Institution: University of Sheffield		
Unit of Assessment: C-21 Sociology		
Title of case study: Making social media platforms safer for people with eating disorders		
Period when the underpinning research was undertaken: September 2017 to March 2018		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Ysabel Gerrard	Lecturer in Digital Media and Society	September 2017–current
Period when the claimed impact occurred: 2018–2020		
Is this case study continued from a case study submitted in 2014? N		

1. Summary of the impact (indicative maximum 100 words)

Through policy interventions at globally popular social media companies, Dr. Ysabel Gerrard's research has made platforms safer for people with eating disorders. Social media companies have been heavily criticised for lax regulation of posts about eating disorders, but Gerrard's research has directly led Instagram and Tumblr to improve their policies and better protect vulnerable users. This research has transformed how potentially harmful posts are advertised on social media and changed the rules on what people are allowed to do and say. Extensive international media coverage of the findings has raised public awareness of eating disorders, encouraging people to seek help and, ultimately, save lives.

2. Underpinning research (indicative maximum 500 words)

Rationale: Eating disorders are on the rise in the UK and internationally, particularly among young women. According to Beat, the UK's eating disorder charity, anorexia is the psychiatric condition with the highest mortality rate (**R1**). Before Gerrard's research-led interventions, social media platforms were plagued with content posted by users to promote the worsening of eating disorders like anorexia: so-called 'pro-ana' posts. Social media companies write their own policies and remove content that breaks the rules, via *content moderation.* But their policy teams previously lacked the expertise to understand what happens in eating disorder communities and how they should decide which posts to remove. This resulted in years of public criticism of inaction. Gerrard's research addressed this issue, filling gaps in policymakers' knowledge about how, exactly, people break social media companies' rules to enact harmful behaviours. Following extensive media coverage of Gerrard's research, Instagram and Tumblr approached her to lead transformative changes to their content policies.

Research: This impact case study draws on findings from three publications (**R1, R2, R3**). The underpinning research was conducted during Gerrard's first year at the University of Sheffield, using digital methods such as covert online observations.

Findings: Gerrard's research filled three main gaps in knowledge:

1) **Hashtag moderation:** Social media users often use hashtags (e.g. #proana) to expose their posts to people beyond their 'friend' networks. Gerrard revealed that Instagram and Tumblr failed to restrict access to all hashtags that promoted eating disorders, exposing

vulnerable users to distressing content (**R3**). *This finding highlighted the need for Instagram and Tumblr to restrict users' access to more potentially harmful hashtags.*

2) **Username moderation:** Social media users generally choose a username, and while some people use their real name, others apply pseudonyms. Gerrard's research discovered that Instagram and Tumblr did not restrict access to searches for usernames containing terms that promote eating disorders (**R3**). *This finding highlighted the need to develop new content moderation techniques that prevent users from choosing usernames that could be harmful.*

3) **Recommendation systems:** Social media platforms often show new content to users by collecting data on their browsing habits and using this to show them other things they might be interested in. However, Gerrard's research revealed that Instagram and Tumblr recommend troubling content that promotes eating disorders, including for 'miracle' diet products (**R1, R3, R4**).

Gerrard also published her findings in a *WIRED* magazine article (**R1**), in order to share the findings with technology industry workers from a broader range of companies.

3. References to the research (indicative maximum of six references)

R1. **Gerrard, Y.** and Gillespie, T. (2019, 21st February). When algorithms think you want to die. *WIRED.* Available at: https://www.wired.com/story/when-algorithms-think-you-want-to-die/.

R2. McCosker, A., & **Gerrard, Y.** (2020). Hashtagging depression on Instagram: Towards a more inclusive mental health research methodology. *New Media & Society*, [published online ahead of print]. https://doi.org/10.1177/1461444820921349

R3. **Gerrard, Y. (**2018). Beyond the hashtag: Circumventing content moderation on social media. *New Media & Society, 20*(12), 4492–4511. https://doi.org/10.1177/1461444818776611

R4. **Gerrard, Y.** and Thornham, H. (2020). Content moderation: social media's sexist assemblages. *New Media and Society*. 22(7): 1266-1286. https://doi.org/10.1177/1461444820912540.

4. Details of the impact (indicative maximum 750 words)

Gerrard's research has led to wide-scale policy changes at two of the world's most popular social media companies: Instagram and Tumblr (1 billion and ~400 million active users worldwide respectively, Statista, 2020). [Text removed for publication].

Her work also attracted the attention of the UK Government's Centre for Data Ethics and Innovation (**S2**) stating "*The research highlighted the need for Instagram and Tumblr to change how their recommendation systems work, to avoid suggesting dangerous content to vulnerable users*". Instagram and Tumblr have acted on all of the research recommendations.

Gerrard's findings prompted multiple forms of impact for three beneficiaries: policymakers at Instagram and Tumblr, social media users who are suffering or recovering from an eating disorder, and members of the public. The significance of the impact is as follows:

Instagram: As a direct result of the media coverage Gerrard's research (**R3**) received, she was invited to become a member of Instagram and parent company Facebook's Suicide and Self-

Injury (SSI) Advisory Board, to contribute expertise to content policies about eating disorders (**S1**).

[Text removed for publication]. Gerrard's research brought to the Board:

- First, the research and its subsequent publicity – which influenced a BBC investigation (**S5**) – [Text removed for publication] led to users searching a *"much greater"* range of hashtags. This information has decreased the number of pro-eating disorder posts on Instagram: [Text removed for publication]. Thanks to Gerrard's interventions, this means fewer people are posting dangerous content to Instagram and are instead using the platform's resources to seek help.

- Second, Gerrard's research informed a new Instagram policy banning advertisements for weight loss products making 'miraculous' claims about their power to help people lose weight (**S6**, **S4**). Gerrard was the only expert cited in Instagram's press release, which said: "*We've sought guidance from external experts, including Dr Ysabel Gerrard in the UK, to make sure any steps to restrict and remove this content will have a positive impact on our community of over 1 billion people"* (**S6**). The Girlguiding charity was one of many organisations to highlight the value of the policy, saying "*every step like this helps build a world girls want to live in"* (**S7**).

- Third, the research led Instagram to transform how it recommends content to users on its Explore page (which shows users new images they might be interested in seeing) (**R1, R2, R3**). [Text removed for publication]. As a result, Gerrard worked with Instagram to tighten the criteria a post needs to meet before it can be recommended to people. This will decrease the chance that users will be exposed to posts promoting eating disorders. [Text removed for publication].

Tumblr: Tumblr has changed its content moderation policies in two key ways as a direct result of Gerrard's research. First, after an invited visit to meet with Tumblr's CEO and key policy figures, the company changed how it recommends eating disorder-related content to its users: [Text removed for publication]. Without this research, Tumblr would not have taken urgent steps to reduce the risk that graphic content might appear on a users' feed.

Second, Gerrard authored Tumblr's resources for Mental Health Awareness Month (**S9**). Almost 400 million Tumblr users can access Gerrard's post to learn about eating disorders and be signposted to health services. [Text removed for publication].

5. Sources to corroborate the impact (indicative maximum of 10 references)

S1. Letter, Associate Professor (Faculty of Law and member of Facebook's Oversight Board), *Queensland University of Technology*. This letter explains the originality of the research findings and provides some metrics to demonstrate the scale of positive impact Gerrard had on Instagram in particular.

S2. Research cited in a UK Government Centre for Data Ethics and Innovation report, 'Online targeting: final report and recommendations' https://www.gov.uk/government/publications/cdei-review-of-online-targeting/online-targeting-final-report-and-recommendations (February 2020).

S3. Letter, Public Policy Associate, Facebook. This letter details the positive impact of Gerrard's research on Facebook and Instagram's content moderation policies, mainly how

it has helped to prevent harmful imagery being recommended to the platforms' users, and how the companies classify and control certain kinds of content.

S4. Letter, Public Policy Manager (UK and Northern Europe), Instagram. This letter demonstrates how Gerrard's research has helped to make Instagram a safer place for users, and explains how she will continue to work with the company to implement even more positive changes.

S5. Research cited in a *BBC Trending* radio programme. 'Do Instagram hashtags promote eating disorders?': https://www.bbc.co.uk/programmes/w3csws7n (December 2018). This interview led the *BBC* to conduct its own investigation into harmful hashtags on Instagram, and as a direct result Instagram expanded its list of banned hashtags.

S6. Media coverage portfolio. This demonstrates the significant international national impact of the research and impact on public opinion. The *Guardian* article (**p.1**) includes the official Instagram press release, made by Emma Collins (Instagram Public Policy Manager).

S7. Twitter post by the Girlguiding charity: https://twitter.com/Girlguiding/status/1174641536983871490. This message is one of many praising Instagram's new policy on cosmetic surgery and weight loss-related advertisements, for which Gerrard was the only named expert contributor.

S8. Letter, Director of Social Impact and Public Policy, *Tumblr*. This letter details the positive impact of Gerrard's research on Tumblr's content moderation policies, specifically commenting on how her interventions have helped to make the platform safer for vulnerable users.

S9. Gerrard's contribution to Tumblr's Mental Health Awareness Month Post it Forward campaign, 'Eating Disorders': https://postitforward.tumblr.com/post/184849352618/far-too-often-we-equate-the-perception-of-others (May 2019). This is a post that Gerrard authored for Mental Health Awareness Month. In it, she signposts Tumblr's userbase to eating disorder resources, provides information about the condition and shares users' stories of recovery.