

Supplementary written evidence submitted by Dr Stephanie Alice Baker

Policy Recommendations: Select Inquiry into Influencer Culture

Dr Stephanie Alice Baker

Thank you for the opportunity to provide further written evidence to the Select Inquiry into Influencer Culture. These recommendations pertain directly to the session on misinformation and disinformation, which I provided oral evidence at on 25 November 2021.

Terminology	Definition
Misinformation	False or misleading information believed to be true.
Disinformation	False or misleading information intended to deceive or cause harm.
Influencer	A content creator who builds an online following on social media for social, economic or political gain.

- 1) **Tech platforms should work together to limit the spread of misinformation and disinformation online.** We need to move away from thinking about individual platforms to consider how platforms operate within the broader information ecosystem. Many influencers producing disinformation attempt to enact what I refer to as the ‘Pied Piper Effect’ by using mainstream social media platforms to build an online audience before directing their followers to more conspiratorial content on personal websites, newsletters and encrypted messaging services, such as Telegram (see Baker, 2021). Disinformation is also shared on mainstream platforms, however, it is often concealed in the form of questions, memes and personal anecdotes to avoid detection by fact checkers (see Baker et al., 2020). Some influencers normalise conspiracy theories on mainstream platforms by sporadically publishing disinformation alongside more generic content. Disinformation is often made more alluring on these sites by using aesthetically pleasing images and videos to appeal

to mainstream audiences (Baker, 2022). For example, during the pandemic numerous wellness influencers used the theme of purity to spread disinformation and political extremism online. Ostensibly innocuous claims about the benefits of ‘clean eating’, holistic health and natural remedies become the gateway to anti-vaccine content and, in some instances, xenophobic claims (see Image 4 of the ‘Purity Paradigm’ – Baker, 2021; see also Baker and Walsh, 2020; Walsh and Baker, 2020 on the role of purity in configuring wellness communities on Instagram). Mothers are strategically targeted on social media by anti-vaccine influencers, who exploit popular beliefs about ‘maternal intuition’ to encourage mothers to refuse vaccinating their children (Baker and Walsh, 2022). In our research during the pandemic, we found examples of high-profile, anti-vaccine influencers misappropriating the #SavetheChildren, #SaveOurChildren and #SavetheBabies hashtags to appeal to mothers by merging the anti-vaccine movement and the Save the Children movement as common efforts to protect innocent children from harm (Baker and Walsh, 2022). Following the murder of George Floyd in May 2020, some anti-vaccine influencers strategically targeted Black, Asian and minority ethnic communities by co-opting Black Lives Matter hashtags and using the incident to sow distrust of mainstream science and medicine (see Baker and Walsh, 2022).

Influencers suspended from social media platforms commonly use claims of censorship to appear unjustly persecuted by Big Tech and mainstream authorities (see Baker, 2021). By depicting themselves as persecuted heroes, these influencers are able to mobilise loyal online followings of like-minded individuals willing to defend truth, freedom and justice. The persecuted hero motif often becomes the basis for influencers publishing books and films alleging to document conspiracy theories involving these self-described “martyrs” and “whistle blowers”. Several high-profile influencers have profited from using the persecuted hero motif to disseminate medical misinformation on e-commerce sites and streaming services, such as Amazon and Gaia, during the pandemic. For example, Robert F. Kennedy Jr. and

Judy Mikovits's books became bestsellers on Amazon and Gaia still sells David Icke's conspiracy films, despite the fact that his social media accounts were suspended in 2020 for spreading coronavirus medical misinformation (see Baker, 2020a, 2020b). In addition to enabling influencers to profit from medical misinformation, *by distributing their products (e.g. books, films), these companies give alt. health influencers a degree of legitimacy.* This was an issue a colleague and I identified in our research on medical misinformation prior to the COVID-19 pandemic where we demonstrated how cancer frauds were both profiting from, and legitimised by, major publishing and technology platforms, such as Penguin and Apple (Baker and Rojek, 2019, 2020). Content moderation is difficult to implement at speed and scale. *The aim of regulation should not be to remove every piece of content that features misinformation; but to remove the incentives to produce and share disinformation online.* Specific actions tech platforms could take include:

- a) Creating specific guidelines for public figures and influencers to inform them of their responsibilities when sharing content on their platforms. Establishing guidelines would help influencers understand what actions are permitted on these platforms and the consequences for violating their policies. Platforms should have strict policies to prevent influencers from advertising fraudulent products on their sites, especially when they pose imminent physical harm to consumers (e.g. products promoting ingesting bleach, colloidal silver and Ivermectin intended for animal use as treatments for COVID-19). Penalties for influencers spreading medical misinformation could include downranking and demonetisation to remove the financial incentives to spread false and misleading information online. Suspensions could be enforced for coordinated inauthentic behaviour and repeat offenders. There should also be clear guidelines for celebrities, influencers and politicians about their role in amplifying disinformation online, especially those with verified accounts. These guidelines should indicate the procedures platforms will take to hold influencers to account for spreading misinformation and disinformation on their sites. For example, repeat offenders losing their verification status. Influencers should also be informed about

the ways they might unknowingly be used to amplify misinformation. These guidelines could be published by the platforms and reinforced by the agents who manage the top and mid-tier influencers.

- b) Implementing common community guidelines to address the covert strategies some influencers use to spread misinformation and disinformation. Platforms should also work together to improve content moderation practices. This includes increasing human fact checkers in different regions with fluency in different languages and dialects. Companies also require more robust approaches to moderate visual content. At present, most disinformation evades detection from automated content moderation as it takes the form of memes, images and videos. *Content moderation will not be improved by simply hiring more human fact checkers, tech platforms need to work together to improve the quality of fact checking practices.* Community guidelines need to address the covert strategies some influencers use to spread disinformation including coded language, asking questions, sharing memes and personal anecdotes (Baker et al., 2020).
- c) Reducing virality and visibility at scale by introducing friction across the major platforms. Social media enables misinformation to be amplified at speed and scale, in some instances going viral. Influencers are often instrumental in making content go viral. For example, the conspiracy theory film, *Plandemic*, was amplified by lifestyle and wellness influencers, many of whom had verified accounts (see Baker, 2020a). Introducing friction when content appears problematic would enable fact checkers to examine a post's veracity before it goes viral. Friction could take the form of labels providing clear, reliable information (e.g. succinct articles representing data in visual form), limiting shares and disabling trending hashtags when they are being used to disseminate false and misleading narratives. Hashtags are especially important to moderate as they enable like-minded groups to affirm each other's beliefs and contribute to a common narrative (Baker and Walsh, 2018). This occurred during the pandemic with the viral spread of the #plandemic hashtag, which is still active on Instagram and Twitter despite these platforms announcing they would

disable the hashtag #plandemicthemovie to combat the spread of misinformation about COVID-19. Trending hashtags give the appearance of truth by virtue of capturing public attention. Tech platforms could limit the spread of disinformation by sharing potentially problematic content they identify with each other before a narrative gains traction.

d) Enforcing permanent suspensions for influencers involved in extreme violations across all major platforms and publishing sites. We require a more coordinated approach by tech platforms to remove dangerous users from their sites. Just as the major tech platforms – Facebook, Google, LinkedIn, Microsoft, Reddit, Twitter and YouTube – took the unprecedented move in March 2020 to work together to combat the spread of misinformation about COVID-19 and elevate authoritative content on their platforms (see Baker et al., 2020), tech platforms could work together to prevent influencers from disseminating harmful disinformation online. At present, many of the Disinformation Dozen are only suspended from one of the major social media sites. This includes Facebook and Instagram, despite both being owned by Meta. If platforms are serious about providing safe spaces for users to communicate online, we require a cross-platform approach to content moderation. A cross-platform approach to content moderation would prevent influencers migrating to other platforms when their accounts are suspended from a platform, thereby, limiting the spread of disinformation online. Influencers should have the right to appeal suspension. This process should be simple, include a human response and the right for redemption following a period of suspension.

2) **An independent oversight board should regularly audit platforms to ensure the safety of their algorithms and that their community guidelines are being enforced.**

a) These results should be published by an independent regulator rather than self-published by tech platforms, who are essentially ‘marking their own homework’. At present, the transparency reports published by tech platforms only provide country

statistics on removal requests without any insight into what posts have been removed, by whom and why (Baker et al., 2020). In addition to conducting independent research into the recommendation algorithm to assess the effects of post promotion and targeted advertising (e.g. which users they are connecting, who profits from disinformation and what content is removed or downranked), the oversight board should publish statistics about the extent to which tech platforms enforce their community guidelines. Publishing reports about the efficacy of content moderation practices in statistical form would hold tech platforms to account without violating the privacy of individual users.

- b) Regulators should inform platforms when their community guidelines have not been adequately enforced and platforms should have a specific timeframe to respond to these issues. There should be financial penalties for platforms knowingly failing to enforce their community guidelines. Publishing these results will hold platforms to account (see 2a).
 - c) Results should be archived to see how content moderation has been enforced. e.g. what content has been removed or downranked, by whom, and why.
 - d) A select group of independent researchers should be given access to harmful content (including those posts removed by fact checkers), so we can learn more about how influencers produce disinformation, which accounts are targeted and the impact of content moderation strategies, such as reporting, notifications and labels. Given the rate of technological change, content moderation strategies will need to evolve. Independent research into the efficacy of content moderation practices will provide timely, iterative feedback into effective ways to tackle the spread of misinformation and disinformation online.
- 3) **There should be strict penalties for influencers who repeatedly spread misinformation, including those who use more covert strategies such as asking questions, sharing memes and personal anecdotes to sow uncertainty and doubt.** Fact-checking to determine accurate information is not sufficient to reduce

the spread of misinformation. *Influencers trade off sharing opinions rather than facts*, in the context of health using their personal journey of self-transformation to stand in for professional expertise (Baker, 2021). Wellness influencers sharing medical misinformation often use disclaimers to avoid regulation and responsibility for the opinions they share online (Baker and Rojek, 2019, 2020).

- a) Repeated instances of influencers sharing false and misleading content online in the form of personal opinions, memes, questions and anecdotes should be subject to the same criteria as fact-based claims. This would reduce the capacity for influencers to use these covert strategies to foster fear, uncertainty and doubt.
- b) Public figures should also be regulated as they can amplify the spread of misinformation and disinformation. Public figures contributing to the spread of misinformation was an issue prior to the pandemic as demonstrated by the strategic use of political influencers to amplify the viral video campaign, KONY 2012, on social media (Baker, 2014). Public figures have also contributed to the spread of conspiracy theories related to COVID-19. During the pandemic, there have been numerous clinical trials to study repurposing existing drugs as potential treatments for COVID-19. While many of these treatments appear effective in high doses in-vitro, these results do not necessarily translate in human studies. Over the course of the pandemic, several high-profile public figures have exaggerated the efficacy of potential treatments for COVID-19, including hydroxychloroquine and Ivermectin, by presenting these drugs as miracle cures for the virus (see Baker and Maddox, 2022). Narratives of this kind imply that government healthcare agencies are conspiring to prevent the public from accessing these drugs to enable pharmaceutical companies to profit from manufacturing vaccines for COVID-19 at scale. Such claims amplify medical misinformation and undermine efforts to vaccinate the population (Baker et al., 2020). These claims also correlate with an increase in prescriptions for these drugs and can result in real-world harms with some wellness influencers misleading the public by selling Ivermectin during the pandemic for 'animal use' (Baker, 2021; Baker and Maddox, 2022). In these

instances, Ivermectin was spelled Iv.er.mectin to avoid detection from fact checkers.

- c) Regulation should not be limited to so-called 'super spreaders', many of whom are mid-tier influencers with 50-500,000 followers; any content creator who profits financially from advertising products on social media should be regulated. This is because *the influence of an influencer is not measured simply by their follower count, but by their capacity to earn the trust and admiration of a loyal community of followers.* Some of the most influential online users are micro or nano influencers with a relatively small number of followers (Baker, 2021). These influencers, nevertheless, tend to be trusted and admired by their followers and therefore are in a powerful position to influence them. This is a particularly serious issue when it comes to influencers targeting marginalised groups already distrusting of authority (Baker and Walsh, 2022).

- 3) **Tech platforms should consider creating a separate symbol to verify medical doctors and public health professionals.** Facebook suggests that verification signifies the authenticity of the user rather than endorsement. However, *verification can easily be misinterpreted by users as a sign of credibility.* Verification is especially susceptible to contribute to the spread of misinformation as some of the leading anti-vaccine advocates have verified accounts. David Icke's Twitter account was also verified prior to his suspension towards the end of 2020. Using a separate symbol to verify medical professionals would help users to identify trusted medical sources with smaller online followings. Those medical professionals with verified accounts could be subject to specific guidelines to ensure they are not using their influence to spread misinformation and disinformation.

- 4) **Tech platforms should consider elevating the voices of non-political medical authorities.** Misinformation flourishes in contexts of uncertainty and distrust (Baker, 2020a). When people distrust the government and medical establishment, they seek alternatives for scientific and medical advice. Alt. health influencers exploit people's

distrust of government healthcare agencies to promote their own products and services (see Baker, 2021). This is why the decision by the major tech platforms to elevate authoritative content from government healthcare agencies online is unlikely to be successful as a broad public health strategy (Baker et al., 2020). *Misinformation is not merely an information problem; it is a relationship problem.* Internet users are not simply in filter bubbles, they are in echo chambers with many online users inherently distrustful of the mainstream media, political and scientific institutions (Baker and Rojek, 2019; Baker, 2021). Elevating authoritative content is not sufficient to combat the current infodemic. Tech platforms should also consider elevating the voices of non-political medical professionals, who could potentially reach those disillusioned with the mainstream media, political and scientific institutions.

References

Baker, S. A. (2014). Mediation as Moral Education: Kony 2012—Can Social Tragedies Teach? In: *Social Tragedy: the Power of Myth, Ritual, and Emotion in the New Media Ecology* (pp. 149-175). Palgrave Macmillan: New York.

Baker, S. A. (2020a). Tackling Misinformation and Disinformation in the Context of COVID-19. *Cabinet Office C19 Seminar Series*, 8 July.

Baker, S. A. (2020b). Influencing the 'infodemic': how wellness became weaponised during the pandemic. *Lockdown: Mental Illness, Wellness, and COVID-19*.

Baker, S. A. (2021). Alt. Health Influencers: how wellness culture and web culture have been weaponised to promote COVID-19 conspiracy theories and far-right extremism. *European Journal of Cultural Studies*.

Baker, S.A. (forthcoming, 2022). *Wellness Culture: How the Wellness Movement has been used to Empower, Profit and Misinform*. SocietyNow Series, Emerald Publishing.

Baker, S. A. & Maddox, A. (2022). From COVID-19 Treatment to Miracle Cure: the role of influencers and public figures in amplifying the hydroxychloroquine and Ivermectin conspiracy theories during the pandemic. Special Issue – Conspiracy. *M/C Journal*.

Baker, S. A. & Rojek, C. (2019). The Belle Gibson scandal: The rise of lifestyle gurus as micro-celebrities in low-trust societies. *Journal of Sociology*, 56(3), 388-404.

Baker, S. A. & Rojek, C. (2020). *Lifestyle Gurus: Constructing Authority and Influence Online*. Cambridge: Polity.

Baker, S. A., Wade, M. & Walsh, M. J. (2020). The Challenges of Responding to Misinformation During a Pandemic: content moderation and the limitations of the concept of harm. *Media International Australia*, 177(1), 103-107.

Baker, S. A. & Walsh, M. J. (2018). 'Good Morning Fitfam': Top posts, hashtags and gender display on Instagram. *New Media & Society*, 20(12), 4553-4570.

Baker, S. A. & Walsh, M. J. (2020). You Are What You Instagram: clean eating and the symbolic representation of food. In: *Digital Food Cultures*, 53-68.

Baker, S.A. & Walsh, M.J. (2022). 'A Mother's Intuition: It's real and we have to believe in it': how the maternal is used to promote vaccine refusal on Instagram. *Information, Communication & Society*.

Walsh, M. J. & Baker, S. A. (2020). Clean eating and Instagram: purity, defilement, and

the idealization of food. *Food, Culture & Society*, 23(5), 570-588.

* Please note that the article 'Alt. Health Influencers' (2021), was formerly referred to as 'Influencing the infodemic: the intersection between wellness, conspiratoriality and far-right extremism (2021b)' in the written evidence I submitted to Parliament in May 2021 based on a conference paper I delivered in 2020 with the same title. It was accepted for publication on 9 November 2021 by the *European Journal of Cultural Studies*.