Philosophy & Technology (2022) 35:70
https://doi.org/10.1007/s13347-022-00567-7

RESEARCH ARTICLE

Social Media and its Negative Impacts on Autonomy

Siavosh Sahebi[1] · **Paul Formosa**[1,2]

Received: 29 March 2022 / Accepted: 14 July 2022
© The Author(s) 2022

Abstract
How social media impacts the autonomy of its users is a topic of increasing focus. However, much of the literature that explores these impacts fails to engage in depth with the philosophical literature on autonomy. This has resulted in a failure to consider the full range of impacts that social media might have on autonomy. A deeper consideration of these impacts is thus needed, given the importance of both autonomy as a moral concept and social media as a feature of contemporary life. By drawing on this philosophical literature, we argue that autonomy is broadly a matter of developing autonomy competencies, having authentic ends and control over key aspects of your own life, and not being manipulated, coerced, and controlled by others. We show how the autonomy of users of social media can be disrespected and harmed through the control that social media can have over its users' *data*, *attention*, and *behaviour*. We conclude by discussing various recommendations to better regulate social media.

Keywords Autonomy · Social Media · Manipulation · Exploitation · Addiction · Relational Autonomy

1 Introduction

In recent years, there has been a growing interest in the literature into the ways in which digital technologies, and specifically social media, impact on the autonomy of its users (e.g. Smith, 2020; Susser et al., 2019). While the impacts of new technology on autonomy have been explored in areas such as Artificial Intelligence (Floridi et al., 2018) and social robotics (Formosa, 2021), much of the literature that explores the impacts of social media on autonomy fails to engage in depth with the philosophical literature on autonomy. This has resulted in a failure to consider the full range of impacts that social media might have on the various aspects of autonomy.

✉ Paul Formosa
Paul.Formosa@mq.edu.au

[1] Department of Philosophy, Macquarie University, Sydney, NSW, Australia

[2] Centre for Agency, Values and Ethics, Macquarie University, Sydney, NSW, Australia

Published online: 27 July 2022

A deeper consideration of these impacts is therefore needed, given the importance of both autonomy as a moral concept and social media as a feature of contemporary life. This paper will explore these issues as follows. First, we survey different conceptions of autonomy found in the philosophical literature and argue that autonomy is broadly a matter of developing autonomy competencies, having authentic ends and control over key aspects of your own life, and not being manipulated, coerced, and controlled by others. We then detail what we mean by social media and focus on the control that it can have over users' *data*, *attention*, and *behaviour*. We outline three related but distinct autonomy harms this control can inflict on the users of social media: (1) disrespecting their autonomy; (2) interfering with the exercise of their autonomy; and (3) impairing the maintenance or development of their autonomy competencies. Finally, we end with recommendations for better regulating social media to limit the harm it can have on the autonomy of its users.

2 Autonomy

The philosophical literature concerning autonomy is vast and multi-faceted (see e.g. Klenk & Hancock, 2019; Darwall, 2006; Formosa, 2011, 2013; Killmister, 2013a, 2013b; Friedman, 1986; Lavazza & Reichlin, 2018; Meyers, 1987; Schneewind, 1998; Mackenzie, 2014a, 2014b; Walker & Mackenzie, 2020; Raz, 1986). To make sense of this vast literature, it is helpful to note that theories of autonomy are typically divided into procedural, substantive, and relational views.

Procedural theories of autonomy hold that specified content-neutral procedures are necessary and sufficient conditions for autonomy. On such views, "what one decides for oneself can have any particular content" (Dworkin, 1988, p. 12). The relevant procedures typically involve choosing in accordance with one's own second-order desires (Frankfurt, 1971), values (Watson, 1975), practical identity (Korsgaard, 1996), or after engaging in critical reflection. What matters for autonomy on such views is that the proper process is followed, whatever the content of the choice. We can better understand such views by considering the challenges they face in dealing with cases of so-called oppressive socialisation. Oppressive socialisation is "a type of socialisation that reliably discourages or fails to properly develop the required autonomy competencies in agents subject to that socialisation" (Formosa, 2013, p 203). Autonomy competencies are those skills, capacities, and powers that agents need to be able to act autonomously, such as the ability to reason and critically reflect on their values, imagine different alternatives, develop a conception of the good, and regard themselves as self-directing agents worthy of respect. Oppressive socialisation poses a problem for procedural accounts since the second-order desires, values, or practical identity of a person may be caused by their oppressive social setting. Furthermore, oppressive social milieus can also influence the critical reflection and reasoning skills that an agent uses when making and reflecting on their own choices (Killmister, 2013a, 2013b). For example, an agent that chooses in a procedurally correct manner by acting in terms of her practical identity after engaging in critical reflection may nonetheless lack autonomy if her practical identity is the result of false beliefs, such as beliefs about the inferiority of women, that

are the product of her oppressive socialisation and are objectively flawed (Killmister, 2013a, 2013b, p. 514). Procedural historical accounts of autonomy, such as the view defended by Christman (1991), attempt to deal with this problem by arguing that a person's values can be seen as autonomous only if that person would not resist the acquisition of those values *were* they to become aware of the historical process that led them to hold those values, and they were able to engage in self-reflection that was minimally rational and not based in self-deception. In cases where a person is brainwashed, manipulated, or oppressed into holding a certain value or end, this condition is presumably not met, and the person thereby fails to count as autonomous.

In contrast, substantive accounts of autonomy set out substantive conditions that must be met for a choice or person to count as autonomous. Substantive theories of autonomy can be further classified as weak or strong (Mackenzie & Stoljar, 2000). Weak accounts set out certain conditions which "somewhat limit and inform the content of" choices that can count as autonomous while still maintaining a significant degree of content neutrality (Formosa, 2013, p. 16), whereas strong accounts fully determine (or at least more strongly limit) the content of autonomous choices. On weak accounts of autonomy, the conditions which must be met for choices to count as autonomous typically include the "evaluative attitudes of self-respect, self-love and self-esteem" (Formosa, 2013, p. 17). We can think of these self-attitudes as part of a broad suite of autonomy competencies, along with related skill such as critical reflection, that must be present to a sufficient degree for a person and their choices to count as autonomous. While strong substantive theories of autonomy can deal with the problem of oppressive socialisation by offering objective normative conditions that provide a standard independent of oppressive milieus, they face the strong burden of identifying and justifying objectively good ends and values. Further, the lack of any significant degree of content neutrality in strong accounts can be seen as positively limiting autonomy, since this greatly limits the ends that agents are able autonomously to set as their own.

Weak substantive theories of autonomy, in focusing on a suite of autonomy competencies and related skills and self-attitudes, are best understood in the context of a relational approach to autonomy. Relational accounts argue that autonomy, and its associated competencies, is a "socially constituted capacity", since socialisation plays a necessary role in its development and utilisation (Rogers et al., 2012, p. 23). The development, cultivation, maintenance, and expression of these autonomy competencies are socially scaffolded. However, this development and utilisation can be undermined if interpersonal relationships, as well as social and political structures, are oppressive, exploitative, or unjust. According to Mackenzie's (2014a, 2014b) relational account, there are three dimensions to autonomy: self-determination, self-governance, and self-authorisation. Self-determination concerns one's freedom to make the relevant choices which reflect control over one's life; for example, having the freedom and opportunity to choose between a multitude of options in life. Self-governance pertains to one's decision-making abilities being reflective of one's identity and values; for example, by making decisions which reflect the preferences of agents which are authentically their own. Our preferences and values are 'authentically' our own if, roughly, we would either endorse them upon critical

reflection or acknowledge them and take responsibility for them (Walker & Mackenzie, 2020). Self-authorisation necessitates taking authority over one's identity, values, and choices; for example, having the self-esteem to internalise values and to act in accordance with them (Mackenzie, 2014a, 2014b, pp. 17–19). Self-authorisation thus also has a social aspect insofar as it requires us to see ourselves as a respect-worthy agent entitled to set our ends amid a community of fellow agents. We will work with a weak substantive relational account of autonomy below, since such an account emphasises the key role played by the social environments, of which social media is an increasingly important part, within which agents act. It also allows us to focus on the ways that social environments impact, positively or negatively, the key autonomy competencies of self-respect, self-love, and self-esteem, and the beliefs and critical reasoning capacities that underwrite the exercise of agents' autonomy.

To explore these impacts, it will also prove useful to briefly differentiate programmatic or global as opposed to episodic impacts on autonomy. Programmatic (Meyers, 1987) or global autonomy (Mackenzie, 2014a, 2014b) focuses on a person's overall life and ability to carry out their life plans. In contrast, episodic autonomy applies to a "particular situation" and is "confined to a single action" (Meyers, 1987, p. 265). We utilise these distinctions below as they help us to assess whether personal autonomy is being impacted by social media on a global level or only in one domain or choice. We also explore below cases where a person's autonomy is being disrespected without their autonomy competencies being harmed (e.g. when their choices are ignored but their autonomy competencies are not harmed), cases where their autonomy competencies are undermined or hindered even if their choices are respected (e.g. their self-respect is undermined but their choices are acted upon), and cases when their autonomy is both harmed and disrespected. Finally, given our focus on the potential dangers of social media, we focus here only on the negative impacts of social media on autonomy. Of course, there are also many important positive benefits for autonomy that social media has the potential to make possible, such as allowing us to make new self-esteem boosting social connections, learn new information, and provide tools to help us to better realise our social ends. An example of such positive autonomy benefits of social media is the Arab Spring of the early 2010s, where social media enabled citizens to pursue their self-given ends by giving them tools to organise, connect with one another as well as with journalists and those abroad, and have their voices heard (Comunello & Anzera, 2012; Frangonikolopoulos & Chapsos, 2012; Howard et al., 2011). However, given our focus, we will not be exploring these positive impacts further here, although it should be noted in any overall assessment that these exist.

3 Social Media and its Negative Impacts on Autonomy

Social media includes "applications, such as blogs, microblogs like Twitter, social networking sites, or video/image/file sharing platforms" (Fuchs, 2017, p. 34) that provide "digital infrastructures that enable two or more groups to interact" (Srnicek, 2017, p. 31). Social media is a digital environment in which individuals, groups or organisations can interact and communicate with one another via the sharing of

content and information, in text, photos, images, videos, or other digital formats. It is informative to differentiate traditional media and social media to better understand how the different mechanisms function and why they may be more problematic on social media than with its more traditional counterpart. Social media platforms personalise content based on the individual user's digital footprint as well as having the ability to target specific portions of the population based on data analysis (Benkler et al., 2018; Mittelstadt, 2016). This gives social media companies and their customers a significant advantage over traditional media since they can tailor their messaging and advertising to suit the users they are aiming to reach, as opposed to traditional media where all those who consume it receive the same content, messages, or advertisements. This has resulted in advertising money shifting from traditional media to social media. In Australia, for example, the money spent on social media advertising increased from 6% in 2012 to 52% in 2017 (Malesev & Cherry, 2021, p. 68). Meanwhile global digital advertising spending increased by 12.7% in 2020 to $378.16 billion, while traditional media advertising spending saw a decline of 15.7% in the same period (Cramer-Flood, 2021). A further distinction between social media and traditional media is the ability for social media companies and their customers to enter the "private, invisible realm", by entering "into an increasingly personalised, private transaction" which is removed from a public realm that is potentially accessible and open to all, and which thereby avoids much of the public scrutiny to which traditional media is subject (Benkler et al., 2018, p. 273). This leads to a related dimension, namely that as Benkler et al., (2018) argue, "disinformation campaigns and legitimate advertising campaigns are effectively indistinguishable on leading internet platforms" (p. 273). This is not only due to the increased sophistication with which malicious actors disseminate (mis)information and content, but also due to the comparative lack of public scrutiny which social media platforms are subject to because of the personalised and private nature of users' relationships with the platform.

As noted in the previous section, socialisation is a key influence on autonomy, and particular patterns of socialisation not only exist within the context of social media, but are arguably exacerbated by it. For example, socialisation tends to influence how social agents perceive the world, be it the standards of beauty that society aspires to, what defines success, or what career paths are seen as desirable. This is also evident within the realm of social media where the exhibition of wealth and material possessions influences perceptions of beauty, social status, and social capital through systems such as 'likes' and 'shares' which exist within social media (Fuchs, 2017, p. 36). Furthermore, a key component of socialisation is education and knowledge, which allows for the socialised to make informed decisions. Social media plays a central role in the dissemination of information and knowledge to users, with studies finding that 68% of American adults 'at least occasionally' access their news through social media (Matsa & Shearer, 2018). This is troubling given the rise of misinformation and fake news on social media (Ha et al., 2021). To explore these issues in a systematic way, we now consider the impacts of social media on autonomy through its control over user *data*, *attention*, and *behaviour*. This focus on control is important since autonomy can be understood as requiring (among other things) freedom from control by others. In exploring these areas in Sects. 4 to 6, we

first outline the relevant form of control and then return to the impacts this has on autonomy at the end of each section.

4 Control over Data

As users navigate through and interact with one another on social media, they generate incredible amounts of data which are recorded and stored by the owners of the platforms. Zuboff (2019, p. 100) refers to users of such platforms as "human natural resources". In the case of platform capitalism, including social media, it is the users and their activities which provide the materials, data, which lead to profits. Platform capitalism refers to a form of capitalism focused on building digital infrastructure (i.e. the platform) that enables two or more groups to interact, and which typically builds on network effects to try to keep its users inside the platform (Srnicek, 2017). Platforms have become "an efficient way to monopolise, extract, analyse, and use the increasingly large amounts of data that [are] being recorded" (Srnicek, 2017, p. 31). This data is subsequently used by social media companies to generate profit, typically through selling advertising (although the precise details of their means of generating value in the forms of profit or capital differs among the platforms, these details aren't important here). As Fuchs (2017) notes, social media platforms "sell big data for advertising purposes. They are the world's largest advertising agencies that operate as big data collection and commodification machines" (p. 54). The loss of control by users over their data has, we shall show, important autonomy implications.

The first key issue this raises for autonomy is that there seems to be (at least the potential for) exploitation of users by social media companies. Fuchs (2011) argues that the concept of exploitation extends to social media users since they engage in the production of data and user generated content which is then used by social media companies to generate profit. This counts as exploitation because social media companies "do not (or hardly) pay the users for the production of content" or data and yet they generate profit, or surplus value, from this (Fuchs, 2011, p. 297). This argument is extended by Fuchs (2012) who argues that the "labour that generates audience commodity is exploited because it generates value and products that are owned by others [i.e., the social media platforms and not users who create the content], which constitutes at the same time an alienation process" (p. 705).

While *any* extraction of surplus value by the owner of capital may count as "exploitation" in a technical Marxist sense (Wood, 1995, p. 147), such exploitation only seems to be a morally problematic expression of disrespect for autonomy when further conditions are met. Typically, these further conditions include, one, the benefits rendered in exchange for the labour that produces the surplus value is unfair or unreasonably low or, two, when the labourer (in this case, the user of social media platforms) cannot reasonably refuse to participate in the exchange (Valdman, 2009) or only consents because some vulnerability of theirs is preyed on (consciously or not) by their exploiter (Wood, 1995). Exploitation counts as disrespectful of autonomy when it fails to treat the other agent as an equal party to determining consensual terms of interaction, and instead leverages a vulnerability to excessively benefit

oneself (Formosa, 2017, p. 107). Without committing to any particular account of exploitation, we can explore these various aspects to assess the extent to which social media users might have their autonomy wrongly disrespected through being exploited.

First, some may argue that users do receive a reasonable benefit in the form of a 'free' service in exchange for the value being generated from their data. Against this, Fuchs (2012) argues that social media users do not receive a "universal medium of exchange" which can then be used as they wish (i.e. money), but instead users are given "access to particulate means of communication whose use serves their [the corporations'] own profit interests" (p. 703). Whether this counts as unfair or unreasonable compensation for the value created by users is, as is often the case with such claims, difficult to determine, but there are certainly legitimate concerns here. In terms of the second point, there are also legitimate concerns over users' ability to refuse using the platform since they "may miss certain social contact opportunities" and therefore "suffer social disadvantages in society" (Fuchs, 2012, p. 704). If, at least in certain social circles, it is not reasonable for people to refuse to use popular social media platforms since otherwise they will miss out on too many important social opportunities, then it may not be reasonable for people in certain social circles to refuse to agree to whatever terms they are offered by social media companies. This sets them up to be exploited. Further, people's need for social contact and their fear of missing out (FOMO) (Kuss & Griffiths, 2017), combined with their vulnerability to accepting privacy terms without reading them (Nissenbaum, 2011), both point to specific human vulnerabilities that might be preyed upon by social media companies to exploit their users and thereby disrespect their autonomy.

However, even if users explicitly 'consent' to use social media, this does not mean that they have autonomously consented to these terms since their consent could be forced as a result of exploitation or they may not be properly informed. This later concern feeds into the response that social media users provide consent to have their data collected and used when they sign up and agree to a platform's terms of services. However, as noted above, humans are vulnerable to being exploited by such agreements as they often lack the capacity for informed and free consent. As Nissenbaum (2011) notes, the "ubiquitous regime of offering privacy to individuals on a 'take it or leave it' basis" (p. 35) leaves users with no real choice outside of not using the service, which we have already seen may not be reasonable to expect. Further, with the use of social media platforms becoming essential for many professional activities, there are further financial and economic costs, as opposed to social ones, which means that users may feel that they have no choice but to join the platform otherwise they will be missing out on too much. This can be in terms of missing out on key information or news, as it pertains to both personal news from family and friends or current affairs news, or the fear of being socially or professionally excluded for not being part of a social network (Kuss & Griffiths, 2017). This raises the question of whether users' decision to join social media platforms, if it is forced upon them, can count as an autonomous one to begin with.

Nissenbaum (2011) raises further concerns over what she refers to as the Transparency Paradox, which articulates the problem that faces informed consent within this context. This brings into question how accurate of an understanding users have

regarding their data and their control over it. The Transparency Paradox highlights the fact that if the terms of service – the information which is provided to users as part of their agreement with the service or platform – are too simple, then they fail to provide users with accurate information pertaining to the collection, control, and use of personal data. Alternatively, if those terms are too complicated, then users will not be able to understand the agreement, both due to a lack of technical and legal expertise and the time needed to clearly understand the agreement in its entirety. Both options raise serious concerns about whether users can be informed enough to be able to autonomously consent to social media use and the subsequent extraction of their data.

Of course, one might object that this stipulates an unrealistic standard for informed consent. Comparisons to medical ethics, where the focus on informed patient consent is central (O'Neill, 2002), are helpful. Surely we do not expect patients to have as detailed an understanding of their proposed medical treatment as their physician in order for their consent to count as informed? But, as O'Neill (2002) argues, it is because patients lack a complete understanding of what they consent to that trust is needed to buttress consent in medical contexts. However, the key difference is that patients can have strong grounds for trusting their physicians, in part due to their physicians' duty of care towards them and their professional obligations to maintain their medical competence. But it is precisely a duty of care and professional obligations that are lacking in the relationship between a social media user and their provider, and thus in the absence of strong grounds for trust, informed consent must do all the normative work. And the Transparency Paradox shows us that it cannot do all that work, at least for the vast majority of users who lack the relevant technical and legal expertise.

A further concern is the extent to which anonymisation is impossible in the age of big data. To this end, Barocas and Nissenbaum (2014, p. 51) argue that "anonymity is impossible" since, due to the vast amounts of data controlled by platforms, there still exists information which "uniquely distinguishes a person enough to associate those records to a specific individual" (p. 50). It is the combination of data from multiple sources which leads to this loss of privacy, and this further extraction of information from the combination of data is unlikely to have been autonomously consented to by informed users. This leads us to the next issue of concern: surveillance.

Zuboff (2019) defines surveillance capitalism as "a new economic order that claims human experience as free raw material for hidden commercial practices of extraction, prediction, and sales". Zuboff (2019) goes on to state that while Marx's critique of capitalism illustrated it as a "vampire that feeds on labor", instead of feeding on labor, "surveillance capitalism feeds on every aspect of every human's experience" (p. 9). One of the main consequences of surveillance is that users of social media experience a loss of privacy, especially in moments where they consider themselves to be alone and in private, such as when scrolling through Facebook (now part of Meta) alone in their own room (Molitorisz, 2020). As Reiman (1995) argues, a loss of privacy can have dire consequences. Specifically, it can lead to the "risk of extrinsic loss of freedom", meaning users are vulnerable to having their behaviours controlled by others, and the "risk of intrinsic loss of freedom",

meaning users may be susceptible to social pressures and unable to make important choices for themselves (Reiman, 1995, pp. 35–38). Further, the surveillance, or fear that they *might* be subject to surveillance, that users experience can curtail and prevent users from authentically exercising their autonomy in what they say and do on social media sites (and beyond). For example, concerns about surveillance might prevent a social media user from liking or posting certain content that expresses who they authentically are out of fear of social repercussions from others or concerns that their deeply personal information could be used by social media sites to target advertising at them. Evidence of this can be seen in a study looking at college-aged social media users in the USA which found that "imagined surveillance" on social media "steered their self-presentation practices" (Duffy & Chan, 2019; p. 119). Further, concerns around government surveillance of social media could also limit autonomy. Evidence of this exists in countries such as China where university students noted that their concerns over mass surveillance had an impact on their engagement with the digital world, including social media sites, where students report practicing self-censorship in these spaces (Shao, 2020). Social media surveillance has also been used in other countries such as Iran (Fassihi, 2009) and Bahrain (Jones, 2017) as a form of social control.

The loss of autonomous control over intimate data can also impact on the autonomy competencies of self-respect, self-love, and self-esteem. For example, if you attempt to read the terms and services to gain a better understanding of the data collection processes of social media sites, but do not have the technical and legal expertise to understand it, then your self-esteem could be negatively impacted. Further, recognising that social media sites could be expressing disrespect for your powers of self-government by exploiting you as a mere means to obtaining your data can make it harder for you to maintain your self-respect. This last point relates particularly to the self-authorisation aspect of autonomy, whereby we see ourselves as having the authority to set our own values and ends, as it is hard to see ourselves in those terms when we are subject to (potentially) exploitative terms that we do not autonomously consent to and (in many cases) cannot reasonably refuse. Being exploited in this way could impinge on our ability to regard ourselves as entitled to set our own ends. The surveillance aspect of the platform capitalism that fuels social media sites also challenges the self-determination aspect of autonomy, by making it difficult to authentically determine our own ends under the data-hungry gaze of various organisations who generate profit from our data. These various negative impacts can operate at both the episodic levels of autonomy, by interfering with individual choices that we make due to the gaze of surveillance, and the programmatic or global levels, by undermining our overall sense of ourselves as self-authorising agents with robust levels of self-respect and self-esteem. The impact of social media on self-esteem has been extensively researched with studies finding that increased social media usage among adolescents is associated with lower self-esteem (Steinsbekk et al., 2021; Woods & Scott, 2016) and this is further exacerbated by addictive use of social media (Andreassen et al., 2017; Hawi & Samaha, 2017). It should be noted, however, that the "true relationship between social media use and self-esteem is person-specific and based on individual susceptibility and uses" (Cingel et al., 2022, p. 1) and thus the temporality of social media's impact on self-esteem, and thus also

on autotomy competencies, can vary on a case-by-case basis. The impacts of social comparisons and social media addiction on self-esteem will be discussed further below.

5 Control over Attention

By controlling your attention, what you focus on, notice, or look at, social media can control its users, and by controlling its users thereby potentially disrespect and interfere with their autonomy. To critically analyse the ways in which social media uses control over users' attention to impact autonomy, it is important to understand the mechanisms that work to attract and exploit users' attention. The relevant literature concerning attention is wide-ranging and multi-faceted (see e.g. Citton, 2017; Rheingold, 2010; Daugherty & Hoffman, 2014; Boyd, 2010; Feng et al., 2015) and so in this section the focus will be on the specific facets of attention which are directly relevant to the context of social media. Citton (2017) provides a comprehensive account of attention by situating it within the economic model of "an attention economy" (p. 17). Within this context, we see a shift from control over the means of production to control over the means of attention, where attention is the much sought-after scarce resource. This is referred to as the 'Postulate of Limited Resources' which states that "the total quantity of attention available to humans is limited at any given time" (Citton, 2017, p. 48). The scarcity of attention becomes even clearer when we consider it within the context of social media, where the flow of content and information is almost limitless, while the time and attention we have available to us to consume that content is clearly limited. In this context Zulli (2018) argues that "the glance", which they define as "a quick, fleeting, and indiscriminate type of seeing", is a "key feature of what drives our attention economy" and allows us to examine "how digital technologies restructure user and economic behavior" (pp. 137–138). This means that social media companies (and their advertising customers) are in constant competition to attract as much of their users' attention as possible to generate profits. It is in this way that social media companies can design the architecture of their platforms to account for the glance to determine where users' attention is being directed. While it may be argued that users have the power to control what they see through the entities they follow on social media, this simplistic view neglects the complex mechanisms that function within social media, such as the ordering of content, the implementation of algorithms, and the sheer economic power that social media companies have to leverage those with high followers and attentional draw to promote the content that they want viewed by users. Furthermore, Myllylahti (2018) argues that there has been a power transfer from the publishers of news to the platforms which distribute news, which means that social media companies such as Facebook have control over "who publishes what to whom, and how that publication is monetised" (Myllylahti, 2018, p. 239). The control of social media companies over attention becomes even more evident when we consider that "news companies rely on platforms for the audience, and the whole business model of social media platforms is based on harvesting human attention which can be commodified" (Myllylahti, 2018, p. 241).

Algorithms are central to social media's control over attention, as algorithms have a significant impact on what social media users consume and engage with on the platform. It is precisely through the employment of algorithms, which dictate what users see on their social media feeds, that social media companies and their customers control the attention of their users, and through controlling their limited attention, control users in ways that they would not, on reflection, endorse. Recent Facebook leaked documents have shown that particular metrics, such as "meaningful social interactions", are prioritised by Facebook's algorithms since such metrics drive user engagement, keep people on the platform, and therefore bring in more advertising money (Milmo & Paul, 2021). Algorithms can influence users through "content personalisation" in which "content is filtered to fit the user's profile" (Mittelstadt, 2016, p. 4991). As Mittelstadt (2016) points out, algorithms which function in this way dictate what each individual user consumes and engages with, which can "undermine the fairness and quality of political discourse" (p. 4992). One way in which political discourse can be undermined in this way is through the creation of 'echo chambers' which leave users devoid of diverse ideas and beliefs, resulting in users only consuming particular political perspectives and not others (Worden, 2019, p. 240). This has implications for autonomy. For example, a recent Twitter internal report studied the algorithmic amplification of political groups in seven countries (the UK, USA, Canada, France, Germany, Spain, and Japan) on the Twitter home timeline which is determined by personalised algorithms (Huszár et al., 2021). Their findings were twofold: first, they found that among elected legislators the mainstream political right saw statistically significant higher algorithmic amplification compared to the mainstream political left in six of the countries in the study, Germany being the sole exception (Huszár et al., 2021). Secondly, statistically significant algorithmic amplification was found to favour right-leaning media sources within the USA. While Twitter's report claims that they are unaware of the cause of the above findings (Huszár et al., 2021), it is worth noting that the interest of social media companies is not to ensure that its users are provided with diverse viewpoints, political ideologies, or news. Their interest lies in ensuring they can maintain the attention of users to generate revenues through advertising and other means by showing users what they want to see or will cause them outrage. This is how harmful echo chambers are formed. In this case, the amplification of right-wing content and thinking is shown to dominate the political and ideological conversation on Twitter, determined by algorithms outside of the control or awareness of users. This has the strong potential to transform, through control over users' attention fuelled by opaque algorithms, the political views of its users in ways that users would not, after informed reflection on this transformative process, endorse as authentically their own.

Another concern with personalisation is that it can lead to radicalisation through increased exposure to extreme and irrational views, such as unfounded conspiracy theories. Alfano et al. (2020) states that in 2018 YouTube reported that approximately "70% of all watch-time spent on the site was driven by the recommender system" (p. 3). What is most troubling is that Alfano et al. (2020) report that YouTube's algorithms are working to keep people engaged on the platform for as long as possible by exposing some users to conspiratorial content which can result in radicalisation, although the radicalisation pathways differ for different search terms. Recent leaked documents from Facebook confirm the same kind of radicalisation taking place through Facebook's algorithms. Researchers within Facebook released

a report titled "Carol's Journey to Qanon" which found that a test account which was made to represent a 'conservative' mother was recommended conspiracy theory content by Facebook's algorithm within days, despite her interests including seemingly innocent categories such as parenting and Christianity (Mac & Frenkel, 2021; Zadrozny, 2021). What's more, after three weeks the account had become 'a constant flow of misleading, polarising and low-quality content' (Mac & Frenkel, 2021). It is reported that Facebook's staff have raised concerns regarding the polarisation they had perceived to be possible on the social media platform, but as Facebook whistleblower Frances Haugen put it, "the thing I saw at Facebook over and over again was there were conflicts of interest between what was good for the public and what was good for Facebook. And Facebook… chose to optimise for its own interests" (Paul & Milmo, 2021). The reason why radicalisation is a cause for concern when it comes to users' autonomy is that radicalisation can inculcate users with values and ends that they would not endorse after informed reflection. Specifically, many users would be unlikely to endorse those values and ends if they were aware that they acquired them through a process of algorithmic personalisation controlled by an external body looking to maximise its profits through seeking to monopolise their attention with radicalising content.

Another major challenge posed by social media which has the potential to inhibit autonomy is that of "Fake News", which can form part of the content which users focus on due to the control that social media companies have over their attention. Fake news portrays misinformation and inaccurate information about the world in the form of legitimate news, ensuring that the presentation and format of traditionally trusted media sources are followed (Levy, 2017, p. 20). A look at recent political elections can provide us with an insight into the problems caused by fake news. In the 2016 US federal elections, fake news was shared significantly more on social media compared to genuine news (Cohen, 2018, p. 140). During the same period, 115 fake news stories with pro-Trump sentiments saw 30 million shares on Facebook, while fake news stories which were pro-Clinton saw 7.6 million shares (Allcott & Gentzkow, 2017, p.212). Levy (2017) not only acknowledges that fake news has influences in the political sphere, but he stresses that the consumption of fake news, even when one is aware that it is fake news, poses serious concerns. This is due to the source of knowledge and the content of knowledge being separately stored in memory and thus, when the content of knowledge is brought to conscious awareness, it may lead to errors in memory when remembering the source of the knowledge. This results in the potential attribution of fake news to a reliable and trusted news source (Levy, 2017, p. 29). This is problematic for autonomy because it can lead users to falsely assume that they are consuming reliable sources of information which they can safely act upon, rather than (as is the case) information that they would not act upon or endorse if they were critically aware of its source and legitimacy.

Personalised advertising on social media is another way in which these companies influence the attention and behaviours of users, consequently impacting their autonomy. Consumer data being used for targeted advertising may seem harmless and even useful for customers by accurately informing them about offers that will

appeal directly to them. However, users are not being shown advertising designed to benefit them, but rather whatever advertisers want them to see and buy. We can take Winpenny et al.'s (2014) findings as an example of this, where Facebook in the UK exposed "89% of males and 91% of females" between the ages of 15–24 to alcohol marketing every month on average (p. 155). Considering many in this sample are children under the legal drinking age, this seems clearly unethical. It may be argued that this can be easily resolved by banning advertising which exposes minors to harmful products. However, this misses the bigger issue concerning targeted advertising on social media. The reason targeted advertising, unlike its less targeted offline counterpart, is so valuable from corporations' perspective and detrimental for users is that it can micro-target particular users at particular moments. For example, by tracking the profile of users Facebook can target vulnerable teenagers with advertising during moments when they are experiencing negative emotions, such as insecurity or feelings of worthlessness, to sell them products (Susser et al., 2019). By targeting an insecure teenager with products that are designed to superficially increase self-worth, a new watch for instance, social media companies and their customers can gain control over a user's attention to generate profit for themselves by exploiting a vulnerability of the autonomous decision-making processes of the user at a particular instance. This is problematic for autonomy because personalised advertising can be used to control users in ways that they do not understand as it leverages a platform's detailed knowledge of users and its ordering algorithms to systematically influence users outside of their conscious awareness. Personalised advertising can also lead to users partly losing control of themselves since they cannot focus on what they value most because their attention has been intentionally and persistently hijacked by social media companies (Vallor, 2015).

To illustrate this point more concretely, imagine a young woman who uses Facebook and Instagram, much like everyone else her age. As per all forms of media, certain perceptions of beauty and wealth are more evident than most, largely due to the way algorithms work as well as the social capital related to 'likes', 'comments', and 'shares'. Hunt et al. (2011) outline the role of society and culture on perceptions of beauty, which in turn influences consumption behaviours, particularly within the cosmetics industry, to try to meet current perceptions of beauty. What this means is that the young woman in our example is now internalising these perceptions, because this is what she is seeing as the norm on social media, and therefore believes that she must now meet these standards. The way social media's targeted advertising then influences this young woman is by flooding her news feed, in other words, her attention, with advertisements relating to cosmetic products which she can then use to meet these same beauty standards that she has been shown as being the norm and desirable. This young woman's consumption behaviours have been heavily influenced by social media on multiple levels. First, her perceptions of what constitutes beauty is shaped by algorithms and the content she is shown, and secondly, her decision to buy cosmetic products to attain these standards has been driven by targeted advertising which targets her vulnerabilities. Her actions do not look very autonomous, since she has been caused to acquire an ideal of beauty as a standard that she must meet, even though if she were aware of the powerful marketing and algorithmic forces behind her coming to hold this ideal, she would not

endorse that ideal or take responsibility for it upon critical reflection. Further, when she fails to meet these (largely) impossible standards, her self-esteem and worth as a person will likely diminish, impacting her autonomy competencies.

There are several important impacts on autonomy worth drawing out further here. What the user of social media is engaging with, and the order in which they are engaging with it, is determined by something outside of their control. These algorithms can disseminate content that can negatively impact the autonomy competencies of self-respect, self-love, and self-esteem. For example, personalised algorithms or the amplification of content can result in "'mismatched expectations' surrounding who individuals believe themselves to be" which can lead "to embarrassment and anxiety" (Smith, 2020, p. 69), which can clearly impact negatively on self-esteem, self-love, and self-respect. Assessing our findings through the lens of the three aspects of relational autonomy (Mackenzie, 2014a, 2014b) provides further insights. As Smith (2020) notes, for self-determination to be achieved "there must be a legitimate variety of life-choices available for individuals to pursue, free from dominating forms of power and interference" (p. 75). Within the context of social media this means that there must be a diversity of content that users can genuinely engage with which reflects, and sometimes challenges, their interests and desires. However, as is clear in examples such as Twitter's algorithmic amplification or targeted advertising, this is often not the case as algorithms are amplifying radicalising content, fake news, and building echo chambers. When we consider the ways in which social media's control over attention can lead to radicalisation and the formation of false beliefs, we can see how the dimension of self-governance is undermined by this. It is undermined because the decisions, values, and identities which are based on such false beliefs do not authentically reflect the identity and values of the user, at least insofar as the user would not endorse or take responsibility for them if they were critically reflective and aware they were caused to hold them by social media recommendation algorithms designed to monopolise their attention. Similarly, the dimension of self-authorisation is also undermined since the user does not have authority over their values and choices to the extent that these have been caused by social media in ways they may not be aware of or endorse. We can see this in the way that, for example, the amplification of particular perceptions of beauty or the creation of anxiety by social media platforms can result in users no longer taking authority over (some of) their values and choices as these are shaped by alienating external forces.

These impacts operate at both the programmatic and episodic levels of autonomy. The clearest example of episodic autonomy being undermined by social media comes in the form of targeted advertising. By targeting users at vulnerable moments, as determined by the analysis of big data that social media companies have access to, users can be pushed to make decisions that are not ones that they would autonomously make when in less vulnerable states. More broadly, radicalisation, the formation of false beliefs, and the amplification of certain ideals can impact users not only at the level of individual choices, but also at the programmatic level in terms of the person they want to be and the sort of life they value living. For example, a social media user who is gradually radicalised through going from occasionally watching videos on gurus to compulsively watching videos on extreme conspiracy theories (Alfano et al., 2020), can have their practical identity thereby radically rewritten by

social media. This changes who they are as a person at a deep level, thereby impacting their global autonomy negatively (at least insofar as they would not endorse or take responsibility for these changes were they critically reflective and aware of the manipulative algorithmic process that led to those changes). Likewise, evidence that social media usage can lead to (or be associated with) increased depression and anxiety, confirmed by leaked documents from Facebook (Gayle, 2021), also highlights the impact that social media sites can have on global autonomy, given the global negative impacts that depression and anxiety have on a person's agential powers that underwrite their autonomy.

Drawing this together, users of social media can start to lose control over what they believe and pay attention to. They can be pushed toward various political extremes without knowing it, be exposed to false views which they may incorrectly recollect as having reliable sources, be influenced by autonomy-undermining social norms, have their vulnerabilities exploited to sell products, and have their autonomy competencies eroded and harmed at both episodic and programmatic levels.

6 Control over Behaviour

This section explores the mechanisms which allow social media companies to control (intentionally or unintentionally) the behaviours of users to fulfil the ends of others, be it advertisers, political groups, or the social media companies themselves. The focus of this section will be on manipulation and addiction to illustrate the ways in which social media can control the behaviours of users and thereby negatively impact their autonomy.

The literature on manipulation is extensive (e.g. Klenk & Hancock, 2019; McCornack, 1992; Rudinow, 1978; Susser et al., 2019; Susser et al., 2019; Terrenghi et al., 2007; Van Dijk, 2006; Zarsky, 2019). Susser et al. (2019, p. 3) argue that "at its core, manipulation is hidden influence – the covert subversion of another person's decision-making power". This can be done by "exploiting the manipulee's cognitive (or affective) weaknesses and vulnerabilities in order to steer his or her decision-making process towards the manipulator's ends" (Susser et al., 2019, p.3). Manipulation can take place on a rational and deliberative level, through influencing beliefs, desires, values, and critical thinking, and on an affective or emotional level, by exploiting emotions such as fear or disgust, to control the behaviours of others. Manipulation of these two processes consequently impacts the ways in which people make decisions, and thus changes the behaviour of users towards the ends of the agent doing the manipulating. The final important point is that manipulation exploits the vulnerabilities of agents through hidden or covert means. A manipulated agent is controlled by others and is thus clearly not acting autonomously.

However, we first need to differentiate manipulation from other forms of influence. Two key forms of influence are persuasion and coercion. Persuasion is an overt appeal to someone to either do or abstain from something by influencing "their capacity for conscious deliberation and choice" (Susser et al., 2019, p. 14). Coercion, such as the threat of "your money or your life", is also an overt influence on another. However, coercion is done via the withdrawal of options, leaving

the coercers choice the only (reasonable) one remaining for the decision maker. In both methods of influence, the decision-making abilities of the agent are not undermined since they are overt forms of influence. These forms of influence differ from manipulation in that "to manipulate people is to displace them as the decider" and to "undermine or disrupt the ways of choosing that they themselves would critically endorse if they considered the matter in a way that is lucid and free of error" (Susser, et al., 2019, p. 16). Manipulation can occur through *deceiving* ("causing them to have false beliefs"), *tempting* ("creating a desire for what they lack reason to want"), and *inciting* ("causing an inappropriate emotional response") (Susser, et al., 2019, p. 19).

We can see each of these three dimensions of manipulation at work within social media. The deception of users is achieved through fake news or misleading advertising, causing users to form false beliefs which impact their decision making. The tempting of users is achieved through the perpetuation of often unreasonable standards of success, wealth, and beauty. By propagating certain representations of what it means to be successful, wealthy, or beautiful through algorithms, as well as mechanisms such as 'likes', 'shares', and 'comments', social media can create certain inauthentic desires (since users would not on critical reflection endorse or take responsibility for them) that nonetheless impacts on their decision making. For example, Reaves et al. (2004) report that the "exposure to the thin ideal" which is the result of conscious editing and "cosmetic retouching" of photos manipulates women as it "tends to reduce body satisfaction, increase self-consciousness, and reduce self-esteem" (p. 58). This means that women who are continually exposed to such (unrealistic and often unhealthy) ideals feel reduced self-esteem, which is an important autonomy competency, as they do not see themselves as fitting those perceived beauty standards. In this way a new desire has been formed within the women exposed, one that they have no reason which is authentically their own for wanting outside of the manipulation caused by the perpetuated beauty standards. While this is not a new phenomenon, as has been shown through Benson's (1991) pre-social media work on oppressive socialisation, social media exacerbates an existing problem due to the reach it can accomplish as well as its personalised and targeted nature. Finally, by exposing users to content that is designed to induce extreme emotional responses, social media can emotionally manipulate users into changing their behaviours. For example, Benkler et al. (2018) reports that "households earning less than $40,000 were most heavily targeted with advertisements focusing on immigration and racial conflict" (p. 274). By targeting specifically lower income individuals, advertisers can "elicit fear and loathing or to intimidate voters from turning out" (Benkler et al., 2018, p. 274). This is a clear mechanism within social media which can control the behaviours of users through emotional manipulation to achieve, not the authentic ends of some users (i.e., ends that they would endorse or take responsibility for after informed critical reflection), but the outcomes desired by advertisers. However, it is possible that at least some of the people emotionally manipulated by social media may in fact be manipulated to act in ways that they *would* have acted in any case (e.g. those who *already* held certain negative views about immigration and racial conflict before exposure to social media). But insofar as those agents are being caused to act in that way due to emotional manipulation that they are unaware of, as

opposed to being caused to act that way due to their considered convictions, we can still raise concerns about the *exercise* of their episodic autonomy, since it is emotional manipulation and not an authentic exercise of their agency that determines their behaviour.

One study shows how Facebook causes emotional manipulation. Kramer et al., (2014) report that emotional contagion occurs by manipulating the content on Facebook users' News Feed. They report that "when positive expressions were reduced, people produced fewer positive posts and more negative posts; when negative expressions were reduced, the opposite pattern occurred" (p. 8788). They conclude that the emotions that Facebook users are exposed to on their News Feed influences their own emotions (Kramer et al., 2014, p. 8788). The significance of this study comes to light when we consider the ability of social media companies to target advertising to users. For example, investigations by journalists found that Facebook advised advertisers on how they can target vulnerable teenagers, as young as fourteen, at moments when they feel "worthless" and "insecure". This is achieved by "monitoring posts, pictures, interactions and internet activity in real-time", which allows Facebook to know when teenagers "feel 'stressed', 'defeated', 'overwhelmed', 'anxious', 'nervous', 'stupid', 'silly', 'useless', and a 'failure'" (Susser et al., 2019, pp. 2–6). In this way, social media companies not only control the behaviours of users by actively manipulating their emotional responses, but then covertly, outside the awareness of users, exploit these vulnerable emotional responses to manipulate users into buying the products of their customers. This also fits with existing research around "nudge" theory, although here the nudging is not towards ends that are in the users' best interests, such as the best retirement plan (Thaler et al., 2012), but rather towards ends that are not the users' own at all, thereby raising clear concerns about a lack of user autonomy.

Moving on to examples of behavioural manipulation, a study by Bond et al., (2012) found that "messages directly influenced political self-expression, information seeking and real-world voting behaviour" in 61 million Facebook users during the 2010 US congressional elections (p. 295). The reason why this example goes beyond mere influence and constitutes manipulation is due to its covert nature, where the messaging is outside of the awareness of users, and it involves the exploitation of users' deliberative functions, specifically, deliberation regarding their civic duty pertaining to political decision making. They further report that the influence went beyond the users who directly received messaging, influencing "users' friends, and friends of friends" (Bond et al., 2012, p. 295). These findings again show the ways in which social media can manipulate users to change their behaviours and pursue ends that are not authentically those of the user.

Addiction is another way in which social media can control the behaviour of users. There are numerous studies throughout the literature which find that social media addiction is a global phenomenon which is widespread in both young people and adults (see e.g. Guillot et al., 2016; Kross et al., 2013; Primack et al., 2017; Rezaee & Pedret, 2018; Shakya & Christakis, 2017; Hou et al., 2019; Kuss & Griffiths, 2017; Bhargava & Velasquez, 2021). Social media companies utilise three mechanisms to fuel social media addiction (Bhargava & Velasquez, 2021). The first is the use of intermittent variable rewards which is a design feature found within

the loading of the home screen and in the "pull-to-refresh" function of social media platforms such as Twitter, both of which are designed to produce intermittent variable rewards within the user, similar to that of a slot machine (Bhargava & Velasquez, 2021, p. 327). Second, social media platforms are designed to exploit the "desire for social validation and social reciprocity" through "social reward schemes" such as "like" buttons or the "share" functionality (Bhargava & Velasquez, 2021, pp. 326–327). Third, social media platforms have taken away any natural stopping cues by using the 'infinite scroll', depriving users of any stopping cues which may give them the opportunity to make the decision to step away from the platform for a time (Bhargava & Velasquez, 2021, pp. 326–327). Further, Kuss and Griffiths, (2017) report that "higher levels of FOMO [Fear of Missing Out] have been associated with greater engagement with Facebook" and that FOMO is "associated with social media addiction" (p. 8). This leads to the next significant component of social media which works to fuel addiction and control the behaviours of users, the use of pop-up notifications, which functions on the same principle of intermittent variable rewards outlined above (Bhargava & Velasquez, 2021). Social media has not only been able to utilise the notification system to induce FOMO, as users feel that they must attend to every notification as soon as it pops up, but the notification system has become a way of attracting the attention of users away from other activities back towards the social media platform.

When social media platforms covertly manipulate the beliefs, emotions, and decision-making processes of users by exploiting their vulnerabilities, then this expresses a clear disrespect towards users' right to make their own autonomous choices. Social media also manipulates the emotions of users in ways which impact their autonomy-constituting evaluative attitudes. The example we looked at of targeting advertising at young teenagers when they are feeling vulnerable highlights how social media companies can exploit the vulnerability of users' autonomy competencies, and associated evaluative attitudes of self-respect, self-love, and self-esteem, to shape their behaviours and choices. Further, social media manipulation negatively impacts all three dimensions of autonomy. Self-determination is undermined by the fact that the manipulation is taking place for the sole reason to direct users to make choices which suit the ends of the social media companies and their customers, not their users. Therefore, the user is not making choices which reflect control over their life at that time. This also highlights the fact that episodic autonomy is clearly impacted, since during times of poor emotional well-being, social media companies can exploit and manipulate users. Self-governance is impacted by the fact that the manipulation on social media is done covertly and can occur outside of the awareness of the user, hindering their ability to critically reflect on the choices they are making. Finally, self-authorisation is completely undermined when users are manipulated on social media since the user lacks authority over their identity, values, and choices when their decision-making faculties have been hijacked and manipulated through the covert exploitation of their vulnerabilities. Furthermore, global autonomy, concerning one's overall life, is most clearly impacted by the addiction which is fuelled by social media through the implementation of mechanisms which lead to social media addiction, as well as a constant fear of missing crucial and significant events or content, and the need to always be online and available. In these ways,

users' overall life is being impacted and shaped by social media and its behavioural controls.

7 Discussion and Recommendations

Social media can help us to achieve our ends and exercise our autonomy by facilitating social interactions with others that we authentically value. But there are also significant dangers to autonomy posed by social media. These dangers and costs need to be weighed against other benefits, such as economic gains, that social media brings. However, our focus here has been exclusively on the harms to human autonomy that can occur through social media in three ways: (1) *disrespecting* users' autonomy; (2) interfering with the *exercise* of users' autonomy at episodic and global levels; and (3) lowering users' degree of *autonomy competencies*. Some of these harms are caused by other users of social media (such as bullying), but many rely on systemic features of social media platforms. These three types of autonomy harms clearly lead to discussions around social media regulation, which has recently become an important political topic given both the importance of autonomy and the influence of social media on it. While many of these broader regulatory discussions go well beyond our focus on social media's impacts on autonomy, our more detailed focus has something to contribute to this wider debate.

However, in the below discussion, it is important to keep in mind that there are dangers to both too much regulation, which can stifle innovation and the various benefits of new technologies, as well as under regulation, which can lead to harm and a failure to realise the benefits of new technology due to concerns about its safety (Petit, 2017). Further, proposed and actual regulatory and legal frameworks around social media differ from country to country, such as Europe's General Data Protection Regulation (GDPR) (see e.g. Vese, 2021), and cover a range of issues, such as freedom of expression (Nelson et al., 2021; Vese, 2021), that reach beyond our focus. Our discussion here thus remains by necessity at a high-level of abstraction to retain the relevant generality and focus on autonomy harms only. In so doing, we consider recommendations for users, social media platforms, and regulators.

Many of the existing recommendations around social media regulation have some relevance for our discussion here. This includes the following: requirements for reliability ratings to combat fake news and conspiracy theories (Vese, 2021); attempts to limit doom scrolling and auto-play of video and other content (Sharma et al., 2022); changing business models away from advertising and thus a focus on the attention economy (Alfano et al., 2020); banning targeted advertising (Roth, 2022); preventing the spread of extremist and terrorist content on social media (West, 2021); limiting of state-run political surveillance of users (Obia, 2021); clarifying whether platforms or users have responsibility for user generated content (Obia, 2021); and the need for user education, digital literacy, and epistemic vigilance (Alfano et al., 2020).

In terms of end users, better education around social media and the cultivation of relevant digital virtues, such as epistemic vigilance, will be important in limiting some of the negative autonomy impacts outlined here. These virtues can be

cultivated through "ongoing, conscious and intentional" (Alfano et al., 2020, p. 20) monitoring of, and self-education about, the influences of social media on one's beliefs, values, and ends. This can be aided by using various means to limit that influence further, such as time limiting access or taking a 1-week break from social media (which can improve well-being, depression, and anxiety (Lambert et al., 2022)), using ad-blockers and preventing tracking, and accessing a broad range of credible new sources outside of social media. Users also need to consider the way that their actions on a social media platform, such as liking or sharing certain content, influences the autonomy of other users of the platforms in potentially negative ways. However, given the pressures to use social media and the network effects this can have (Srnicek, 2017), individual efforts are not a complete solution. A better understanding of how social media impacts upon the development of autonomy competencies, its impacts on our self-attitudes (beyond self-esteem (e.g. Lambert et al., 2022), which has been widely studied), the veracity of our beliefs, the way we reason and think, what we value and care about, who we interact with and are influenced by, and the choices we make is therefore crucial. Many of these require, to be effective, regulations of the form proposed above, such as a move away from platform features that encourage a focus on monopolising users' attention, including personalised advertising and continuous streams of content. In assessing these proposed regulations, it is important to consider the extent to which exploitative and manipulative practices can disrespect, interfere with, and hinder users' autonomy and related autonomy competencies. For example, regulatory moves to stop continuous feeds of content or auto-play of videos will help users to autonomously choose what content they wish to engage with and limit the ability of social media platforms to exploit users' FOMO to monopolise their limited attention spans for advertising purposes. A similar autonomy-focused analysis could be supplied for many of the various regulatory proposals noted above. Limiting autonomy harms is therefore an effective normative framework to justify and analyse many of the proposed regulations of social media platforms outlined above, even if it does not cover all such proposals.

8 Conclusion

By drawing on the philosophical literature on autonomy, we have provided a more comprehensive analysis of the various negative impacts of social media on the autonomy and autonomy competencies of its users. This led us to focus on the *control* that social media can have over users' *data*, *attention*, and *behaviour*, which can result in disrespect of users' capacity for autonomy, interference with exercises of their autonomy, and harms to their autonomy competencies. Some of these harms involve local episodic impacts on autonomy, whereas others, such as social media addiction, can have a much more pervasive and global impact. Finally, we considered briefly various recommendations to better regulate social media and demonstrated how a focus on autonomy harms can help to justify and make sense of some of these proposals.

Author contribution Both SS and PF contributed to all aspects of the preparation and writing of this manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Data Availability Not applicable.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing interests SS declares they have no financial interests. PF declares they have previously received funding for another unrelated project from Facebook.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Alfano, M., Fard, A. E., Carter, J. A., Clutton, P., & Klein, C. (2020). Technologically scaffolded atypical cognition. *Synthese*. https://doi.org/10.1007/s11229-020-02724-x

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 Election. *Journal of Economic Perspectives, 31*(2), 211–236.

Andreassen, C. S., Pallesen, S., & Griffiths, M. D. (2017). The relationship between addictive use of social media, narcissism, and self-esteem. *Addictive Behaviors, 64*, 287–293.

Barocas, S., & Nissenbaum, H. (2014). Big data's end run around anonymity and consent. *Privacy, Big Data, and the Public Good: Frameworks for Engagement, 1*, 44–75.

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in american politics*. Oxford University Press.

Benson, P. (1991). Autonomy and oppressive socialization. *Social Theory and Practice, 17*(3), 385–408.

Bhargava, V. R., & Velasquez, M. (2021). Ethics of the attention economy. *Business Ethics Quarterly, 31*(3), 321–359.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature, 489*(7415), 295–298.

Boyd, D. (2010). Streams of content, limited attention. *Educause Review, 45*(5), 26.

Christman, J. (1991). Autonomy and personal history. *Canadian Journal of Philosophy, 21*(1), 1–24.

Cingel, D. P., Carter, M. C., & Krause, H. V. (2022). Social media and self-esteem. *Current Opinion in Psychology*, 101304.

Citton, Y. (2017). *The ecology of attention*. Polity Press.

Cramer-Flood, E. (2021, April 29). Worldwide digital ad spending 2021. *EMarketer*. https://www.emarketer.com/content/worldwide-digital-ad-spending-2021

Cohen, J. N. (2018). Exploring Echo-Systems. *The Journal of Media Literacy Education, 10*(2), 139–151.

Comunello, F., & Anzera, G. (2012). Will the revolution be tweeted? A conceptual framework for understanding the social media and the Arab Spring. *Islam and Christian-Muslim Relations, 23*(4), 453–470.
Darwall, S. (2006). The value of autonomy and autonomy of the will. *Ethics, 116*(2), 263–284.
Daugherty, T., & Hoffman, E. (2014). eWOM and the importance of capturing consumer attention within social media. *Journal of Marketing Communications, 20*(1–2), 82–102.
Duffy, B. E., & Chan, N. K. (2019). "You never really know who's looking": Imagined surveillance across social media platforms. *New Media & Society, 21*(1), 119–138.
Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge University Press.
Fassihi, F. (Dec 3, 2009). Iranian crackdown goes global. *Wall Street Journal*. https://www.wsj.com/articles/SB125978649644673331
Feng, L., Hu, Y., Li, B., Stanley, H. E., Havlin, S., & Braunstein, L. A. (2015). Competing for attention in social media under information overload conditions. *PLoS One, 10*(7), e0126090.
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines, 28*(4), 689–707.
Formosa, P. (2011). From discipline to autonomy: Kant's Theory of Moral Development. *Kant and Education* (pp. 163–176). Routledge.
Formosa, P. (2017). *Kantian ethics, dignity and perfection*. Cambridge University Press.
Formosa, P. (2013). Kant's conception of personal autonomy. *Journal of Social Philosophy*, *44*(3), 193–212.
Formosa, P. (2021). Robot autonomy vs. human autonomy. *Minds and Machines, 31,* 595–616. https://doi.org/10.1007/s11023-021-09579-2
Frangonikolopoulos, C. A., & Chapsos, I. (2012). Explaining the role and the impact of the social media in the Arab Spring. *Global Media Journal*, *7*(2).
Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, *68*(1), 5–20.
Friedman, M. A. (1986). Autonomy and the split-level self. *The Southern Journal of Philosophy, 24*(1), 19–35.
Fuchs, C. (2012). Dallas Smythe today - The audience commodity, the digital labour debate, marxist political economy and critical theory. *TripleC, 10*(2), 692–740.
Fuchs, C. (2011). Web 2.0, Prosumption, and surveillance. *Surveillance & Society*, *8*(3), 288–309.
Fuchs, C. (2017). *Social media: A critical introduction*. Sage.
Gayle, D. (2021). Facebook aware of Instagram's harmful effect on teenage girls, leak reveals. *The Guardian.* https://www.theguardian.com/technology/2021/sep/14/facebook-aware-instagram-harmful-effect-teenage-girls-leak-reveals
Guillot, C. R., Bello, M. S., Tsai, J. Y., Huh, J., Leventhal, A. M., & Sussman, S. (2016). Longitudinal associations between anhedonia and internet-related addictive behaviors in emerging adults. *Computers in Human Behavior, 62*, 475–479.
Ha, L., Andreu Perez, L., & Ray, R. (2021). Mapping recent development in scholarship on fake news and misinformation, 2008 to 2017. *The American Behavioral Scientist, 65*(2), 290–315.
Hawi, N. S., & Samaha, M. (2017). The relations among social media addiction, self-esteem, and life satisfaction in university students. *Social Science Computer Review, 35*(5), 576–586.
Hou, Y., Xiong, D., Jiang, T., Song, L., & Wang, Q. (2019). Social media addiction. *Cyberpsychology*, *13*(1), Article 4.
Howard, P. N., Duffy, A., Freelon, D., Hussain, M. M., Mari, W., & Maziad, M. (2011). Opening closed regimes. *Available at SSRN 2595096*
Hunt, K. A., Fate, J., & Dodds, B. (2011). Cultural and social influences on the perception of beauty. *Journal of Business Case Studies, 7*(1), 1–10.
Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2021). Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences*, *119*(1).
Jones, M. O. (2017). Social media, surveillance and social control in the Bahrain uprising. *Westminster Papers in Communication and Culture*, *9*(2), 69-92.
Killmister, S. (2013a). Autonomy and false beliefs. *Philosophical Studies, 164*(2), 513–531.
Killmister, S. (2013b). Autonomy and the problem of socialization. *Social Theory and Practice, 39*(1), 95–119.
Klenk, M., & Hancock, J. (2019). Autonomy and online manipulation. *Internet Policy Review*, *1*.
Korsgaard, C. M. (1996). *Creating the kingdom of Ends*. Cambridge University Press.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, 111*(24), 8788–8790.

Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D. S., Lin, N., & Ybarra, O. (2013). Facebook use predicts declines in subjective well-being in young adults. *PloS One, 8*(8), e69841–e69841.

Kuss, D. J., & Griffiths, M. D. (2017). Social networking sites and addiction. *International Journal of Environmental Research and Public Health, 14*(3), 311.

Lambert, J., Barnstable, G., Minter, E., Cooper, J., & McEwan, D. (2022). Taking a one-week break from social media improves well-being, depression, and anxiety. *Cyberpsychology, Behavior, and Social Networking, 25*(5), 287–293.

Lavazza, A., & Reichlin, M. (2018). Of meatballs, autonomy, and human dignity. *AJOB Neuroscience, 9*(2), 88–95.

Levy, N. (2017). The bad news about fake news. *Social Epistemology Review and Reply Collective, 6*(8), 20–36.

Mac, R., & Frenkel, S. (2021). Internal alarm, public shrugs. *The New York Times.* https://www.nytimes.com/2021/10/22/technology/facebook-election-misinformation.html

Mackenzie, C. (2014a). The importance of relational autonomy and capabilities for an ethics of vulnerability. In C. Mackenzie, W. Rogers, & S. Dodds (Eds.), *Vulnerability* (pp. 33–59). Oxford University Press.

Mackenzie, C. (2014b). Three dimensions of autonomy. In A. Veltman & M. Piper (Eds.), *Autonomy, Oppression and Gender* (pp. 15–41). Oxford University Press.

Mackenzie, C., & Stoljar, N. (2000). Autonomy refigured. In C. Mackenzie & N. Stoljar (Eds.), *Relational Autonomy* (pp. 3–34). Oxford University Press.

Malesev, S., & Cherry, M. (2021). Digital and social media marketing-growing market share for construction SMEs. *Construction Economics and Building, 21*(1), 65–82.

McCornack, S. A. (1992). Information manipulation theory. *Communication Monographs, 59*(1), 1–16.

Matsa, K. E., & Shearer, E. (2018). News use across social media platforms 2018. *Pew Research Center*, *10.*

Meyers, D. T. (1987). Personal autonomy and the paradox of feminine socialization. *The Journal of Philosophy, 84*(11), 619–628.

Milmo, D., & Paul, K. (2021). Facebook harms children and is damaging democracy, claims whistleblower. *The Guardian.* https://www.theguardian.com/technology/2021/oct/05/facebook-harms-children-damaging-democracy-claims-whistleblower

Mittelstadt, B. (2016). Auditing for transparency in content personalization systems. *International Journal of Communication, 10*, 4991–5002.

Myllylahti, M. (2018). An attention economy trap? *Journal of Media Business Studies, 15*(4), 237–253.

Molitorisz, S. (2020). *Net Privacy*. NewSouth Publishing.

Nelson, M. R., Ham, C. D., Haley, E., & Chung, U. C. (2021). How political interest and gender influence persuasion knowledge, political information seeking, and support for regulation of political advertising in social media. *Journal of Interactive Advertising, 21*(3), 225–242.

Nissenbaum, H. (2011). A contextual approach to privacy online. *Daedalus, 140*(4), 32–48.

Obia, V. A. (2021). Are social media users publishers? *Makings, 2*(1), 1–13.

O'Neill, O. (2002). *Autonomy and trust in bioethics*. Cambridge University Press.

Paul, K., & Milmo, D. (2021). Facebook putting profit before public good, says whistleblower Frances Haugen. *The Guardian.* https://www.theguardian.com/technology/2021/oct/03/former-facebook-employee-frances-haugen-identifies-herself-as-whistleblower

Petit, N. (2017). Law and regulation of artificial intelligence and robots—Conceptual Framework and normative implications. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2931339

Primack, B. A., Shensa, A., Sidani, J. E., Whaite, E. O., Lin, L. yi, Rosen, D., … Miller, E. (2017). Social media use and perceived social isolation among young adults in the U.S. *American Journal of Preventive Medicine*, *53*(1), 1–8.

Raz, J. (1986). *The morality of freedom*. Clarendon Press.

Reaves, S., Bush Hitchon, J., Park, S.-Y., & Woong Yun, G. (2004). If looks could kill. *Journal of Mass Media Ethics, 19*(1), 56–71.

Reiman, J. H. (1995). Driving to the panopticon. *Santa Clara Computer and High-Technology Law Journal, 11*(1), 27.

Rezaee, F., & Pedret, A. (2018). A critical review of silicon valley solutions for smartphone addiction. *Idealogy Journal of Arts and Social Science, 3*(3), 189–196.

Rheingold, H. (2010). Attention, and other 21st-century social media literacies. *Educause Review, 45*(5), 14–24.
Rogers, W., Mackenzie, C., & Dodds, S. (2012). Why bioethics needs a concept of vulnerability. *IJFAB: International Journal of Feminist Approaches to Bioethics*, *5* (2) 11–38
Roth, E. (2022). European Parliament approves initial proposal to ban some targeted ads. *The Verge*. https://www.theverge.com/2022/1/23/22897574/european-parliament-eu-digital-services-act-big-tech
Rudinow, J. (1978). Manipulation. *Ethics, 88*(4), 338–347.
Schneewind, J. B. (1998). *The invention of autonomy*. Cambridge University Press.
Shakya, H. B., & Christakis, N. A. (2017). Association of Facebook use with Compromised well-being. *American Journal of Epidemiology, 185*(3), 203–211.
Sharma, B., Lee, S. S., & Johnson, B. K. (2022). The dark at the end of the tunnel. *Technology, Mind, and Behavior.* https://doi.org/10.1037/tmb0000059
Smith, C. H. (2020). Corporatised identities ≠ digital identities. In C. Burr & L. Floridi (Eds.), *Ethics of Digital Well-Being* (pp. 55–80). Springer.
Srnicek, N. (2017). *Platform capitalism*. Polity Press.
Steinsbekk, S., Wichstrøm, L., Stenseng, F., Nesi, J., Hygen, B. W., & Skalická, V. (2021). The impact of social media use on appearance self-esteem from childhood to adolescence. *Computers in Human Behavior, 114*, 106528.
Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet Policy Review*, *8*(2).
Terrenghi, L., Kirk, D., Sellen, A., & Izadi, S. (2007). Affordances for manipulation of physical versus digital media on interactive surfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1157–1166.
Thaler, R. H., Sunstein, C., & Balz, J. P. (2012). Choice architecture. In E. Shafir (Ed.), *The Behavioral Foundations of Public Policy*. Princeton University Press.
Shao, C. (2020). The surveillance experience of Chinese university students and the value of privacy in the surveillance society (Doctoral dissertation, The University of North Carolina). https://doi.org/10.17615/tr7a-yr90
Valdman, M. (2009). A theory of wrongful exploitation. *Philosophers' Imprint, 9*(6), 1–14.
Vallor, S. (2015). Moral deskilling and upskilling in a new machine age. *Philosophy & Technology, 28*(1), 107–124.
van Dijk, T. A. (2006). Discourse and manipulation. *Discourse & Society, 17*(3), 359–383.
Vese, D. (2021). Governing fake news. *European Journal of Risk Regulation*, 1–4. https://doi.org/10.1017/err.2021.48
Walker, M., & Mackenzie, C. (2020). Neurotechnologies, relational autonomy, and authenticity. *IJFAB, 13*(1), 98–119.
Watson, G. (1975). Free agency. *The Journal of Philosophy*, *72*, 205–220.
West, L. J. (2021). Counter-terrorism, social media and the regulation of extremist content. In S. Miller, A. Henschke, & J. Feltes (Eds.), *Counter-Terrorism* (pp. 116–128). Edward Elgar Publishing.
Winpenny, E. M., Marteau, T. M., & Nolte, E. (2014). Exposure of children and adolescents to alcohol marketing on social media websites. *Alcohol and Alcoholism, 49*(2), 154–159.
Wood, A. W. (1995). Exploitation. *Social Philosophy and Policy, 12*(2), 136–158.
Woods, H. C., & Scott, H. (2016). # Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *Journal of Adolescence, 51*, 41–49.
Worden, K. J. (2019). Disengagement in the digital age. *Moral Philosophy and Politics, 6*(2), 235–259.
Zadrozny, B. (2021). 'Carol's journey': What Facebook knew about how it radicalized users. *NBC News.* https://www.nbcnews.com/tech/tech-news/facebook-knew-radicalized-users-rcna3581
Zarsky, T. Z. (2019). Privacy and manipulation in the digital age. *Theoretical Inquiries in Law, 20*(1), 157–188.
Zuboff, S. (2019). *The age of surveillance capitalism*. Profile Books.
Zulli, D. (2018). Capitalizing on the look. *Critical Studies in Media Communication, 35*(2), 137–150.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.