they otherwise enjoyed: "At the moment, it makes me want to quit everything I do online." Many women spoke of it affecting their mood, and felt threatened going online: "I can't even look at social media because I'm so scared that I'll see more sexism. It's really affecting my mental health." Some ended up feeling like they were responsible for the abuse, and seemed unhappy being women as a result: "I'm aware that my gender is weaker, and I'm miserable about that already."

THE IMPACTS OF GENDERED ONLINE ABUSE

THEME OF POST	POST COUNT
Women experiencing body shaming	9
Abuse online threatening women's careers (e.g. when their career requires them to be online)	5
Women being sexualised by others	8
Abusers telling them that online abuse is not a problem - either that it is non-existent or 'not that bad'	17
Misogynistic jokes being made online	2
Misogyny intersecting with other forms of discrimination e.g. racism, transphobia	8
Women experiencing exhaustion and a desire to self-censor to avoid abuse	18
Women being targeted with threats of violence	2
Women being blamed for the abuse they receive	11
Women bearing witness to the abuse that women in the public eye receive	6

Themes identified in 90 posts discussing women's experiences of gendered abuse online, with post count³

3 1 post = Other category

Many of the women we observed discussing their experiences were on the cusp of leaving social media and internet forums for good, or wanted to drastically limit their content and make their accounts private. These women spoke of being pushed off gaming sites and current affairs and political discussions in particular. One gamer tweeted, "It's so demoralising being a woman in what's seen as a man's world" alongside screenshots of messages calling her a "hideous fucking slut" and a "stupid bitch". Even when some tried to curate their feed to sports/animal content, these were often still intercepted with sexist abuse: "I was on a sports subreddit and all the comments said things like 'if your wife doesn't get your commitment to the game, DIVORCE HER'."

Online 'humour' is threatening women's safety

Women are subject to threats as 'jokes', and are then further abused if they call out this behaviour. We observed on Reddit in particular complaints about a culture of posting misogynistic memes and "jokes" that threatened women. One user said that she saw "memes about literally murdering women just because they were women" and received death threats for calling out this "humour" on Reddit. Another user noted a pattern where, in response to men complaining about their girlfriends, other users would pile in with insults and threats about the woman in question, such as "AWALT" (all women are like this) and "leak her nudes".

Gendered abuse intersects with many other forms of abuse such as racism, homophobia and transphobia

While we were searching for misogynistic abuse, it became clear that this frequently intersected with other forms of abuse, including racism and transphobia. Sometimes this was overt, with women called the N word online. Other times these were tools of abuse used to degrade women's appearances, such as, "she's nothing special with her natural [Afro-Caribbean] hair" or "she looks like a bloke wearing a wig".

Women's privacy is being violated and their personal information used to abuse them

One woman on Twitter spoke of photos she had uploaded being used on a pro-anorexia site where other users (mainly women) would shame her body. On Reddit, women users complained about men encouraging each other to upload revenge pornography of their partner if she wronged him. Others said that when male users disagreed with

what a women user said, they would "scour women's profiles for 'dirt' on them to insult and degrade them".

Women are attacked for speaking out about their experiences - creating a vicious cycle of abuse that women can only break by disengaging from online spaces

We saw commonly that women speaking out faced skepticism, half-hearted recognition of sexism combined with criticism of the individual at best, and at worst, further abuse. Across the platforms, many conversations about women's online abuse were often hijacked by users countering this with claims of widespread "misandry" (hatred of men, a term often used to attack feminists) and "double standards". The negative responses claimed that either online abuse is not real and that women have nothing to complain about, or blamed the woman for putting herself online or in the public eye in the first place.

Cycle of silence

Women users who speak out against gendered abuse online are often met with further abuse, disbelief and criticism: "You get more attacked for making general comments about men than for treating women as a subspecies." The alternative to speaking up is to become more reticent, as some fans noted an influencer doing, "I've noticed she's so careful about everything she says because she's scared of the hate." Abuse is inevitable for many women internet users who often see themselves in a lose-lose situation: they are criticised whether they speak out against the abuse, or if they remain silent.

Women are blamed for the abuse that they receive: abuse is portrayed as inevitable, and so it is the responsibility of the target to keep themselves safe

These victim-blaming comments were particularly noticeable when the women were high profile, such as politicians, actors or influencers. When women used Reddit and Youtube to discuss the criticism Emma Watson received for her gender equality campaign HeForShe, many users considered online abuse "fair game": "it's the internet, what do you expect". In Daily Mail comments on articles about former Love Islanders discussing the effects the abuse has had on them, many users criticised those posting the abuse, but also the women for not knowing better: "Don't want to be trolled? Don't become an "influencer"!" There was even a sense in which women were seen as attention seeking and desperate for talking about this issue and the pitfalls of the celebrity spotlight: "She's desperate to be

back in the *Love Island* limelight but she's past it. There are new girls now. Bye!"

While the victim-blaming comments acknowledged trolling and abuse as an issue at least in part, they somewhat paradoxically went hand in hand with comments denying that sexist abuse was a problem at all. Many of these were centred around the premise that "real" abuse is physical; that the "virtual" world isn't real: "Who seriously believes these rape threats?" They also implied that women, especially politicians and football pundits, were too sensitive to handle legitimate criticism, and were "conflating abuse on the internet with actually holding her to account".

Discussions about women's safety are seen as unfairly neglecting men's experiences

Many conversations by women about sexist abuse were taken over by male users who objected to the "double standards" they perceived existing between men and women: "Every time a man says anything it's "mansplaining", but no one ever calls women out for "femsplaining", and that sums it up that women are sexist." Sometimes male users expressed anger that the online abuse directed towards women got attention, when online abuse directed towards men, as well as physical violence, did not: "Sorry love, this "harassment" isn't violent. Violence is being kicked or punched or held at knife point, it's not just someone going off on one in your comments." They did not see abuse as gendered; they believed that everyone online gets abused: "The internet doesn't hate on women, it hates everybody".

Women users sometimes felt their online harassment was minimised by those saying it was not "in the real world" and they could just leave it behind. "Touch grass" is a phrase used online to tell people to go off their computer, but one user noted that it is disproportionately used to shut down "someone challenging sexism and racism online". Many women users' posts challenge the notion of being able to exit the abuse online, as they feel very real spillover effects of online abuse into their non-virtual lives: "I muted the conversation and went back to bed, when I woke up in the morning, I just wanted to cry. I feel so low."

POLICY RAMIFICATIONS

Current systems of redress put the burden on women to manage the abuse they receive online and are failing to meet even minimum standards of supporting them in doing that.

We observed discussions of women reporting their harassment to the online services they use, using existing reporting tools and providing evidence of the abuse they are receiving, only for no action to be taken. "I am sick and tired of this community. I submit a report about the non-stop gendered abuse I've gotten, including screenshots, and not a thing gets done, despite this evidence?" Users did not seem to see reporting the abuse to the police as an option. The exception was celebrities who received death threats and who shared that they worked with the police to find the abusers.

Social media platforms provide a source of solidarity and a space for women to share their experiences

Although we saw discussion of women having extremely negative experiences on social media platforms, we also saw them turning to other platforms to share those experiences more safely, such as women using Reddit or Twitter to discuss and share screenshots or anecdotes from their interactions on gaming platforms. Sometimes they shared advice for avoiding abuse, such as hiding their gender online: "I deleted everything that would suggest I was female online. I changed my icon, my pronouns. When I joined the game, I was treated completely differently and I could play without worrying about being stalked."

Many also use social media to call out the sexist abuse high profile celebrity women receive online, as well as to talk about the sexist abuse they witness or experience in offline life. We observed comments in response to women speaking out about being attacked online expressing overwhelming, even unanimous, solidarity with the woman, offering sympathy and support. These comments typically came from their fans' accounts, who appear almost entirely younger and female.

Licence to publish

Demos – Licence to Publish

The work (as defined below) is provided under the terms of this licence ('licence'). The work is protected by copyright and/or other applicable law. Any use of the work other than as authorized under this licence is prohibited. By exercising any rights to the work provided here, you accept and agree to be bound by the terms of this licence. Demos grants you the rights contained here in consideration of your acceptance of such terms and conditions.

1 Definitions

a 'Collective Work' means a work, such as a periodical issue, anthology or encyclopedia, in which the Work in its entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole. A work that constitutes a Collective Work will not be considered a Derivative Work (as defined below) for the purposes of this Licence.

b 'Derivative Work' means a work based upon the Work or upon the Work and other pre-existing works, such as a musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Work may be recast, transformed, or adapted, except that a work that constitutes a Collective Work or a translation from English into another language will not be considered a Derivative Work for the purpose of this Licence.

c 'Licensor' means the individual or entity that offers the Work under the terms of this Licence.

d 'Original Author' means the individual or entity who created the Work.

e 'Work' means the copyrightable work of authorship offered under the terms of this Licence.

f 'You' means an individual or entity exercising rights under this Licence who has not previously violated the terms of this Licence with respect to the Work, or who has received express permission from Demos to exercise rights under this Licence despite a previous violation.

2 Fair Use Rights

Nothing in this licence is intended to reduce, limit, or restrict any rights arising from fair use, first sale or other limitations on the exclusive rights of the copyright owner under copyright law or other applicable laws.

3 Licence Grant

Subject to the terms and conditions of this Licence, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) licence to exercise the rights in the Work as stated below:

a to reproduce the Work, to incorporate the Work into one or more Collective Works, and to reproduce the Work as incorporated in the Collective Works;

b to distribute copies or phono-records of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission the Work including as incorporated in Collective Works; The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. All rights not expressly granted by Licensor are hereby reserved.

4 Restrictions

The licence granted in Section 3 above is expressly made subject to and limited by the following restrictions:

a You may distribute, publicly display, publicly perform, or publicly digitally perform the Work only under the terms of this Licence, and You must include a copy of, or the Uniform Resource Identifier for, this Licence with every copy or phono-record of the Work You distribute, publicly display, publicly perform, or publicly digitally perform. You may not offer or impose any terms on the Work that alter or restrict the terms of this Licence or the recipients' exercise of the rights granted hereunder. You may not sublicense the Work. You must keep intact all notices that refer to this Licence and to the disclaimer of warranties. You may not distribute, publicly display, publicly perform, or publicly digitally perform the Work with any technological measures that control access or use of the Work in a manner inconsistent with the terms of this Licence Agreement. The above applies to the Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Work itself to be made subject to the terms of this Licence. If You create a Collective Work, upon notice from any Licensor You must, to the extent practicable, remove from the Collective Work any reference to such Licensor or the Original Author, as requested.

b You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended for or directed toward commercial advantage or private monetary compensation. The exchange of the Work for other copyrighted works by means of digital file sharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works.

c If you distribute, publicly display, publicly perform, or publicly digitally perform the Work or any Collective Works, you must keep intact all copyright notices for the Work and give the Original Author credit reasonable to the medium or means You are utilizing by conveying the name (or pseudonym if applicable) of the Original Author if supplied; the title of the Work if supplied. Such credit may be implemented in any reasonable manner; provided, however, that in the case of a Collective Work, at a minimum such credit will appear where any other comparable authorship credit appears and in a manner at least as prominent as such other comparable authorship credit.

5 Representations, Warranties and Disclaimer

a By offering the Work for public release under this Licence, Licensor represents and warrants that, to the best of Licensor's knowledge after reasonable inquiry:

i Licensor has secured all rights in the Work necessary to grant the licence rights hereunder

and to permit the lawful exercise of the rights granted hereunder without You having any obligation to pay any royalties, compulsory licence fees, residuals or any other payments;

ii The Work does not infringe the copyright, trademark, publicity rights, common law rights or any other right of any third party or constitute defamation, invasion of privacy or other tortious injury to any third party.

b Except as expressly stated in this licence or otherwise agreed in writing or required by applicable law, the work is licenced on an 'as is' basis, without warranties of any kind, either express or implied including, without limitation, any warranties regarding the contents or accuracy of the work.

6 Limitation on Liability

Except to the extent required by applicable law, and except for damages arising from liability to a third party resulting from breach of the warranties in section 5, in no event will licensor be liable to you on any legal theory for any special, incidental, consequential, punitive or exemplary damages arising out of this licence or the use of the work, even if licensor has been advised of the possibility of such damages.

7 Termination

a This Licence and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this Licence. Individuals or entities who have received Collective Works from You under this Licence, however, will not have their licences terminated provided such individuals or entities remain in full compliance with those licences. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this Licence.

b Subject to the above terms and conditions, the licence granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different licence terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this Licence (or any other licence that has been, or is required to be, granted under the terms of this Licence), and this Licence will continue in full force and effect unless terminated as stated above.

8 Miscellaneous

a Each time You distribute or publicly digitally perform the Work or a Collective Work, Demos offers to the recipient a licence to the Work on the same terms and conditions as the licence granted to You under this Licence.

b If any provision of this Licence is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Licence, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

c No term or provision of this Licence shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

d This Licence constitutes the entire agreement between the parties with respect to the Work licenced here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This Licence may not be modified without the mutual written agreement of Demos and You.

DEMOS

Demos is a champion of people, ideas and democracy. We bring people together. We bridge divides. We listen and we understand. We are practical about the problems we face, but endlessly optimistic and ambitious about our capacity, together, to overcome them.

At a crossroads in Britain's history, we need ideas for renewal, reconnection and the restoration of hope. Challenges from populism to climate change remain unsolved, and a technological revolution dawns, but the centre of politics has been intellectually paralysed. Demos will change that. We can counter the impossible promises of the political extremes, and challenge despair – by bringing to life an aspirational narrative about the future of Britain that is rooted in the hopes and ambitions of people from across our country.

Demos is an independent, educational charity, registered in England and Wales. (Charity Registration no. 1042046)

Find out more at www.demos.co.uk

DEMOS

PUBLISHED BY DEMOS OCTOBER 2021

© DEMOS. SOME RIGHTS RESERVED.

15 WHITEHALL, LONDON, SW1A 2DD

T: 020 3878 3955

HELLO@DEMOS.CO.UK

WWW.DEMOS.CO.UK