Becoming oneself online

Social media platforms as environments for self-transformation

Marin, Lavinia

DOI
10.1057/s41599-025-04495-7

Publication date
2025

Document Version
Final published version

Published in
Humanities and Social Sciences Communications

Citation (APA)
Marin, L. (2025). Becoming oneself online: Social media platforms as environments for self-transformation. *Humanities and Social Sciences Communications*, *12*(1), Article 219. https://doi.org/10.1057/s41599-025-04495-7

Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-025-04495-7 OPEN

Becoming oneself online: Social media platforms as environments for self-transformation

Lavinia Marin [1]✉

This paper examines social media platforms as spaces fostering their user's self-transformation. This paper argues that the ethics of (illegitimate) technological influence can be expanded and enriched with a concept of situated agency and an enactive evaluation of adaptability afforded by an environment. The paper proposes a taxonomy to evaluate social media platforms as environments for self-transformation by using the concept of situated agency and a notion of enactive normativity. Using a situated concept of agency, we should look into how an agent is afforded or even pushed to undergo transformative experiences while inhabiting a certain technical environment. To coin this taxonomy, the types of transformative experiences from Carel and Kid were used to show that one agent can effectively have four modes of formative agency in any given situation, namely how one responds to the pressures of an environment. However, since it could be argued that any kind of adaptation is legitimate since we become who we change into, this taxonomy adds an explicit normative dimension pertaining to the environment that fosters the agent's transformation of the self: borrowing from the terminology of enactive theories of cognition, how an environment responds to an agent's adaptation is also important and can make the difference between a hostile environment and a flourishing-conducive one. The paper shows how one can apply this taxonomy of formative agency to evaluate some of the most concerning cases of self-transformation triggered by social media platforms: online self-radicalisation, habit acquisition, and identity rigidification.

[1] Technical University Delft, Delft, the Netherlands. ✉email: l.marin@tudelft.nl

Introduction: The problem of user self-transformation in a digital environment

The problem of illegitimate influence over users of digital platforms has emerged in recent years as a topic of concern for philosophers and ethicists of technology. Illegitimate influence occurs when a user interacts with a technical artefact (an online platform, a smartphone, any algorithmically shaped interface) and when, because of that interaction, the user's opinions, emotions or actions are steered in a certain direction without their noticing or agreeing to this steering (Specker Sullivan and Reiner 2021). Technological influence of users occurs on a spectrum of acceptability, from completely forbidden processes, such as in external interference with elections and voting behaviour (Digital Services Act 2024), to illegitimate cases in which a technological artefact unconsciously shapes a user's values and behaviours, to socially acceptable influences such as in the case of advertising messages in the public spaces or educational interventions like nudging or gamification. For a process of technological influence to count as illegitimate, it should either undermine the user's own interests (hence undermining their autonomy) or be simply indifferent to their interests while pursuing the influencer's own agenda (Klenk 2022). There are many ethically problematic aspects with illegitimate technological influence, such as subverting the user's autonomy, agency and capacity for reasoning (Specker Sullivan and Reiner 2021; Klenk 2022) or their long-term well-being.

In the ethics of technology concerned with illegitimate influence, the ongoing quest has been to find the appropriate normative framework that would allow us to evaluate the legitimate and illegitimate effects of a digital interface over a human user. For example, Laura Specker Sullivan and Peter Reiner have recently proposed a framework to evaluate an online platform's persuasive influence in categories such as nudging, manipulation, deceit, paternalism or maternalism (Specker Sullivan and Reiner 2021). This framework and similar ones focusing on online manipulation (Klenk 2022) or persuasion (Mitchell and Douglas 2024) are well-suited to assess the short-term influence of a platform over a user's beliefs, actions or desires. For something to count as a successful persuasive influence, it needs to afford a clear link to an attempt to influence, and it has to be intentional on the side of the influencer. If I want you to buy product X and I exert all kinds of influence strategies, the success of my attempts will be measured by your buying the product X or expressing the intent to buy X soon. Even when the influence is deemed careless (as in Michael Klenk's account of manipulation as a careless influence – see Klenk, (2022), there is still intent on the part of the influencers for the target to do something or believe something. Yet how should one normatively assess the technological influence when this is not coming from one single actor, but from an environment that the technology user inhabits over a longer time?

Most cases of illegitimate technological influence found thus far in the philosophical scholarship are discernible through their short-term effects: users are influenced either to perform an action, feel an emotion, or change a belief. If the change happens without their own intent, we are dealing with a successful attempt at illegitimate technological influence. However, some types of technological influence are more difficult to evaluate in terms of legitimacy, namely those influences taking place slowly over a longer period of time and when technology constitutes an environment for the user. There is, however a long-term, more insidious indirect influence of technology use that such frameworks cannot capture because one usually assumes intent from the platform side. Most of us live in technologically shaped environments, which fundamentally shape how we experience the world around us: from smart cities to smart homes, to wearable devices, to platforms we access to socialise or consume content: social networking platforms, entertainment, news and streaming services. These apps, platforms and devices constitute a technological ecosystem we seamlessly inhabit to such an extent that we cannot separate the technological from the "natural" in our lives anymore, as our lives are increasingly happening in the online realm (Floridi 2015, p. 2). Considering the existing concerns with illegitimate technological influence, two tricky issues emerge when this influence is experienced as environmental: first, all environments influence their inhabitants; simply put, it is impossible to live in an environment habitually and not be influenced by it in some way, and, secondly, that environmental influences happen in the long run, through minute changes, thus making the normative evaluation of these influences difficult to operationalise.

The transformative potential that a technological environment has for its users has not been researched thus far in the ethics of technology, albeit there have been recent calls to research this phenomenon in more depth from other disciplines, such as communications sciences (Walther & Lew, 2022). The scholarship on ethics of (digital) technologies has mainly focused on the more spectacular cases of change or those that are discernible over a short period. Except for those researching neurotechnologies (Boer et al. 2021) or self-tracking apps (Wieczorek et al. 2022), philosophers and ethicists of technology have yet to examine how the long-term use of technology transforms users. Self-transformation is a continuous process we all undergo, presumably never-ending, usually experienced through minute changes that are difficult to trace (Westley et al. 2013). For a definition of self-transformations in a digital environment, we can turn to a recent definition of this phenomenon as "a change in individuals' self-concepts, personalities, or specific attitudes and behaviors pertaining to individuals' own attributions about themselves" (Walther and Lew 2022, p. 135). This means that the personality or character traits of a technology user need not actually change, if their perception of the self changes, this should count as a self-transformation. How should we assess such a technologically induced transformation from an ethical perspective? Could we embrace who we have become because the environment pushed us into a certain kind of transformation with the same resignation that we would embrace an accident? This is the ethical problem of environmentally triggered incremental self-transformation.

The self and the transformative environment. Some conceptual distinctions

Debates on what gives identity to a human self have a long history in philosophy (Tobia 2022), spanning almost millennia. In the last decades, technology has started to appear in these debates, for example, in theories of the extended self. Such theories hold that the self is extended through tools and artefacts (Heersmink, 2020), implying that "the boundaries of selves are fluid, shifting across biological, artifactual, and sociocultural structures" (Heersmink, 2020, p. 10) or that the self can be seen as a networked set of traits, "a temporal, changeable network of accumulating traits" (Wallace 2019, p. 1), which can also be understood as an environmentally triggered accumulation of embodied experiences that tacitly "sediment" (Colombetti and Bogotá 2024, p. 1) like layers of sand on the bed of a river. While the ontology of the self varies in recent theories (from extended self to relational, networked selves, to embodied sedimentation), what is similar across these theories is the idea that the human self is not individually contained in one's body and brain; rather, the self gets shaped by the world which becomes an important

part of the self. Following this idea of the self as extended and distributed into the world, it makes little sense to say that the environments in which we find ourselves "influence" or shape us; rather, these environments constitute our selves. Extended notions of the self challenge the idea of an illegitimate influence from the outside of the self since the 'outside' becomes rather fuzzy and porous. Especially when technology is the environment in which the self unfolds and is constantly immersed, the notion of illegitimate influence on the self needs urgent reconceptualisation. We need a less dualistic notion of self *versus* the world to account for how the world shapes us and to be able to say when this interference is illegitimate.

Transformations of the self. What kinds of self-transformation can a person undergo in a lifetime? The distinction between transformative and regular changes of the self is useful here, as introduced by Laurie Ann Paul through the concept of "transformative experience" in current philosophical debates. Transformative experiences are those that a human undergoes and radically change who one is, specifically changes in one's priorities and values, for example, becoming a parent, changing professional paths or becoming a vampire (Paul 2014). The philosophical argument advanced by Laurie Ann Paul was that one could not rationally choose to undergo such a transformative experience since a rational decision is based on a weighing of costs and benefits based on one's priorities and values at the moment of decision. Assuming that the transformative experience will change the criteria that make such a choice rational for one person, the choice to undergo a transformative experience catapults oneself into becoming someone else and hence cannot be called rational or a-rational. Relevant to this paper is an argument that occurred in this still ongoing debate surrounding the transformative experience: Carel and Kid pointed out that most of the transformative experiences one undergoes are not happening by choice (2020, p. 205) rather, these experiences are imposed upon us by others intentionally or accidentally, hence we do not have that much agency in undergoing these. Experiences such as going to prison, fighting in a war, or surviving a traumatic event such as a stroke or an earthquake are just as transformative as deciding to become a parent. In addition to expanding the concept of transformative experience to account for the negative experiences that are beyond one's control, Carel and Kid also make room for incremental, small experiences that will sum up to a transformed self: transformative experiences "can also be mundane, unchosen, nonvoluntary or involuntary, and small" (Carel & Kidd, 2020, p. 210). Even daily experiences accumulated over time should count as transformative experiences if, after a while, we find ourselves significantly changed.

Based on Carel and Kid's observation that transformative experiences can happen incrementally and slowly over time (2020), this raises a novel problem about the normative evaluation of the environments that trigger such experiences. When an experience happens accidentally, such as a traumatic one, we have no choice but to accept it and live with the new self we have become. However, when a transformative experience happens because we immerse ourselves in an environment frequently, then we should share some of the responsibility for this transformation. Still, if the environment that triggers this transformation over time is designed and curated to trigger such changes, then some responsibility should be taken by those who enabled that environment. From the smart homes we live in to the smartphones that we use daily to the social media platforms that we spend our time on, our lives are immersed in technological environments that are designed by others to fulfil certain functions. Nevertheless, can such environments also transform who we are? The current technological environments should also be evaluated for their transformative potential as visible over a longer time.

Environmentally triggered transformations fall outside the normative frameworks for assessing illegitimate technological influences. By contrast to manipulation, persuasion or coercion, in environmentally triggered transformations over a long time, one cannot meaningfully attribute intent to the technological environment since multiple influences converge in a way that cannot be traced back to any single actor or group. I will designate such cases with the term *synergistic influence* to designate the cases in which a user is influenced to do or change something but without clear intent from another actor and any clear benefits for someone else. Such technological environments of synergistic influence are found whenever a user interacts with a complex system (a platform, a search engine, an algorithm) via a digital interface, usually over a longer period of time. In this paper, I will focus primarily on the synergistic influence of a technological environment over the user's self-transformation, aiming to provide some ways of normatively assessing when this influence is acceptable and when it becomes harmful or pathological.

Formative agency as a type of situated agency. This paper aims to arrive at some normative criteria to allow the distinction between the illegitimate and legitimate influences of a technological environment on the users' selves. The second aim is to apply this framework to the case of user self-transformation as mediated by social media platforms. Arriving at this normative framework to evaluate the effects of an environment on its inhabitants implies finding a way of assessing when a transformation of the self is pathological, when it is beneficial to the agent, and when it is neither of the two extremes. Any change of self leads to a new self, be it a transformative change experiences as a rupture, or an incremental accumulation of change over time. From the perspective of the individual undergoing the change of the self, we are who we become and there is no vantage point outside the agent from which to designate the new self as alien or authentic, desirable or undesirable. To say that a transformation of the self is pathological or conducive to flourishing, we need to keep the reference point grounded in the agent while also giving them the space to change in whatever way they seem fit.

I propose to take the concept of agency in one's self-formation as the crucial point that helps us flesh out when change is desirable or not. The main question with technological influence is how we can discern between environments that put pressure on us to transform yet maintain our agency in self-formation to some extent, and those that are depriving us of the agency in self-transformation. We need a concept of agency to account for these synergistically and environmentally triggered changes of the self. The concept of *situated agency* (Desmond and Huneman 2022) appears useful for the purposes of this normative analysis. Situated agency is somewhat different from the classical concept of agency, which would emphasise an individual's autonomy in deciding what to do, usually understood as independence from their environment. In a classical view of agency (Dung 2024, pp. 3-8), there are several conditions for someone to have agency: X has agency in regard to A if X autonomously decided to do A, if they achieved A efficaciously (they were able to trigger the change and were not reliant on another agent to initiate it), they aimed to do A with a goal in mind (intentionality and goal-directedness), and A was planned for (Dung 2024, pp. 3-8). Classical concepts of agency tend to focus on short-term actions: agent X wants to do action A, and then they perform it in a limited time window. By contrast, a situated account of agency will emphasise how much of the actions are shaped as a response to the environment in

which the agent finds themselves, namely how agents respond to environmental pressures and adapt in such a way that they still pursue their goals (Desmond and Huneman 2022, p. 1).

With situated agency, what matters is that agents are always constrained in their range of responses by the environment they inhabit, but their environment does not determine them, they are still agentic in how they respond to the environmental cues: "people might be situated agents rather than autonomous agents, but they are nonetheless agents. They can innovate, adopt novel beliefs, and perform novel actions against the background of contexts that influence them" (Bevir 2017, p. 53). Even if an environment shapes the actions of the agents acting in it, we are dealing with situated agency because these actions stem from the agent having their own goals. Furthermore, the style of adapting to environmental pressures is also personal to that agent. In a competition, one may cave under pressure to perform, while someone else may thrive; both competitors are agentic in coping with this pressure. However, when an environment predictably gives rise to similar reactions for a variety of agents, then we can say something about the environment itself. If the responses accumulate over time into changes that lead to self-transformation, we are dealing with an environment conducive to self-transformation reliably and predictably.

In order to evaluate SMPs as environments fostering user self-transformation, I propose to take situated agency as the core dimension and see then which types of agency are afforded by online environments. I will flesh out a more specific concept of situated agency, namely that of formative agency. The concept of formative agency captures the idea that we still have agency in how we transform ourselves, even when the changing experiences are unplanned or unwanted. Having agency in one's self-transformation can be distinguished by looking at the four main ways in which people change who they are throughout their lives: through voluntary decisions (Paul 2014), nonvoluntary experiences happening to them (Carel and Kid 2020), and through slow incremental changes (which can also be divided into voluntary and nonvoluntary) such as picking up habits (Carlisle 2014). If there are four ways of undergoing a transformation of the self, then we should also distinguish between four kinds of agency in self-transformation:

A. The transformative voluntary decisions give rise to the agency of having the option to make a decision (agency in deciding);
B. transformative nonvoluntary events - the agency lies in coping or adapting to the event in a way that is consistent with one's values and life-goals (agency in coping);
C. habitual voluntary micro-experiences - the agency lies here in deciding to pick up that habit (agency in habit-formation);
D. habitual nonvoluntary micro-experiences - the agency is expressed in coping/adaptation to the unwanted habits and making sense of them (agency in coping).

Formative agency can be thus split into four distinctive modes: agency to decide, agency to cope with external events, agency to form habits, and agency to cope with unwanted habits. This four-fold taxonomy of formative agency allows us to describe and evaluate the kinds of (technical) environments that foster self-transformation.

This taxonomy of formative agency will be further refined in the next section when we look at aspects of the self that can be affected by transformative experiences. For now, I will use it to classify the kinds of transformative experiences that regular users of SMPs can undergo.

Formative agency afforded by a socio-technical environment

Concerning the first two types of transformations of the self (A and B), both can be reliably shown to happen on online platforms. Transformative voluntary decisions can happen in an online environment, like in any other offline social environment. There is nothing special about SMPs as spaces for voluntary transformative decisions than other social spaces except for the exposure to a wider public and global audience. Someone can use the affordances of social media to deliberately trigger a transformative experience, for example, by calling out a public figure and testifying about previous harassment or by starting an activist movement online. Non-voluntary experiences can also happen on SMPs, by undergoing social experiences with traumatic effects, such as becoming viral or being cancelled online, which facilitates becoming the target of sudden storms of hate or of moral outrage suddenly directed at one's user profile. Waves of outrage online can be experienced as highly traumatic and hard to defend against or prevent (Sawaoka and Monin 2018), with some people having killed themselves following this kind of online harassment (Cocking and van den Hoven 2018). There are also online nonvoluntary transformative experiences that can lead to flourishing and self-discovery, and some can be fostered through deliberate design choices such as designing emotional affordances - i.e. affordances for emotional experiences that destabilise the self (Gaggioli, 2015) by adding exposure to serendipitous information through design or by creating possibilities for users' own self-experimentation online (Napolitano, 2013). Seeking exposure to intentional serendipitous information is agentic (Copeland 2022), hence in line with the idea of self-induced transformative experiences.

The other two types (C and D) of self-transformation can also be easily documented on social media platforms. Online habits have been studied in depth; however, it is hard to assess how voluntary the picking up of these habits is and how suggestible users are to picking up new habits online. Negative examples of habitual online self-transformation are the users' declining of mental health and well-being, with cases of eating disorders and self-harm frequently attributed - to some extent - to the heavy social media usage (albeit what kind of usage matters quite a lot, see for example an argument made by Osler and Krueger 2022). In these cases, we can speak of an environmental push towards nonvoluntary habits, however, there is little to no discussion of the agents' adaptation and coping to these online self-transformations. Concerning the voluntary habitual self-transformations, some self-tracking devices allow one to publicise the progress in acquiring a habit, making it visible to other online users, often to accountability groups online (Wieczorek et al. 2022) and this is adding a social re-enforcement element to otherwise a solitary decision. In these cases, the user employs social support to pick up a desirable habit such as maintaining a healthy diet, a fitness regime, or ditching an addiction. SMPs act as scaffolds that make the desired habit much easier to acquire due to the increased support and social accountability. The formative agency of the user lies here in choosing to ask for the other's support and following through with this commitment.

For all four cases of transformative experiences, SMPs can easily enable self-transformations for their users, be those voluntary or nonvoluntary, just as in any other offline social environment. This is not yet enough to say what kinds of environments for self-transformations are SMPs and, hence, to evaluate them normatively as hostile or beneficial, because I have not yet established how much formative agency online users actually have. We need a more precise account of the dimensions of the self that are amenable to change through online transformative experiences. This is the aim of the next section: fleshing out this taxonomy of formative agency in more detail.

An enactive notion of normativity for the changing self.

Enactivism belongs to the 4E family of situated cognition theories, and it distinguishes itself by rejecting the Cartesian duality of

mind-body and instead focusing on the interaction as a concept that shows the mark of cognition: "the human mind arises as an emergent phenomenon from the dynamic interplay between brain, body, and environment" (Van Balen 2025, p. 4). For enactive theorists like Thompson and Varela, it is a mistake to call only humans smart or intelligent since any organism that pursues its goals is showing cognition, including a swarm of bacteria going towards a sugar crystal (Varela et al. 2016). Enactivism does not discern between higher and lower cognitive processes, all adaptation of an agent to its environment is "smart" and the mark of cognition. From an enactive perspective, adaptation is the mark of agency of an organism: "Because living systems are adaptive, they must also be regarded as agents … Not only, given their autonomous organisation, living systems define their own identity over and against their environment." (Bogotá 2024, p. 4) It is precisely in how an organism adapts to its environment that one sees its agency and, we might add, its identity. In an enactive conception, self-identity is given by how an organism interacts with others and its environment: "The enactive approach holds that biological and mental phenomena are continuous, which means that it characterizes the identity of cognitive beings by similar principles and concepts as the identity of living beings" (Kyselo 2014, p. 2). The enactive approach, grounded in interaction and in identity as adaptation is useful for selecting the extreme ends of a normative spectrum, between the pathological and the flourishing.

Any environment in which an agent acts needs to allow for some degree of the agent's adaptation, otherwise, it should be considered toxic. What modes of adaptation are then possible? Right from the start, three distinctive modes become discernible: when an environment does not allow the agent to adapt when the environment poses no resistance to the agent (hence, no adaptation is needed), and when the environment poses enough resistance to demand adaptation from the agent. To summarise, the potential for self-transformation of a user situated in an environment can be placed on a spectrum between flourishing (understood here as developing one's potential or capabilities to the maximum extent allowed by the environment), and pathological rigidity at one end (a non-changing self that hinders one's adaptation to a changing environment), and on the other end, pathological flexibility of the environment (when the environment poses no resistance to the self, it requires no adaptation). In addition, on the side of the agent there needs to be also some kind of resistance to the environment, without falling into rigidity. When someone is too influenced by their environment and changes who they are based on cues from the environment, they will develop an unstable identity. This would be pathological flexibility because the self-identity has no time to coalesce into a narrative of the self, as it keeps changing.

We all fall somewhere on a spectrum of self-change: some of us consider our identity to be fixed once we become grown-ups, and we carry the same self-narrative throughout the years in all environments we circulate, while others are more on the flexible end and do respond to the changes in their social environment, changing their self-narrative as they age and gather more experiences. The spectrum is used to illustrate only what counts as pathological and hence to be avoided, namely the two ends: rigidity and over-suggestibility of the subject, and the same holds for the environment: overly rigid and overly loose. The normative evaluation of the environment needs to happen between these two poles of rigidity and looseness.

Any environment proper for self-formation must open up sufficient possibilities for self-formation through adaptation, such that none of these emerges as necessary, and, at every point in time, the agent has at least two options for coping with the situation. When only one option is visible from the individual's point of view, it appears quasi-necessary. For example, when someone is imprisoned, their perceptions of possible futures are constrained by their release date. Life is supposed to start after they get out of prison, which is not usually seen as a space of possibilities, given the constraints and the repetitive nature of activities going on in prison. This makes them insensitive to how they change while being in prison and perhaps less able to be agentic about this change. Thus, while any social environment affords coping and adaptation, hence situated agency, the prison is a particularly restrictive environment for formative agency. In such restrictive environments, one is usually focused on survival and less prone to transform oneself in ways that would lead to one's flourishing.

If an environment presents the agent with too many options for becoming someone else, one can postpone deciding and choosing a course of action. This seems to be the case with social media, which makes it an environment that is disruptive to self-transformation in a unique way. SMPs allow us to manipulate our public persona and how we are perceived by others by performing various identities online (Shin et al. 2009; Erden 2016, p. 3). If we can perform various selves under the guise of profiles and fake online identities, we may believe we are those fake selves. If one of the selves we perform online does not turn out as expected, meaning that it does not gather the desired reactions, one can always decide to pursue another path. We can be trapped in a realm of fantasy with so many possible futures. We can refuse to exercise our formative agency right now and postpone the decision of who we want to become so that we never extinguish a possibility. The fear of choice is the fear of extinguishing possible futures by choosing only one. Lost in fantasy, we can already think we made that choice and are already who we want to be. If we find the right audience online, one can be seen as a famous writer, a truth-seeker, a social justice warrior, an activist, or an organiser of social rebellion. If the right social media bubble is found, others can perform and validate any identity with no discernible friction from the outside world.

Using the conceptual distinctions made thus far, I propose the following taxonomy as way to classify and evaluate any techno-social environment for its transformative potential. Every environment can be charted across two major dimensions: the kind of formative agency it affords, and the flexibility to adaptation an agent's adaptation. The following table illustrates these dimensions, amounting to a 2D conceptual space:

Formative agency modes	Adaptability afforded by the environment
A. Voluntary decision making	X. Loose environment - No resistance to any assertion of the self
B. Nonvoluntary events	Y. Rigid environment - resistance to self change
C. Voluntary habit formation	Z. Flexible environment
D. Nonvoluntary habit acquiring	

In the next section, I will use this framework to illustrate the few known cases of self-transformation on social media and online platforms in general, aiming to show how we can evaluate SMPs as transformative environments after all.

Social Media Platforms as transformative environments

To evaluate how the human self changes through the usage of a techno-social environment such as SMPs, we need some operationalisation of the notion of the self, namely to understand what actually changes. In an enactive understanding, the self-identity of an agent is constituted by "self-generated, self-determined precarious networks" (Kyselo 2014, p. 2), or what has been called a network of patterns of interactions, the patterned-self, a theory

coined by Gallagher and Daly (2018). This theory depicts the self as a dynamic set of patterned interactions between several dimensions of the self, none of which is sufficient to define the self on its own. Gallagher and Daly describe the self "in terms of such dynamical neural and narrative processes, [a]s not a fixed entity but … rather an ongoing production which brings a real but contingent coherence to an evolving (or in some cases, devolving) stream of sensations, thoughts, emotions, desires, memories, and anticipations." (Gallagher and Daly 2018, p. 1). These dimensions of the self include the moral self (our norms and values), the embodied self, past experiences, social interactions and social capacities, affective components, narrative capacities, reflective capacities, and extensions of the self. The gist of this theory is that neither dimension of the self is sufficient to give someone a stable identity, but some of these dimensions need to happen minimally in one person: "minimal embodied, minimal experiential, affective, intersubjective, psychological/cognitive, narrative, extended, and situated" (Kyselo 2014, p. 1).

Technically, all of the self dimensions could change after interacting enough in a social environment, but in the case of social media, which consists mostly of "virtual" interaction, it is hard to imagine how it could change our embodied dimension of the self, and also the cognitive and psychological traits, which seem to be long-lasting traits one grows into. Thus, among these dimensions of the patterned self that can change while being an online user, we should focus mostly on the behavioural aspects, specifically the habits which build the character of a person, the reflective capacities (which allow one to compare one's long-term goals with current actions and steer one's behaviour), the narrative capacities, the normative factors (the values and norms one is committed to), and the technological extensions of the self.

I have already touched on habit formation (included in C and D types of self-transformation), which is also one of the most studied effects of SMP usage. Habits can be picked up voluntarily and non-voluntarily online, and UX designers can deliberately design for habit formation. Meanwhile, the other two aspects of the self have been less discussed thus far. Similarly, significant past experiences that can alter the self are the traumatic ones or the ones conducive to sudden overwhelming positive outcomes (winning a prize, becoming a revered influencer, etc.) - and these are included in all the modes of self-transformation, particularly in A and B. The interesting cases that remain to be fleshed out are the remaining ones: 1) the social relations, 2) normative aspects, 3) reflective capacities, 4) narrative capacities, and 5) the extensions of the self. Using Gallagher and Daly's dimensions of the self (2018), the most relevant self dimensions that could change are one's behaviours (habits), experiences, social relations, normative aspects, reflective capacities, narrative capacities, and the extensions of the self.

1) Social relations. Any social environment in which we are immersed has the capacity to alter our selves and our self-understandings because we encounter others. Social relations are very strong predictors of the capacity to change. The first changes of the self we undergo as infants through the relations with our families; then the educational institutions have a strong effect on our selves. One would think that grown-ups are not that changeable anymore, but there is still identity and self-change, most notably through the social relations one enters in the workplace or one's friends. Online, we are members of groups and communities that stress a certain aspect of our social selves (the membership aspect) and, if we become dependent on these online groups for validation or meaningful socialisation, these have the capacity to change who we think we are. One of the most devastating forms of self-change happens when users radicalise by becoming members of online groups that encourage them to identify with certain beliefs and actions.

2) Normative aspects. The normative aspects that define oneself and are amenable to change are the values and the norms one subscribes to, as well as one's moral worldviews. To what extent do one's values and norms change as one uses online social platforms? Most research has been done on belief change and influence, on nudging (Specker Sullivan and Reiner, 2021) and online manipulation (Klenk 2022). There is still very little research on online conversions, except for self-radicalisation (Alfano et al. 2018) or value change fostered by online interactions (Steinert 2021; Friedrich et al. 2022). Still, if an online environment manages to change the values and norms we ascribe to, this would be a transformative change.

3) Reflective capacities. The reflective capacities help us compare our current selves with our long-term goals and values. If we are disconnected between the desired long-term self and the current self, reflecting on this disconnection should launch us into some corrective action (see also Aydin 2021). Without these reflective capacities, we only have short-term agency and lack the planning dimension, which is core to human agency. Do SMPs foster or sabotage reflective capacities? Insofar as these platforms are designed for maximising user engagement by triggering short-term reactions, the online social platforms seem to be mainly hostile to inducing reflection, as others have argued (Voinea et al. 2020; Bentvelzen et al. 2022; Lutzke et al. 2019). SMPs seem to promote a short-term focus of their users on the here and now and disconnect their users from their long-term goals and values and thus from their desired selves. This means that SMPs are disruptive to the formative agency of users in their voluntary form insofar as these platforms disrupt self-reflection and critical thinking (Steinert et al. 2022). Further research of an empirical nature is needed to clarify to what extent capacities for self-reflection are hindered online and in what ways.

4) Narrative capacities. While for Gallagher and Daly (2018), the narrative capacity is one of the components that make the self, according to Muriel Leuenberger (2024, p. 3-4), the self-narration is the process that manages to integrate the disparate shards of the self into one coherent narrative, using the self-patterns and their dynamic interaction to tell a story of who one is. A self-narrative is "a story telling one's life events from a personal perspective, reflecting character traits, goals, and values" (Leuenberger 2024, p. 6). Self-narratives give unity to the self over time and change, explaining its persistence not by identifying the same features over time but by giving meaning to changes, even in the most radical changes (Leuenberger 2024, p. 10). Thus, a change in the self-narrative of a user will signal a change in one's self-identifying core traits. When transformations of the self are hard to detect, the agent's changes in self-narratives can be taken as a proxy for identity changes, so any experience that changes one's self-narrative should be seen as transformative.

5) Technological extensions of the self. Just as our cognition is extended through various devices and prostheses we may use to store our memory (Clark and Chalmers 1998), it is possible to extend ourselves through the technologies we use. One can think, for example, of how much one's smartphone or smartwatch is part of one's identity. This is not about using technology instrumentally to achieve one's purposes but about how much that technology is part of our self-identity, for example, if we build it into our self-narrative. Some people take pride in being Mac users or iPhone users, and some people use social media to have an altogether different persona, that of an influencer or of a comedian, and this dimension of the patterned self needs the technology to be reliably accessible for the agent, such that the technology can extend the self. It is at least conceivable that the self gets extended through social media, especially for people whose online persona is bigger than their offline identity.

It seems then that SMPs hold the potential to affect significantly some of the dimensions of the patterned self that Gallagher and Daly (2018) conceptualised. On this account, SMPs should be seen as environments that can lead to self-transformation. In the remainder of this section, I will use the taxonomy of self-transformation to evaluate some of the most

remarkable cases in which SMPs change the selves of their users, as documented in previous social media scholarship.

Online self-radicalisation (D-X). The most pathological cases of self-transformation online are those of self-radicalisation, people joining extremist organisation after consuming enough online content. The few documented cases of self-radicalization online (Alfano et al. 2018; Missier 2022; Ledwich and Zaitsev 2020) are spectacular, but many other smaller changes may go unnoticed and yet, cumulatively, may end up with a transformation of who we are. For example, someone can start a fitness routine because an influencer inspired them, and then, years down the line, they become this person that all their friends identify as "health obsessed" who cannot stay late at social events because they plan to go out running the next morning. Such changes in one's self-identity through online interactions (be those with information, with influencers, or online communities) seem trivial, too small to make a momentary change, but possibly significant in the long term. On the proposed taxonomy, self-radicalisation would fall under the categories of D-X: users fall into a non-voluntary habit (D) of watching videos fed to them by the algorithm, and they create fantasy selves in which they get to save the world from whatever threat that organisation has focused on (X - loose environments). By watching these videos, users who self-radicalise form a fantasy self that is not amenable to feedback at least in the first instance, since this change happens in isolation, until they move on and actually contact the members of the organisation they now aspire to be part of.

Unreflective Habit Formation on SMPs (D-Z). Habit formation on social media has been one of the most studied online phenomena, albeit limited to short-term spans (Hu et al. 2018; Anderson and Wood 2021). SMPs are the hosts of "one of the most common – and controversial – forms of habitual behavior in contemporary society" (Bayer et al. 2022, p. 1). From a behavioural psychology lens, habits are frequent behaviours that trigger rewards (Anderson and Wood 2021, p. 85). For example, if the stimulus is the boredom felt while waiting in line at the supermarket, the response behaviour is pulling out one's phone and scrolling through social media feeds, rewarded by short-term entertainment (Bayer et al. 2022, p. 1). Checking a social media feed whenever one has a break can become addictive (van den Eijnden et al. 2016), and for some people, it interferes with their work or offline relations. However, the issue at stake here is not whether a habit gained through SMP usage is beneficial or harmful - which was the main concern for previous studies - but rather the extent to which these habits shape user identities and interfere with self-transformation.

The habits that we build deliberately contribute to forming our self-identity: habits are the building blocks of virtues and habits are signifiers of personal values (Carlisle 2014; Matthews 2021). Habits are actions we perform frequently, often without being aware of them, as these tend to fade into the background after a while. A habit can be smoking or starting to exercise right after waking up. Smoking and exercising will say something about the person who performs them, their values and goals, but what it says depends on the narrative the subject builds around that habit. A smoker who assumes their habit as part of their identity might be sending the message, "I do not care about long-term illness down the line; I am happy to be a smoker now, and this is who I am." Another smoker might be saying, "I would like to quit, but I cannot; smoking is my curse and a sign of weak will". Two narratives about the same habit give rise to different narratives of the self as weak-willed or indifferent to ill health. One cannot disown a habit as easily as one would do with a physical reflex; smoking or exercising does not happen to someone; they choose to build it into their routines. Hence, we can use habits as pointers to aspects of our self-identity. To say that someone is a smoker, a gambler, or a fitness enthusiast is to communicate something about their identity.

To what extent can habits acquired on SMPs have a negative effect on one's formative agency? Sylvie Delacroix, (2022) suggests this is a genuine possibility through the term of "habit rigidification" (p.136), which can occur in online environments because of the platform's tendency to optimise for user engagement. When maximising for user engagement, the information users see is based on "our machine-readable past" (Delacroix, 2022, p. 136): if we clicked on pictures of cats shown in our social media feed, more cats will be shown in the next interactions, under the assumption that whatever interested us in the past will be interesting again, hence we should see more of it. Sylvie Delacroix (2022) argues that environments fostering self-formation need to have embedded structurally several possibilities of "normative experimentation" (p. 133). However, normative experimentation does not happen in a vacuum; rather, we need infrastructures that allow experiments with the self (Jaeggi 2014) and with our values. For this, the infrastructures need to be designed for serendipitous encounters (Ross and Copeland 2022). In addition, we need moments in which we become reflectively aware of the habits we have accumulated down the line through our daily interactions with these platforms. Such moments of reflective awareness need to be designed in the user experience, which is usually meant to be the opposite of reflective, aiming for frictionless interaction (Bentvelzen et al. 2022), which ultimately leads to the accumulation of a load of unreflective habits.

The issue at stake is not the emergence of this or that particular habit for SMP users but the extent to which these habits rigidify their identity and hinder them from developing new self-identities by rearranging features of their self-identity. When habits appear unnoticed in our lives without our endorsement, and then habits become stuck as labels of who we are, then we are dealing with a hindering of our *formative agency* – defined here as the agency to change one's identity through self-transformation. Because I am using a situated concept of agency, formative agency is not something belonging to the agent alone, it is also afforded to an extent by the environment one inhabits. Formative agency is made up of the willingness one has to change and the affordances of the environment that can lead to that change. I may want to become an Olympic-level swimmer, but if I live in a place without swimming pools, my agency to become a swimmer is heavily limited by environmental constraints. On the other hand, if I do not want to become a swimmer, the number of pools in my surroundings will not affect my formative agency in that direction.

Identity rigidification (D-Y). From the three normative dimensions outlined above, it seems that identity rigidification is the most problematic on social media, as outlined by previous scholars. Online users perform several identities online (Walther & Lew, 2022), and sometimes this identity can become inescapable, as it gets rigidified. Digital identity is not a unitary performance of the authentic self, as there are many layers of interacting with others, telling narratives of ourselves and getting feedback from others about the salient features that we managed to get across through digital communication. Identity is always relational, and digital identity even more so, as digital interactions "impact[s] on the way we engage with others and the ways in which we make our voices heard, hear the voices of others, and how much time we give to each" (Erden 2016). SMP daily interactions can lead to an identity rigidification for online users, but this needs to be taken in the broader context of digital identity being already quite fluid and hard to pin down to one clear persona.

Sylvie Delacroix, (2022) argued that information monotony plays a significant part in this rigidification happening online, and that serendipity is sacrificed to monotony. Spaces for self-

experiments need encounters with serendipitous information (Napolitano, 2013) to change who we are. Intentional serendipitous orientation is agentic (Copeland 2022) and in line with the ideas of self-induced transformative experiences. Thus, the issue here is not about the bad luck of falling into unwanted habits or being trapped in online echo chambers but about having the courage to experiment with dimensions of the self, which the environment seems to discourage. On SMPs, most users are algorithmically profiled by machine-learning algorithms that prefer our predictable selves to experimental selves that suddenly stop looking at cats and consuming a kind of political news. The predictability of who we are is encapsulated in user profiles, and it will be disturbing for machine-learning algorithms to have such profiles change.

However, this does not mean that we are actively discouraged from normative experiments. Despite what Delacroix, (2022), Jaeggi (2014), and Napolitano (2013) suggest, there is nothing deterministic in being shown a monotonous stream of information online and then continuing to be the self that consumes and endorses that information. While habit formation and identity rigidification through monotonous exposure to the same kind of information are good clues, we need a more systematic account of this possibility to transform oneself brought on by an environment.

Online identity rigidification may happen after receiving socially skewed feedback online. When an online platform distorts the social feedback from others, it actively fosters skewed relations. This can happen by entering into relations with certain people and filtering the kind of feedback we get from them; we select the kind of image that others reflect back to us. We can enter one-dimensional relationships (parasocial) with role models or influencers where there is no feedback, but we imagine that there is a close relationship there, and we interpret their broadcast messages as aimed at us ("This video speaks to me!"). We can associate with certain people and groups and then ignore their feedback if it is critical. In a real-life situation, it is hard to ignore the feedback that others give to us since it is addressed to us directly. If we do something considered harmful by our close acquaintances, they will signal it to us. But online, our circle of acquaintances is much larger, and we can dismiss them easily by deleting their messages, making their posts invisible, reporting their comments, etc. On SMPs, it is too easy to silence someone by using the affordances embedded in these platforms for reporting and blocking. The ultimate form of unwanted self-transformation is a user's self-radicalisation while watching videos online (Alfano et al. 2018). In these cases, the relations created are parasocial and hence imaginary, yet still strong.

Secondly, when we promote a certain narrative about ourselves through social media posts by performing a self with certain values (such as a sustainable lifestyle, minimalist, healthy, expertise, etc.), this identity tends to become inescapable for the one performing it online because there are always traces left of this identity with which the user can be later confronted by others ("Last year you posted that you were a sustainability advocate, why do you take plane flights now?"). This is the case with most influencers who do not have the luxury of being themselves as changing selves. If I promote myself as an expert in nutrition and healthy diets through my posts or videos, and if I gain a wide follower base through this persona, it becomes hard to distance myself from this identity because I would lose my followers and the online prestige. Still, this is not impossible since the cases of apology videos are quite frequent on YouTube, cases when an influencer is exposed as hypocritical or doing things that are not aligned with their promoted values, and then the influencer "comes out" with an apology video in which they admit their misdeeds and then continue to promote that value after this repentance. Still, the cases where one admits having a fake value and then shifting their public image away from that value are extremely rare. This is because value change and identity transformations will alienate followers, and the influencers do not have enough trust in their new identity to be attractive enough to gain new followers.

A third dimension of self-rigidification concerns the user's epistemic agency, which is the agency that one has in pursuing knowledge, understanding, or justifying opinions (Olson 2015). When we have access to only certain kinds of information that reinforce our worldviews, it does not allow us to challenge what we know and who we are and ultimately explore our values. By promoting engagement online, personalisation algorithms expose users to information that either they would agree with or with information that they would find outrageous and thus re-enforce their convictions that their worldviews are right after all because their moral indignation is shared by so many others (Mihailov et al. 2023; Nguyen and Williams 2020). Storms of collective moral indignation online are not conducive to any epistemic transformation since their main purpose is performative - to perform a certain moral identity in front of others who approve - and not about changing beliefs or becoming reflective about what triggers our indignation. Thus, this dimension of algorithmic design, which is selected for worldview confirmation or outrage (Voinea et al. 2024), is detrimental to any kind of epistemic transformation. While we are exposed to information that contradicts our worldviews quite often online, the ways in which we are exposed to this epistemic difference are unlikely to have a transformative effect.

The threat of identity rigidification due to SMP interactions seems to be more significant than the habit rigidification previously discussed. Identity rigidification interferes with the possibility of changing one's self-narrative, and the main mechanism of action seems to interfere with personal normative experiments by limiting users to a range of predictable and quite limited experiences (Jaeggi 2014). Having the option to experiment with one's life is an important condition of moral agency. Jaeggi (2014) argues for these self-experiments as being not merely aesthetic experiments for the sake of experiments; rather, these are attempts to solve a problem that is novel to us: "the self is less a work of art one makes oneself into than a practical-experimental process one is caught up in" (Jaeggi 2014, p. 189). Experiments occur in our lives as attempts to solve novel problems (Jaeggi 2014, pp. 65-66), and this aspect is precisely agentic. Imagine that your best friend has betrayed you by disclosing your secret to others. The friend's betrayal is a problem you must solve as a moral agent. You can experiment with possible options for moral actions: cutting off that friend, ghosting them, confronting them. If certain actions tell you something about your moral character that could not have been predicted otherwise, these count as experiments with the self. These experiments bring increased self-knowledge and self-transformation, but it seems that, online, these self-experiments are discouraged by design since platforms assume a stable self for their users, the user profile.

Conclusions

How should we see SMPs as environments fostering the users' self-transformation? The taxonomy I have proposed thus far allows for a complex mapping of its transformative potential. If we were to look only at the first dimension, the formative agency, SMPs are like any other offline social environment: complex environments that afford but also hinder formative agency, depending on the user's context. However, if we add the second angle, that of dimensions of the self that can get transformed, almost all the patterned dimensions of the self as outlined by

Gallagher and Daly, seem to be influenceable through social media interactions, except perhaps for the embodied dimensions and the stable character traits.

Concerning the adaptability dimension, an environment can be seen as hostile to self-transformation in a variety of ways. The modality scale opens up the possibility of normative evaluation: either opening up too few options or too many or showing users only necessary actions and leaving few choices for them. An environment that discourages long-term reflection on one's values and goals or which introduces hidden values in a user's life is also hostile. Finally, any environment that makes visible to the users only one facet of their identity (the so-called "identity rigidification"), as if this is all who they are, is also hostile to self-transformation. I have argued that an environment aiming to foster self-transformation needs to open up more than one possibility for its users to perceive their future selves while avoiding opening up too many possible self-transformations. Such an environment should find ways to avoid escaping into the realm of fantasy. Currently, many users employ social media to create identities that they do not actually have, which is possible because of the profile options that are disconnected from the actual identities of users. This seems to be the most dangerous feature of SMPs, unmatched by other social environments that come with their own version of friction by confronting users with limitations in what they can do and imagine. A fantasmatic self is an imagined self that has no connection with the actual self and which also risks becoming entrenched because an online user's networked connections mirror this self back to us and act as if we actually are this fantasy self. This kind of danger is hard to replicate in the offline social realm, where identity checks are embedded in how we relate to others daily. Compared to 'offline' social realms, SMPs have this distinctive quality that they afford to escape into a fantasy self and craft unrealistic self-narratives that ultimately harm the user's actual formative agency.

When evaluating the ethics of socio-technical environments such as online platforms, particularly SMPs, we should pay attention mainly to the incremental types of change, those triggered by environmental pressures slowly accumulating over time, the C and D types in the formative agency dimension. It is still possible to undergo a transformation of the first two types, A and B, even on social media: think of people who are the targets of a wave of online hate and harassment - this is a nonvoluntary traumatic experience that they undergo and cannot escape, with significant effects on their sense of self, or when people decide to change their lives and announce or pledge this to their social network. Meanwhile, the ethics of technological influence is primarily focused on discernible actions (Specker Sullivan and Reiner 2021) by evaluating which actions are fostered, scaffolded, afforded, hindered or promoted by a socio-technical system while seemingly assuming static goals and values for the user side - hence this ethical framework is mostly suited for types A and B of experiences as outlined above. However, the ethics of social media platforms stemming from the ethics of influence will focus on problematising instances when user's interaction with an online platform leads them to undertake actions that are not consistent with the users' values, goals or reasons, such as in cases of technological manipulation (Klenk 2022). This is because most approaches in the ethics of technological influence assume that a) humans are agentic when we do things in the world that we want to do, and b) our values and goals are constant and known to us. Both assumptions are problematic because these rely on a conception of agency as deliberate action based on one's own decisions and self-knowledge about these values and goals.

This paper inquired whether SMPs have the potential to be spaces for the self-transformation of their users, stirred by the observation that this self-transformative dimension has been mostly absent from the scholarship on the philosophy of technology. I argued that this is an important normative dimension of any social space which deserves its own taxonomy.

I have argued in this paper that the ethics of technological influence can be expanded and enriched with a concept of situated agency and an enactive evaluation of adaptability afforded by an environment. Using a situated concept of agency, we should look into how an agent is afforded or even pushed to undergo transformative experiences. I used the types of transformative experiences from Carel and Kid to argue that one agent can effectively have four modes of formative agency in any given situation, namely how one responds to the pressures of an environment. However, since it could be argued that any kind of adaptation is legitimate since we become who we change into, I added an explicit normative dimension pertaining to the environment that fosters the agent's transformation of the self: borrowing from the terminology of enactive theories of cognition, how an environment responds to an agent's adaptation is also important and can make the difference between a hostile environment and a flourishing-conducive one.

This proposed taxonomy, designed for a normative evaluation, is amenable to some empirical confirmation insofar as some aspects of SMPs and the experiences these platforms give rise to for their users can be measured, albeit long-term transformative processes are still difficult to observe. We can meaningfully measure the extent to which a social media platform enforces non-voluntary coping for users with decisions taken by others (with transformative potential) or how it enables the adoption of non-voluntary habits. The main advantage of this framework is that it does not presuppose a certain notion of flourishing as desirable for individuals and instead recognises their situated nature: some agents have a very limited capacity range to be developed in hostile environments. Rather, this framework takes flexibility to change to be a better proxy for how friendly an environment is for the user's self-transformative processes. This aligns with previous work by Delacroix (2022) and Jaeggi (2014) on identity rigidification. However, the framework proposed here is multidimensional and allows for a more complex space for normative evaluation of a social environment such as SMPs.

Data availability

No data were generated or analyzed for this paper which is theoretical in nature.

Received: 29 February 2024; Accepted: 29 January 2025;
Published online: 17 February 2025

References

Alfano M, Carter JA, Cheong M (2018) Technological Seduction and Self-Radicalisation. J Am Philos Assoc 4(3):298–322. https://doi.org/10.1017/apa.2018.27

Anderson IA, Wood W (2021) Habits and the electronic herd: The psychology behind social media's successes and failures. Consum Psychol Rev 4(1):83–99. https://doi.org/10.1002/arcp.1063

Aydin C (2021) Extimate Technology: Self-Formation in a Technological World (1st ed.). Routledge. https://doi.org/10.4324/9781003139409

Bayer JB, Anderson IA, Tokunaga RS (2022) Building and breaking social media habits. Curr Opin Psychol 45:101303. https://doi.org/10.1016/j.copsyc.2022.101303

Bentvelzen M, Woźniak PW, Herbes PSF, Stefanidi E, Niess J (2022) Revisiting Reflection in HCI: Four Design Resources for Technologies that Support Reflection. Proc ACM Interact, Mob, Wearable Ubiquitous Technol 6(1):1–27. https://doi.org/10.1145/3517233

Bevir M (2017) Situated Agency: A Postfoundational Alternative to Autonomy. In Finite but Unbounded: New Approaches in Philosophical Anthropology (pp. 47–66). De Gruyter. https://doi.org/10.1515/9783110523812-4

Boer M, Stevens GWJM, Finkenauer C, De Looze ME, Van Den Eijnden RJJM (2021) Social media use intensity, social media use problems, and mental health among adolescents: Investigating directionality and mediating processes. Computers Hum Behav 116:106645. https://doi.org/10.1016/j.chb.2020.106645
Bogotá JD (2024) Life, sense-making, and subjectivity. Why the enactive conception of life and mind requires phenomenology. Synth 204(3):101. https://doi.org/10.1007/s11229-024-04746-1
Carlisle C (2014) On habit. Routledge
Clark A, Chalmers D (1998) The Extended Mind. Analysis 58(1):7–19
Carel H, Kidd IJ (2020) Expanding transformative experience. Eur J Philos 28(1):199–213. https://doi.org/10.1111/ejop.12480
Cocking D, Hoven J (2018) Evil online. Wiley Blackwell
Colombetti G, Bogotá JD (2024) The Tacitly Situated Self: From Narration to Sedimentation and Projection.Topoi. https://doi.org/10.1007/s11245-024-10044-9
Copeland S (2022) Metis and the art of serendipity. In The Art of Serendipity (pp. 41–73). Cham: Springer International Publishing
Delacroix S (2022) Habitual ethics? Hart Publishing, an imprint of Bloomsbury Publishing
Desmond H, Huneman P (2022) The integrated information theory of agency. Behav Brain Sci 45:e45. https://doi.org/10.1017/S0140525X21002004
Digital Services Act: Ensuring a Safe and Accountable Online Environment (2024). https://doi.org/10.2812/802073
Dung L (2024) Understanding Artificial Agency. The Philosophical Quarterly. https://doi.org/10.1093/pq/pqae010
Erden YJ (2016) Digital identity: Finding me. AISB Convention, University of Kent
Floridi L (Ed.) (2015) The Onlife Manifesto. Springer International Publishing. https://doi.org/10.1007/978-3-319-04093-6
Friedrich O, Gerlek S, Seifert J, Schleidgen S (2022) Value change through information exchange in human–machine interaction. Prometheus 38(1). https://doi.org/10.13169/prometheus.38.1.0057
Gaggioli A (2015) Transformative Experience Design. In A. Gaggioli, A. Ferscha, G. Riva, S. Dunne, & I. Viaud-Delmon, Human Computer Confluence (pp. 97–122). De Gruyter Open. https://doi.org/10.1515/9783110471137-006
Gallagher S, Daly A (2018) Dynamical Relations in the Self-Pattern. Front Psychol 9:664. https://doi.org/10.3389/fpsyg.2018.00664
Heersmink R (2020) Varieties of the extended self. Conscious Cogn 85:103001. https://doi.org/10.1016/j.concog.2020.103001
Holland B, Linch A (2016) Cultivating human and non-human capabilities for mutual flourishing. In The Oxford handbook of environmental political theory (pp. 413–428). Edited Oxford: Oxford University Press
Hu T, Stafford TF, Kettinger WJ, Zhang X"Paul, Dai H (2018) Formation and Effect of Social Media Usage Habit. J Computer Inf Syst 58(4):334–343. https://doi.org/10.1080/08874417.2016.1261378
Jaeggi R (2014) Alienation (F. Neuhouser, Trans.). Columbia university press
Kyselo M (2014) The body social: An enactive approach to the self. Front Psychol 5. https://doi.org/10.3389/fpsyg.2014.00986
Klenk M (2022) (Online) manipulation: sometimes hidden, always careless Rev Soc Econ 80:85–105. https://doi.org/10.1080/00346764.2021.1894350
Ledwich M, Zaitsev A (2020) Algorithmic extremism: Examining YouTube's rabbit hole of radicalisation. First Monday. https://doi.org/10.5210/fm.v25i3.10419
Leuenberger M (2024) A Narrative Pattern-Theory of the Self. In M. Herrmann (Ed.), Personhood, Self-Consciousness, and the First-Person Perspective (pp. 127–14). Brill | mentis
Lutzke L, Drummond C, Slovic P, Arvai J (2019) Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. Glob Environ Change 58:101964. https://doi.org/10.1016/j.gloenvcha.2019.101964
Matthews S (2021) Habit. In H. LaFollette (Ed.), The International Encyclopedia of Ethics (1st ed., pp. 1–11). Wiley. https://doi.org/10.1002/9781444367072.wbiee944
Mihailov E, Voinea C, Vică C (2023) Is Online Moral Outrage Outrageous? Rethinking the Indignation Machine. Sci Eng Ethics 29(2):12. https://doi.org/10.1007/s11948-023-00435-3
Missier CA (2022) Fundamentalism and the search for meaning in digital media among Gen Y and Gen Z. The J for Deradicalization (JD), 33:255–285
Mitchell T, Douglas T (2024) Wrongful Rational Persuasion Online. Philos Technol 37:35. https://doi.org/10.1007/s13347-024-00725-z
Napolitano CM (2013) More than Just a Simple Twist of Fate: Serendipitous Relations in Developmental Science. Hum Dev 56(5):291–318. https://doi.org/10.1159/000355022
Nguyen CT, Williams B (2020) Moral Outrage Porn. J Ethics Soc Philosophy 18(2). https://doi.org/10.26556/jesp.v18i2.990
Olson D (2015) A Case for Epistemic Agency. Logos Epistem 6(4):449–474. https://doi.org/10.5840/logos-episteme20156435
Osler L, Krueger J (2022) ProAna Worlds: Affectivity and Echo Chambers Online. Topoi 41(5):883–893. https://doi.org/10.1007/s11245-021-09785-8
Paul LA (2014) Transformative experience. Oxford University Press
Ross W, Copeland S (Eds.) (2022) The Art of Serendipity. Springer International Publishing. https://doi.org/10.1007/978-3-030-84478-3
Sawaoka T, Monin B (2018) The Paradox of Viral Outrage. Psychological Sci 29(10):1665–1678. https://doi.org/10.1177/0956797618780658
Shin R, Lopes W, Claycomb, Ahn G -J "A Framework for Enabling User-Controlled Persona in Online Social Networks," 2009 33rd Annual IEEE International Computer Software and Applications Conference, Seattle, WA, USA, 2009, pp. 292-297, https://doi.org/10.1109/COMPSAC.2009.46
Specker Sullivan L, Reiner P (2021) Digital Wellness and Persuasive Technologies. Philos & Technol 34(3):413–424. https://doi.org/10.1007/s13347-019-00376-5
Steinert S (2021) Corona and value change. The role of social media and emotional contagion. Ethics Inf Technol 23(S1):59–68. https://doi.org/10.1007/s10676-020-09545-z
Steinert S, Marin L, Roeser S (2022) Feeling and thinking on social media: Emotions, affective scaffolding, and critical thinking. Inquiry, 1–28. https://doi.org/10.1080/0020174X.2022.2126148
Tobia K (2022) Experimental philosophy of identity and the self. Bloomsbury academic
Van Balen B (2025) Can communication Brain-Computer Interfaces read minds?: Demystifying Brain-Computer Interface mediated mindreading with enactivism. Phenomenology and the Cognitive Sciences. https://doi.org/10.1007/s11097-024-10044-5
Varela FJ, Thompson E, Rosch E (2016) The embodied mind: Cognitive science and human experience (revised edition). MIT Press
Van Den Eijnden RJJM, Lemmens JS, Valkenburg PM (2016) The Social Media Disorder Scale. Computers Hum Behav 61:478–487. https://doi.org/10.1016/j.chb.2016.03.038
Voinea C, Vică C, Mihailov E, Savulescu J (2020) The Internet as Cognitive Enhancement. Sci Eng Ethics 26(4):2345–2362. https://doi.org/10.1007/s11948-020-00210-8
Voinea C, Marin L, Vică C (2024) Digital Slot Machines: Social Media Platforms as Attentional Scaffolds. Topoi. https://doi.org/10.1007/s11245-024-10031-0
Wallace K (2019) The network self: Relation, process, and personal identity (1 [edition]). Taylor & Francis
Walther JB, Lew Z (2022) Self-transformation online through alternative presentations of self: A review, critique, and call for research. Ann Int Commun Assoc 46(3):135–158. https://doi.org/10.1080/23808985.2022.2096662
Westley FR, Tjornbo O, Schultz L, Olsson P, Folke C, Crona B, Bodin Ö (2013) A Theory of Transformative Agency in Linked Social-Ecological Systems. Ecol Soc 18(3):art27. https://doi.org/10.5751/ES-05072-180327
Wieczorek M, O'Brolchain, F., Saghai, Y., & Gordijn, B. (2022). The ethics of self-tracking. A comprehensive review of the literature. Ethics & Behavior, 1–33. https://doi.org/10.1080/10508422.2022.2082969

Acknowledgements

This work is part of the research programme Ethics of Socially Disruptive Technologies, which is funded through the Gravitation programme of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.004.031).

Author contributions

Lavinia Marin is the sole author of this paper.

Competing interests

The author declares no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

Not applicable.

Additional information

Correspondence and requests for materials should be addressed to Lavinia Marin.