Ethical Dilemmas for @Celebrities: Promoting #Intimacy, Facing #Inauthenticity, and Defusing #Invectiveness

Marc Cheong

Abstract

The rise of social-media-mediated celebrity culture raises several philosophical concerns. Therefore, it is not uncommon to see, for example, Hollywood actors being placed in the same bracket as YouTube artists and Instagram influencers. The increased perceived ‘connectivity’ afforded by social media allows online celebrities to reach more fans and increases the perceived engagement or intimacy in the fan-celebrity relationship. In this paper I argue that this online relationship, which is beneficial to celebrities (for brand development) and social media companies (in profiting from the high engagement), has the potential to cause harm to both the celebrities as well as their fanbases (and sometimes innocent third parties). Firstly, I argue that celebrities’ attempts to be more intimate or authentic with their fanbase can turn out to be – more often than not – acts of Sartrean bad faith. Celebrities who tend to project a certain self-styled ‘online authenticity’, given the context of social media’s attention economy, can run the risk of treating a fan as a means-to-an-end, rather than an autonomous for-itself. Secondly, I argue that this phenomenon can catalyse real-world harms, by drawing upon network theory and social psychology. I argue that the connectedness of social media, ‘blending into the crowd’, and social media technologies’ ethos of optimising-for-engagement, leads to real harms to human well-being when fans do not engage critically and mindfully with online celebrities. Examples include Trump-esque celebrity politicians trying to drum up populist influence; celebrities feuding online encouraging fans’ conduct of harmful trolling and harassment; and vilification of those who do not subscribe to the same perceived worldview of a celebrity’s fandom. Following my arguments within, I shall outline an online celebrity’s ethical duties in their social media engagement, with emphasis on the duties that celebrity status comes with, particularly in preventing harms that transcend the online into the offline.

Keywords:

celebrity, social media, authenticity, crowd behaviour, influence, existentialism.

1 My thanks to the presenters at the Philosophy of Fame & Celebrity Workshop especially Alfred Archer, Catherine Robb, Georgie Mills, Matthew Dennis, Pilar Lopez-Cantero, Roos Slegers, and Wojciech Kaftanski for their ideas.
1.0 Introduction

The Internet has become a fertile medium for celebrities to thrive, expand their fame (and fan base), as well as monetise their engagement, amongst other things. The Oxford Dictionary defines a celebrity as both (I) “a famous person, especially in entertainment or sport” as well as (II) “the state of being well known” (“Celebrity” 2010). At the end of the 20th century, online coverage of celebrities was limited to fan pages\(^2\) of ‘offline’ or traditional (Schouten, Janssen, and Verspaget 2020) celebrities\(^3\), who were \textit{already} famous in their own right.

The rise of online social media has revised the interpretation of both of the above definitions. For starters, definition (I) is constantly redefined, as \textit{celebrityhood} in the age of social media is not only limited to traditional real-world artists, actors, or authors. ‘Entertainment’ now also encompasses Internet-powered media including but not limited to, online video platforms (TikTok for quick video shorts, Twitch for online game streaming), image sharing (Instagram), or other combinations thereof. Even non-human characters such as ‘Grumpy Cat’, made famous by an internet meme circa 2012\(^4\), are defined as \textit{Internet celebrities}.

The ‘well known’ part of the second definition (II) now encompasses celebrities who are made famous \textit{because of} the Internet and online social media; whereas the pre-21st century conception of celebrities are those who are already famous \textit{in spite of} the existence of the Internet\(^5\). The former case includes Instagram ‘influencers’, YouTube vloggers\(^6\), and Twitter users. These so-called influencers, or ‘Internet-facilitated’ celebrities:

\(\ldots\) gained fame by successfully branding themselves as experts on social media platforms (Khamis, Ang, and Welling 2017) \[b\]y enthusiastically sharing self-generated content on topics like beauty, fitness, food, and fashion \textit{to} gain \[\ldots\] a large follower base” (Schouten, Janssen, and Verspaget 2020).

For the purposes of this paper, I note upfront that \textit{celebrity} is a ‘superset’ or an umbrella term covering both offline celebrities (such as movie stars who maintain an online presence and engage with fans) and Internet-facilitated celebrities (such as influencers gaining a fan base both off- and online only \textit{after}, and \textit{because of}; their success on Instagram). Hence, the paper’s exploration as a whole encompasses both traditional/offline celebrities who have online presences and Internet-facilitated influencers.

In this paper, I argue that the online relationship between celebrities and their fans has the potential to cause harm, not only directly to the celebrities and the fans but also to potential innocent third parties (or the existential Other, as I shall expand on later). My argument looks at two complementary aspects: existentialist ‘bad faith’, as well as real-world harm to human well-being and flourishing. This paper is divided into three major sections: Section 2 covers the preliminaries underpinning Internet celebrity, including the social

\footnotesize{\(2\) See, e.g., the GeoCities pages characteristic of the 1990s, containing rudimentary (by today’s standards) fan-pages for Brad Pitt, Ellen DeGeneres, etc. https://geocities.restorativland.org/Hollywood/.

\(3\) For the purposes of this paper, we will use the terms ‘traditional celebrity’ and ‘offline celebrity’ interchangeably.

\(4\) See https://en.wikipedia.org/wiki/Grumpy_Cat

\(5\) For the purposes of this paper, we will use the terms ‘influencer’ and ‘Internet-facilitated celebrity’ interchangeably.

\(6\) A portmanteau of video and bloggers.}
technologies which enable (hitherto) non-celebrities to attain (Internet-mediated) celebrity status and traditional celebrities to move their fan base online. I present my first argument in Section 3, where I untangle celebrities’ online personae through the lens of existentialist philosophy, by refuting any normative claims of *social media authenticity* being an ideal for celebrities online: per Sartre et al, the assumption of the *pour-soi-en-soi*. My argument contains examples on how fans can easily be treated as a means-to-an-end, rather than an autonomous for-itself (*pour soi*), which is a manifestation of Sartrean ‘bad faith’ and *inauthenticity* (Sartre 1969). Section 4 contains my second argument, that social media technologies’ ethos of optimising-for-engagement and the sheer design of social networks can catalyse real-world harms, through the lens of network theory and social psychology, with an analogy to Soren Kierkegaard’s denouncement of the mass media (N. Jansen 1990): several real-world examples will be presented to illustrate such harms. In Section 5, I will recap the two major arguments above in order to come up with some desiderata for *ethical* social media engagement from celebrities, drawing upon existentialist ideals as well as from an applied perspective. As existentialist philosophy does not prescribe a unified “possibility of existentialist ethics” (Daigle 2006; Warnock 1967) based on the variety of ideas by its proponents, plus criticisms of existentialism’s inherent subjectivity, I instead look at its unifying theme: of the responsibility of an celebrity to maximise freedom (*a la* de Beauvoir). To this end, the desiderata for ethical behaviour can be ‘reverse-engineered’ by studying the detrimental effects of social media; and tracing it back to the celebrity behind the screen, with a goal of maximising fans’ and fellow social media users’ wellbeing, off- and on-line.

## 2.0 Preliminaries - from Social Networks to Social Media

Before delving into the main arguments, this section will briefly discuss the technical, architectural, and social aspects of online social media, which as we’ve seen earlier, both affords celebrityhood to influencers, and promotes ongoing engagement with offline celebrities.

Online social media sites during the turn of the mid 2000s capitalised heavily on online social networking (OSN)-based functionality. OSNs, modelled upon real-world human interpersonal connections, contains two key parts (Cheong 2013; Krishnamurthy 2009): it allows users to both form *connections* with other users and also generate *content/media* for consumption by other people (i.e. the social media aspect). Early OSNs were *symmetric*, where “both members participating in a dyadic social connection renders their connection symmetric... in terms of influence, esteem, etc.” (Cheong 2019). Examples of which include, popularly, the ‘friendship’ component of Facebook (Gannes 2010), where both parties to a Facebook friendship have a bi-directional ‘link’ with one another on the website.

Acknowledging that the symmetric mode of OSNs were not conducive to ‘professional’ users such as celebrities who want to reach out to many people, akin to dynamics of a fan club (Dunbar 2016; Agence France-Press 2016), social media sites pivoted to an *asymmetric* mode. Facebook’s Pages feature is an example of a pivot (both friends and Pages exist on Facebook at time of writing); and other platforms are built exclusively on asymmetric...
connections, such as Instagram, Twitter, YouTube, and TikTok. This asymmetric mode means that “a user A can unilaterally engage user [C], but user [C] need not necessarily reciprocate…” (Cheong 2019): in other words, A can ‘follow’, say, an Instagram personality C to keep up with new photos produced by A, but with no reciprocal commitment in the opposite direction. Asymmetric social media also allows a certain public consumption of C’s content without having to subscribe or otherwise engage as registered users (A’s) of the platform. Further, an “asymmetrical relationship can also take the form of unidirectional messaging or rebroadcasting of another’s content” (Cheong 2019) such as the Retweeting (RT-ing) of content by celebrities.

Celebrities on asymmetric OSNs thus benefit from two kinds of emergent asymmetric effects. The Asymmetry of Attachment, means that users in a network who already have many links to others (say YouTube subscribers) tend to gain more social network connections than those without many pre-existing links, per network theory (Barabasi and Albert 1999) – the ‘rich get richer’ effect. A given piece of media, by a user with a high follower count, tends to have a high probability for an arbitrary user will access it via their existing social network, or via ‘algorithmic recommender systems’ (such as the YouTube Autoplay feature). Bearing in mind that social media sites emphasise, and optimise for, engagement, highly connected (or followed) users are likely to be recommended. Herein, I define engagement to encompass any form of user-centred participation that can be quantitatively measured - from liking a YouTube video, to sharing a TikTok clip, to retweeting a tweet, to merely watching a video (thereby increasing its view count).

The second effect, the Asymmetry of Influence, reflects “how and why are particular actors more influential than others in terms of attracting attention” (Cheong 2019), by eliciting responses/comments, ‘likes’, shares/Retweets, and overall engagement. Celebrity C is not equivalent to a real-world friend of A’s, but clearly has something to offer to A – be it entertainment value, sense of being heard, esteem, information, etc. – for such a unilateral engagement to take place. This ‘intimacy at a distance’ allows online celebrities to reach more fans directly – via direct @user messages on Twitter and ‘likes’ on Facebook, for example – without the need for an intermediary. As a result, it increases the perceived engagement or intimacy in the fan-celebrity relationship: from a “realness that renders their narratives, their branding, both accessible and intimate” for influencers (Khamis, Ang, and Welling 2017), to the sustainment of direct “mediation, visibility, and attention” by traditional celebrities embracing the Internet (Marwick 2013), to an ‘ambient intimacy’ of presence on social media (O’Reilly and Milstein 2009; Cheong 2013).

3.0 Authentic Inauthenticity

To start my first argument, let’s turn to a definition of what normative values are favoured by celebrities online, in particular online celebrities or influencers. In an online setting, celebrities often emphasise authenticity\(^8\) in

---

\(^7\) For consistency and convenience, the symbol A is used to denote an everyday, non-celebrity user of any given asymmetrical social media platform, while C is used to denote a celebrity on said platform.

\(^8\) Some celebrities do not outwardly emphasise authenticity in their social media engagement: consider one which gets their PR agency to write their tweets and manage their Facebook page on their behalf, for instance.
their social media engagements (Duffy and Hund 2015; Turner 2006; Street 2004). The “question of authenticity” is raised by Rojek (2001) when celebrity presence is staged in the media: this is no different, I argue, from when a celebrity has to regulate the performative ‘me’ (Goffman 1959) they present to their fans (or the potential Other). A key difference between the online celebrity and the ‘traditional’ offline celebrity is that in the former, social media is ubiquitous and omnipresent: there is a never-ending need to engage with one’s followers, even after all stage-managed public appearances have finished for the day.

Celebrities have an ‘I’ - in other words, a veridical self - which is often different from the ‘me’ celebrities want to exhibit to their fan ‘Others’. Cases of traditional celebrities exhibiting dissonance between the ‘me’ and the veridical ‘I’ have been documented in Rojek (2001).

3.1 The Self, the in-itself, the for-itself, and their eternal tension on social media.

To contextualise this argument, I shall briefly turn to prior work on social media and existentialist authenticity (Cheong 2019, 2018; Dolezal 2012; Lopato 2016; Jose 2019; Dowden 2017). In the Sartrean\(^9\) existentialist context, curating one’s portrayal on online social media is an act invoking the tension of the ‘in itself’ (French: \textit{en soi}) -- \textit{facticities}, worldly objects, “kind[s] of being we might ascribe to, say, a pair of glasses” (Dowden 2017); versus the ‘for itself’ (French: \textit{pour soi}) – transcendence, consciousness, or summarised succinctly by Dowden (2017) as “autonomous, embodied subjects like you or I might be”. The tension exists because one’s real-world self is \textit{never} the self they choose to portray online to the Other: parts of it may be missing by omission, curation, edition, or revision (Lopato 2016; Cheong 2018):

“...the person who thinks [by travelling the world, posting their experiences on ...social media sites... and claiming that they are living authentically] “just as a matter of fact” is excluding from view the ability to transform [their existence] through [actually, say, striving to continually improve oneself...] \textit{...a denial of transcendence or freedom.}”

paraphrasing Varga and Guignon (2014), \textit{emphasises mine.}

Our curation of ourselves online in evoking authenticity is affected by “a mere awareness that anyone from across the globe might have a chance to look at [our]... witty tweet or insightful Instagram post” (Cheong 2019). In other words, our self-curated online portrayal – as an amalgamation of our posts, profiles, and activity on social media – are not just subject to the gaze by an abstract ‘Other’ (especially per Sartrean existentialist philosophy), but in other words, there is a realistic possibility that our entire online portrayal may be seen by many concrete ‘others’. This can come about by, e.g., algorithmic recommender systems elevating our content in social media sites’ search results or having a piece of content that we have buried deep in our online timeline

---

\(^9\) I shall borrow the use of the term Other (capital O) from phenomenology, particularly, inspired by the existentialist work by Sartre and de Beauvoir. It refers to the “point of view which is not my own” (Cheong 2019; Lopato 2016; Cox 2008; Dolezal 2012) -- in the context of this paper, the fans’ (or potential fans’) point of view.

\(^{10}\) I use a Sartrean analysis here as it is easily accessible; the distinction between the pour-soi and en-soi is also seen or alluded to by other existentialist philosophers, such as Heidegger and de Beauvoir.
becoming viral\textsuperscript{11}. As such we are tempted to portray our best selves (Goffman 1959) on social media: involving
e.g. smiling faces, a sense of accomplishment, and curation of facticities (Byrne 2019) to match the expectations
of the ‘Other’.

3.2 How do Celebrities Fare?

The prior discussion becomes apparent when celebrities are expected to present a narrative authenticity (Duffy
and Hund 2015) when they are engaging with the Other online\textsuperscript{12}. To explore this, I turn to a Sartrean
“existential reading of Kim Kardashian-West’s International Women’s Day selfie\textsuperscript{13} of 2016” (Dowden 2017) to
explain this phenomenon:

“The image depicts Kardashian-West in her bathroom, in a private and domestic sphere, taking an
image of her reflection in a mirror. … Kardashian-West takes control over her own image by taking a
selfie, and in this way the image is an expression of being-for-itself, of authentic being and of autonomy.
However, there is a paradox in that at the same time, the image exhibits an awareness of the male gaze,
and of the presence of Kardashian-West’s own body in the presence of the other, or her audience, which we
might consider a form of being-for-others”. (Dowden 2017, 127).

Celebrities tend to project a certain vulnerability (Dowden, 2016) or online authenticity, in their engagement
online. It is no wonder that celebrities resort to, e.g. “carefully curated personal sharing” behaviour (Duffy and
Hund 2015), for the purposes of, say, increasing fan engagement, or in the case of political celebrities, increasing
support for their policies (Turner 2006; Wood, Corbett, and Flinders 2016). That can hardly be characterised as
authentic, as they can be self-objectifying: in Sartre’s conception (Sartre 1969), in a for-itself-in-itself
(pour-soi-en-soi) mode of being that can hardly be characterised as authentic.

A salient example during the COVID-19 pandemic illustrates this in action. In March 2020, celebrities who
tried to record a collaborative cover of John Lennon’s ‘Imagine’ were criticised as a “naïveté ... blinding them to
the grossest sin ... the smug self-satisfaction, the hubris of the alleged good deed... [with]... an empty and
profundly awkward gesture from a passel of celebrities ... [bordering]... on delusion” (Caramanica 2020). The
portrayal of celebrities being purportedly empathetic with the rest of the population during the COVID-19
pandemic was panned as they failed to recognise their privileged positions. Celebrities who have different

\textsuperscript{11} A simple example of this is the Rickroll Internet meme, where an almost 20-year-old music video by Rick Astley (‘Never
Gonna Give You Up’) was made popular as an Internet meme involving a ‘bait and switch’. See

\textsuperscript{12} Online social media-based communications promote an (ersatz or otherwise) authenticity amongst its users. See, for
example, from (i) an overarching perspective on how Twitter’s early ethos was to promote genuine engagement that is
characteristic of the veridical ‘I’ (Cheong 2013); and (ii) how entrepreneurs using social media have a perceived need to
curate their online persona for the Other (Byrne 2017).

\textsuperscript{13} An excellent paper by Philips illustrates how selfies “reveal the hierarchical and prejudiced nature of visual (and wider)
culture as it relates to representations of women’s bodies” (Phillips 2020), and will be beneficial to readers to contextualise
Kim Kardashian-West’s case study.
Another example, writeup TikTok: a Facebook platform but limited to media modern not is Take across social little button the features users would ‘a dopamine’ like that as creation such give hit to attention was as consume It your possible? from outset...

exploited creating ‘a founders that vulnerability were something knew feeds designed news – (2013). that such media platforms sad social – Social and et optimise known to engagement Whiting al. Williams 2020; media is (Alfano Marin 4.1 Asymmetry and Rapid Connectivity?

Asymmetry in Section 2.0. How media does invectiveness? will modern-day idea with, site the revisit social of trigger begin I To I

4.0 Asymmetry as a catalyst for Invectiveness

How does a modern-day social media site trigger invectiveness? To begin with, I will revisit the idea of asymmetry in Section 2.0.

4.1 Asymmetry and Rapid Connectivity?

Social media is known to optimise for engagement (Alfano et al. 2020; Marin 2020; Whiting and Williams 2013). The sad news is that social media platforms – especially user-facing algorithms, such as the ones used to personalise news feeds – are designed to make us ‘repeat customers’ at best, and even addicts, at worst. As Solon (2017)’s article illustrates, by paraphrasing Sean Parker, a ‘founding president’ at Facebook:

“Facebook’s founders knew they were creating something addictive that exploited ‘a vulnerability in human psychology’ from the outset... when Facebook was being developed the objective was: ‘How do we consume as much of your time and conscious attention as possible?’ It was this mindset that led to the creation of features such as the ‘like’ button that would give users ‘a little dopamine hit’ to encourage them to upload more content...” (2017).

This phenomenon is not limited to Facebook but is endemic across modern social media platforms. Take another example, TikTok: a writeup from the Lowy Institute characterised its rapid growth in the West as being

---

14 This has parallels of de Beauvoir’s analysis of the ‘Narcissist’ in The Ethics of Ambiguity (De Beauvoir 1962)
due to “the explosive success of its specialised recommendation system... [which] captures even our passive and subtle behavioural patterns to teach its algorithms about us in real time, as we consume videos” (Zha 2020).

Engagement on social media can be measured\(^\text{15}\) by quantitative metrics, such as time spent on social media, number of posts/retweets/likes, and so forth: enabling it to be leveraged for, amongst other things, commercial value from efficient targeting of advertisements to their users, continuing use of the platform, etc. Celebrities, I argue, have an advantage with these optimisations being prioritised; consider Section 2.0’s definition of the Asymmetry of Influence (Cheong 2019). Consider popular users\(^\text{16}\) such as US President Barack Obama (@BarackObama on Twitter, with over ~130 million followers, while his account is following only about ~589,200 users); soccer player Cristiano Ronaldo (@cristiano on Instagram, with ~342 million followers / 480 following), or dancer Charli D’Amelio (@charlidamelio on TikTok, with ~124.3 million followers / 1235 following). Hypothetically speaking, even if a mere 0.001% of users who follow such celebrities act on their social media postings, say by considering their viewpoints or ideological leanings (Zollman 2012; Cheong 2019); the net result is that this figure still amounts to a few thousand users (Leins and Cheong 2020)\(^\text{17}\)! This asymmetry is not afforded to the average user, as they do not have this critical mass of influence: social networks tend to follow a power-law distribution, which means that the majority of users will tend to have a few followers only, and only a select few can command many followers\(^\text{18}\).

4.2. Celebrity Politicians (or Politician Celebrities) - direct effect on the polis

As noted in the introduction (Section 1.0), the definition of traditional celebrity now encompasses politicians and decision makers. I argue, with a contemporary example, that the interconnectedness of the social media ‘crowd’ can provide an avenue for celebrity politicians to drum up real-world, offline, influence - using an appeal-to-authenticity.

---

\(^{15}\) This quantification of engagement is in itself problematic due to many reasons. Thi Nguyen (2021) summarises this clearly in the case of Twitter, which can easily be generalised across social media platforms with the same features: “Twitter shapes our goals for discourse by making conversation something like a game. Twitter scores our conversation. And it does so, not in terms of our own particular and rich purposes for communication, but in terms of its own pre-loaded, painfully thin metrics: Likes, Retweets, and Follower counts. And if we take up Twitter’s invitation and internalize those evaluations, we will be thinning out and simplifying our own goals for communication.”

\(^{16}\) All statistics current as of time of writing, from the official Twitter and Instagram websites respectively. The summaries are at [https://en.wikipedia.org/wiki/List_of_most-followed_Twitter_accounts]; [https://en.wikipedia.org/wiki/List_of_most-followed_Instagram_accounts]; and [https://en.wikipedia.org/wiki/List_of_most-followed_TikTok_accounts]

\(^{17}\) From the analysis (Leins and Cheong 2020) on the 2020 Twitter hack, it becomes apparent, from a structural sense, that if nefarious actors were to compromise “just a single high-profile account with over a hundred million followers, hackers could theoretically influence more than 1000 users... [If] this was a misinformation campaign, the original owner of the Twitter account could suffer a blow to their credibility.” This illustrates the power of a high-profile celebrity when it comes to rapidly engaging many individual users: unprecedented before the rise in social media use.

\(^{18}\) For technical details, see e.g. (Cheong 2013; Barabasi and Albert 1999).
As a primer, let’s briefly review the conception of celebrity politicians per Street (2004)’s pioneering work. A common way a real-world politician can attain celebrity status is “engage[ment]... with the world of popular culture in order to enhance or advance their pre-established political functions and goals” (Street 2004). This includes the adoption of social media strategies, as “…[c]onnecting directly with voters via post-broadcast or new media is a growing trend…” (Wood, Corbett, and Flinders 2016).

I shall illustrate this with US President Donald Trump’s use of Twitter as a prime example. Trump has been recognised as a celebrity politician (Street 2019; Schneiker 2019). By using the Twitter handle @RealDonaldTrump to tweet his thoughts, Trump can directly reach followers with fiery, often false or misleading, rhetoric. With close to 89 million Twitter followers, before being permanently suspended by Twitter in early 2021, he was able to reach his followers, influence, and change popular opinion (Jetter, Lewandowsky, and Ecker 2020) on a large scale. His tweeting behaviour has been documented to “dole out criticism and ad hominem attacks... [and] In the immediate aftermath of Trump’s tweets, levels of severe toxicity and threats increased” (Brown and Sanderson 2020). From a public health perspective, Trump’s controversial use of “…Twitter to promote hydroxychloroquine as a treatment for COVID-19” (Chicago Tribune 2020) illustrates further the damaging influence of online celebrity.

Even when social media companies resorted to fact-checking Trump’s social media posts – which were ascertained to be factually incorrect (Ingram 2020) – by attaching a label to that effect; the factcheck itself increased Trump’s influence as a celebrity. “To Trump and his supporters, the labels are at turns an afterthought, a speed bump or a badge of hono[u]r from social media companies that want to deny him re-election by enforcing their rules” (Thomas 2020). The nett effect is that when politicians gain celebrityhood online, with unfettered access to their supporters on social media, they are able to influence hearts and minds and turn public opinion against the tide.

4.3. The Effect of ‘The Crowd’

We have seen in Section 4.2 that due to the technological advantage of social media (including its structural advantage of a highly connected network of followers, compared to pre-social media days) celebrities have an accelerated ability to influence the ‘crowd’. On the other hand, the ‘crowd’, I now argue, has the ability to do damage unto one of its own.

To complement social network theory, my argument now focuses on two interrelated areas: (i) social psychology and (ii) existentialist philosophy, with emphasis on crowd phenomenology as well as Kierkegaardian critiques of the ‘press’ as opinion leader.

See also (Siegel 2020) for an in-depth review of online hate speech, with emphasis on online social media sites.
Firstly, social media, as argued in Section 4.1, enables the rapid engagement of individuals as a collective, which isn’t at all possible in traditional pre-social-media circumstances. Social psychology has long studied the emergent behaviour of crowds and how individuals are ‘lost’, so to speak, when in a crowd setting. Examples are numerous: ‘safety in numbers’ an individual assumes in a crowd (Le Bon 1897); individuals acting out as their behaviour is unrestrained due to anonymity (Festinger, Pepitone, and Newcomb 1963); and the ‘diffusion of responsibility’ (Darley and Latané 1968). Pre-social media, a real-world presence, or some form of catalyst is required to get everyone on the same page, so to speak. The ubiquity of social media has accelerated the formation of the ‘crowd’, since one’s mere ‘following’ of a celebrity on TikTok (for instance) opens up avenues for their engagement with other fans.

From an existentialist philosophical outlook, the ‘crowd’ has been denounced as an undesirable state. Kierkegaard (and fellow existentialists) have denounced the ‘crowd’ as ‘untruth’ (Tuttle 1996), contra the individual (Golomb 1995; Nerina Jansen 1990). Taking, say, Kierkegaard’s work in focus: reasons include the dilution of responsibility of the crowd, as the “crowd either renders the single individual wholly unrepentant and irresponsible, or weakens his responsibility by making it a fraction of his decision” (Kierkegaard 1846), creating an “abstract whole” created when mindlessly engaging with social-media-as-crowd’ (Kierkegaard 1962). Part of his dislike of the ‘crowd’, I argue, is based on Kierkegaard’s experience of being attacked by the popular press of the day (The Corsair newspaper) as the arbiter of public opinion: see e.g., (Nerina Jansen 1990).

Comparing Kierkegaard’s experiences with modern day social media, this “abstract whole” has a far unprecedented potential for growth, expanding almost at the rate of 500 million per year, or about 200-500 times the entire population of Denmark20 in Kierkegaard’s day. In the present day, any social media user can potentially be in Kierkegaard’s position, if unwanted attraction is somehow foisted upon oneself; or otherwise become one of the anecdotal ‘nameless’ supporters of The Corsair, by following the tide of popular opinion. With those views in mind, how do celebrities online catalyse invectiveness and therefore harm – with or even without their consent or intention – amongst their online followers? The following subsections will illustrate just two cases of how social-media-mediated crowds can entangle other individual users, sometimes inadvertently.

4.4. Unexpected Invectiveness Towards ‘Bystanders’, even when Disavowed by Celebrities.

This invectiveness effect not only affects those engaging with a celebrity via social media, but also, in certain circumstances, harm innocent bystanders who have nothing to do with the initial social media ‘crowd’ in the first place, especially on high-profile celebrity online feuds (more about this in Section 5.0).

---

In this section, my argument will cover the vilification of those who do not subscribe to the fans’ perceived worldview of/around a celebrity. These Others can be real-world bystanders, or other social media users, which have nothing to do with the celebrity in the first place. This phenomenon will be illustrated via two examples.

First is the abuse by purported social media fans of tennis star Novak Djokovic, directed towards a real-world individual, despite the fact that this event happened in the real, offline, world. In 2020, an umpire, in a match Djokovic was in, “got hit in the throat with a ball from his racquet while she was doing her job” (White 2020), resulting in a default against Djokovic. The backlash towards the umpire, however, was rife on social media:

> “People purporting to be fans of Serbia’s world No 1 men’s tennis player Novak Djokovic have sent death threats and other vile abuse to ... [the umpire] ... How has it come to this that someone who is the victim is pilloried as a pariah, referred to as “Tennis Karen”, mocked for overreacting, sent abuse over her dead son and death threats of her own?... There is a strong chance that many are just in it for the social media pile-on, the mob frenzy, the madding crowd, the sense of power that brings and the corresponding permission to go even further. 
>
> *The anonymity of social media allows everyone the chance to stick the boot in with little to no chance of consequence.*” (White 2020), *emphasis mine.*

Despite Djokovic’s denouncement and rejection of his fans’ actions, the damage resulting was clear: an act of cyberbullying, fuelled by the nameless ‘crowd’ against an innocent third-party, which was accelerated by social media.

Another example of this can be seen in politics, especially during the tense atmosphere of the COVID-19 pandemic. During the recent COVID-19 pandemic in Australia, Victorian Premier Dan Andrews’ daily press conference is “...watched and evaluated by an audience at home,” as reported by The Guardian (Wahlquist 2020). However, a small segment of Twitter users have been abusive to reporters and writers from both sides of the political spectrum, simply because they had views critical of Andrews: such as a journalist who “has attracted the ire of Andrews supporters for criticising several elements of the state’s coronavirus response” (2020). Interestingly, the presence of political journalist, “political operative” and “celebrity pundit” (Muller 2020) Peta Credlin at Andrews’ press conferences - to ask difficult and “politically motivated” (2020) questions - similarly has Twitter users either defending or attacking her *modus operandi* online.

In the cases above, the original celebrities themselves are *not* part of the social media engagement, and obviously do *not* condone online acts of bullying. The ‘crowd’ on social media, fuelled by the Asymmetries (Section 4.0)

---


22 See e.g., https://twitter.com/search?q=peta%20credlin%20press%20conference for examples of invective and/or divisive tweets from both sides of the political spectrum.
has the ability to cause damage, to both innocent third parties and those caught in the maelstrom of social media chatter. This phenomenon is not new: social media posts or offline actions becoming ‘viral’ and attracting uninvited attention have existed for quite some time.\(^2\)

5.0 Discussion: Towards an Ethics of Online Celebrity?

Thus far, I have explored the structural and systemic factors which can accelerate potential harms of social media engagements (and the online fan-celebrity relationship) in Section 2.0. I have, in Section 3.0, discuss the recent trends of celebrities invoking ‘authenticity’ as a normative ideal on social media, when said use runs anathema to the existentialist’s ideal of authenticity (especially Sartre’s and De Beauvoir’s). The prior section (Section 4.0) has discussed how harms can manifest from social media activity/engagement to different stakeholders: from those who choose to engage with said celebrity online, to completely innocent bystanders in the real world; as well as current and future social media users (due to the underlying changes in their epistemic and social environment, in the case of, say, Trump).

I now attempt to propose some ethical considerations for celebrities’ usage of social media, based on my prior analyses. Upfront, this attempt faces a big challenge: the subscription to existential ideals necessitates a “subjectivist view of morality, which is inconsistent with any attempt to assert objective ethical values” (Crowe 2006). However, one can agree that the real-world harm to social media users, as documented in prior sections, causes harm, and therefore requires mitigation. I will propose ethical duties that online celebrities need to adopt, and choices they can make, in order for such harms can be ameliorated. I do this by unpacking the human, technical, and structural factors behind celebrity-catalysed harms.

5.1. Performative Acts, Authenticity, and Harm?

Existentialists, especially Sartre, deem *bad faith* on one who attempts to ‘roleplay as authentic’, celebrities inclusive (as seen in Section 3.0-3.2). The stereotypical example is in the ‘waiter who plays the role of a waiter’ vignette presented by Sartre in *Being and Nothingness* (Sartre 1969), which when applied to social media authenticity, can be characterised as a denial of free transcendence to do what’s best, by assuming and locking in a facticity of ‘authentic celebrityhood’.

Whether or not a celebrity is behaving authentically is beyond the scope of this discussion. What is concerning is that existentialists such as de Beauvoir, in developing a “conception of authenticity” (Varga and Guignon 2014), caution against exploitation of others as they can curtail others’ freedom and stymie others’ life-projects (De Beauvoir 1962). Hence, a conception of ethical, freedom-producing authenticity finds its practical application in the backdrop of my analysis; it has hitherto been employed in areas ranging from medical ethics, in forging an

---

\(^2\) See e.g., [https://en.wikipedia.org/wiki/Internet_meme](https://en.wikipedia.org/wiki/Internet_meme) for a history.
effective doctor-patient relationship based on “mutual trust and responsibility” (Arnason 1994); to a “necessary starting point for” ethics in organisations (Freeman and Auster 2011), insofar as considering employees’ “human freedom [as...] a supreme status [and ...] accorded the respect it is due in the design and practices of contemporary business” (Jackson 2005).

Going back to the notion of ‘authenticity’ in the celebrity-fan relationship on social media, it is indeed problematic if the end result is to use their fans as mere objects, with an implicit impingement of the Other’s freedom. This can range from acts as simple as getting fans to increase quantifiable engagement counts (likes, retweets, etc) and increasing sales and profits from fan engagement; all the way to fans being involved in unethical behaviour. With this in mind, what can we establish?

5.2. Establishing Boundaries to the Performative ‘Game’

I have no substantive objection to a celebrity-fan relationship which is a Goffman-esque (1959) performance, a la pre-social media – apart from the unresolvable tensions between ‘celebrity authenticity’ and ‘existential authenticity’ – as long as fan expectations are set accordingly. Quoting Jackson (2005), “the authenticity with which a [person]... approaches moral situations depends on the degree of consciousness he or she has of the various choices at stake”. Celebrities thus need to consider the real risks of harm from the range of affordances allowed by social media, e.g., ability for people to send abusive messages at a rapid scale.

As a first desideratum, celebrities need to make clear the performative notion24 of ‘celebrity online’ as similar to a kind of ‘game’, where both celebrities and their followers consciously know that it is all a mere performance and not real, but mutually choose to play along, while establishing clear boundaries. One very good example of establishing boundaries in the performative game is the Federal Trade Commission’s requirement for online ‘influencers’ on social media to make relevant disclosures when endorsing products (Carpenter and Bonin 2021; Federal Trade Commission 2019). This requirement somewhat restores autonomy of the fan in the online celebrity-fan relationship as “it allows people to weigh the value of ... endorsements” (Federal Trade Commission 2019) made by celebrities in their online communiques.

Of course, one major caveat to this is that celebrities are required to adhere to external norms, in this case, those prescribed by legislation. Can celebrities come up with an implicit norm25 to ensure that the ‘game’ is played

---

24 I’d like to especially thank Simon Coghlan for these ideas.

25 Interestingly, when online celebrities clearly and opaquely transgress fans’ expectations of the fan-celebrity relationship, and when enough of their fans acknowledge this transgression, their fans will actually self-regulate by reassessing the celebrity-fan relationship and reasserting their autonomy. Take the example of Fine Brothers Entertainment, a successful YouTube celebrity duo who, in 2016, tried to trademark the “reaction video” format, causing numerous YouTube subscribers to unsubscribe en masse as a form of backlash. However, for the sake of completeness, it is important to note that FBE was itself subject to online abuse by fans during the 2016 controversy; and in 2020-2021, there were claims by former employees that FBE was engaging in discriminatory practices. See https://en.wikipedia.org/wiki/Fine_Brothers_Entertainment.
correctly by fans? This is an important question as the following subsection highlights situations where the performative game is transgressed, and how celebrities should react.

5.3. Online Celebrity and Ethical Duties

The fanbase self-regulation and external norms on celebrities (due to legislation) discussed in the prior section are diametric opposites to the disastrous pile-ons on social media. Where the previous section outlines relatively lesser harms to the fans, we now consider cases where online invectiveness has the ability to cause significant harms, and ethical requirements for online celebrity-fan engagement. In the existentialist’s credo, it might not be an online celebrity’s fault that one of their fans are caught up in a disastrous, yet unpredicted and unsanctioned, pile-on – however the celebrity, constituting the other half of the celebrity-fan relationship, is obliged to take responsibility (Cox 2012; Jackson 2005).

Through the lens of a celebrity’s performative acts, there is a need for them to be cautious about their actions, and constantly be reminded of the fact that technology affords a rapid means to engage a mass audience, a la Kierkegaard’s experiences with the press. As a worst-case scenario, take the example of Donald Trump (first raised in Section 4.2): the effects of Trump’s invective tweets in provoking the January 2021 US Capitol insurrection has led Twitter Inc to intervene by permanently suspending Trump’s Twitter account, as it was deemed to have a “risk of further incitement of violence” (Twitter Inc. 2021).

One might argue that Trump’s actions were politically motivated. What if, however, the objective behind invective social media activity by an online celebrity is part of their Goffman-esque performative acts? Take the example of celebrity feuds online, where several celebrities take part in a public, online, feud, and where their fans actively take part in the online discourse by supporting their chosen celebrity and piling on the other26. Almost inevitably, fans will be caught up amongst the maelstrom and can find themselves targeted in a meta-feud (e.g., a separate Twitter ‘pile-on’ between fans themselves). Another more dangerous variant is when an online celebrity unjustly directs their attention to another non-celebrity user, and in doing so, implicitly mobilising their fans in the attack. In 2018, a freelance writer was targeted by Nicki Minaj and her fans after merely posting a tweet “about Ms. Minaj’s recent musical direction” (Coscarelli 2018); what resulted was unjustified chaos. Minaj’s “superfans... attacked [the freelance writer]... Then, galvanizing them further, Ms. Minaj chimed in...” to the point where the writer “received thousands of vicious, derogatory missives across Twitter, Instagram, Facebook, email and even her personal cellphone” leaving her “physically drained” and “mentally depleted” (2018).

---

26 Examples of this include: the 2016 feud between Taylor Swift and Kim Kardashian, which spilled across social media platforms, with the hashtag #KimExplosedTaylorParty (sic); a few other celebrity users and many more fans of either side participated in the pile-on. See e.g., https://globalnews.ca/news/2830882/kim-kardashian-kanye-west-reignite-taylor-swift-feud-on-social/ and https://www.nytimes.com/2016/07/19/arts/music/taylor-swift-kim-kardashian.html.
Online celebrities need to realise that the damage they can inflict goes far beyond that of the original celebrity target of their ire. “The new media playing field... still tilts toward the powerful. ... as more charged interactions can leave those who have earned a star’s ire... reeling as eager followers take up the celebrity’s cause” (Coscarelli 2018). Is the performative aspect of the initial feud worth having the possibility of fans’ freedom and well-being impinged? As for calling out an individual online, this is not only an abuse of the online celebrity status, but actual mental, physical, and emotional harms levied on their target, especially in an “entertainment media landscape that blurs the line between journalistic coverage and promotion” (Holmes 2018), amplified by the reach and velocity of social media is concerned. However, as of now, there is no formal sanctioning mechanism which could be used to establish ethical conduct. One way to go about it is for fellow high-status online celebrities to condemn such behaviour (which in itself raises the possibility of escalating feuds); or for celebrities who have initially made an erroneous choice to choose to take responsibility, e.g., the artist Lizzo who publicly apologised for inappropriately putting a delivery worker “on blast” via Twitter (Holmes 2019).

Now, what if, however, harms against individuals (on- and off-line) were entirely unintentional, or unforeseen? The Djokovic example from Section 4.4 comes into question. Even though the actual incident was in the real-world, and the online pile-on was entirely caused by the fans themselves, Djokovic is right in rejecting and denouncing his fans’ actions. What else can be done? For one, Djokovic can go further by publicly ostracising said abusive fans, by e.g., blocking them on social media. A counterargument to this would be that doing so would be superogatory; but to the existentialist, it is a fundamental choice to be considered.

Online celebrities need to normalise a culture of ethical behaviour and set expectations of their fandom online. They need to realise that, online fans have different motivations: from innocently co-creating a Goffman-esque performative act; to attempts (albeit severely misguided) in protecting their celebrity’s image (Holmes 2018; Coscarelli 2018); to merely acting recklessly due to the anonymity and protection afforded by the ‘social media crowd’ that is unprecedented in non-online forms of celebrity fandom (Section 4.3). Back to the question of authenticity, sometimes the tension between a celebrity’s veridical self and public presentation needs to be embraced 27: a better form of existential authenticity can be presented by acknowledging the presence of both to their fans and followers and having the celebrity’s veridical self ‘break the fourth wall’ in addressing ethical concerns related to the performative online self. An analogy would be for an actor, whose onscreen action-hero character unfortunately causes indirect harm (say, for kids who are injured in real life after copying the stunts they see on TV), to address her fans as her veridical self in order to discourage such behaviour.

6.0 Conclusion

My paper has argued that the structural and technological features of social media have created an environment that triggers a need for celebrities to uphold a purportedly authentic public persona, even if they have to (ironically) curate an existentially inauthentic image of themselves by doing so.

27 My thanks to the presenters at the Philosophy of Fame & Celebrity Workshop for this particular point.
I have also argued that the two asymmetries of social media have led to unexpected (and accelerated) formations of ‘crowds’ online, which not only cause harms to a celebrity’s interlocutors online, but also to innocent bystanders who have nothing to do with the entire discourse whatsoever: reminiscent of innocent bystanders studied in traditional social psychology, as well as existential philosophy, to a certain extent.

I have concluded by underlining ethical responsibilities that online celebrities need to be cognisant of, when engaging their fans both off- and on-line, by emphasising that destructive side effects of online celebrity fandom have no prior precedent in the offline world pre-social media, and that the notion of authenticity in an online celebrity’s public persona needs to be reconsidered, with an emphasis on making the right choices to maximise others’ freedom and protect their right to not be harmed.

References


———. 2018. “Experimental Investigations of #authenticity Online.” In 2018 Australasian Association of
Philosophy/New Zealand Association of Philosophy Conference.


Communication Is Not Fulfilling.” *Philosophy & Technology.*


Marwick, Alice E. 2013. *Status Update: Celebrity, Publicity, and Branding in the Social Media Age.* Yale University Press.


Siegel, Alexandra A. 2020. “Online Hate Speech.” *Social Media and Democracy.* https://doi.org/10.1017/9781108890960.005.


Thomas, Phil. 2020. “Trump Threatens Twitter and Facebook over ‘illegal Situation’ as US Coronavirus Death


