Enactive Principles for the Ethics of User Interactions on Social Media

How to Overcome Systematic Misunderstandings Through Shared Meaning-Making

Marin, L.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Enactive Principles for the Ethics of User Interactions on Social Media: How to Overcome Systematic Misunderstandings Through Shared Meaning-Making

Lavinia Marin[1,2]

Abstract

This paper proposes three principles for the ethical design of online social environments aiming to minimise the unintended harms caused by users while interacting online, specifically by enhancing the users' awareness of the moral load of their interactions. Such principles would need to account for the strong mediation of the digital environment and the particular nature of user interactions: disembodied, asynchronous, and ambiguous intent about the target audience. I argue that, by contrast to face to face interactions, additional factors make it more difficult for users to exercise moral sensitivity in an online environment. An ethics for social media user interactions is ultimately an ethics of human relations mediated by a particular environment; hence I look towards an enactive inspired ethics in formulating principles for human interactions online to enhance or at least do not hinder a user's moral sensitivity. This enactive take on social media ethics supplements classical moral frameworks by asking us to focus on the relations established through the interactions and the environment created by those interactions.

1 Introduction

Early January 2020, a Tweet went viral claiming that a second impeachment procedure of then-president in office Trump would lead to him losing all the post-presidential benefits. The tweet's initiator, a user, hereafter called "B. C.", did not envision the massive reach of this Tweet to roughly over 175.000 shares and 750.000 likes since he was not an influencer by any standard with only 200 followers. When asked later why he had sent the tweet containing a false statement, B. C. replied that 'he had seen the information pop up somewhere on his Facebook feed and "it made me feel good" (…) "I don't want to mess up the world. I just wanted to make me feel good."'[1] This example of a tweet

with unpredictable effects points to the pervasive issue of unintended moral harm that arises online. But was this tweet morally harmful by any standard?

The harm generated by this Tweet may be considered insignificant given that mainstream mass-media channels quickly and firmly debunked the claim. Yet, one cannot be sure that the users who retweeted and liked the initial Tweet followed mass-media channels and thus were aware of its debunking. There is a chance that many users remained misinformed after the initial exposure to this tweet. Given that B. C. had no intention to harm others since he was unaware of the misinformation, should he be held morally responsible for the viral effects of that tweet? Misinformation sharing is primarily an epistemic kind of harm; specifically, it hinders people from forming justified true opinions about a matter, hence it is not *primarily* a moral issue. However, this is a moral issue because one needs access to accurate information to form moral judgments (Floridi 2013). Secondly, the viral Tweet stirred various feelings for its audience, ranging

✉ Lavinia Marin
lavimarin@gmail.com

1 Ethics and Philosophy of Technology, TU Delft, Delft, the Netherlands

2 Education, Culture and Society, KU Leuven, Leuven, Belgium

---

from euphoria to despair. These feelings were ungrounded, based on a false proposition. To understand the wrongness, imagine the following scenario: someone told the same users that their favourite soccer team had won the UEFA championship. The users would feel elated, cheering and celebrating, until they actually checked the official stats and then would be disappointed about the lie, perhaps furious. Such emotional turmoil has moral consequences since people rely on emotions to form moral intuitions and to decide what stance to take about a particular contentious issue (Steinert and Roeser 2020). Therefore, this unnecessary sway of emotions can make users desensitised to their emotions and moral intuitions. In the long term, it could lead some to become indifferent about a topic to avoid the disappointment of a roller-coaster of emotions. Consequently, this self-imposed indifference may create a kind of moral desensitising to the very issue at stake. This would be undesirable since we would like people to be morally sensitive to the issues that matter to them.

Social networking platforms (such as Twitter, Facebook, Instagram, Reddit, LinkedIn, Snap Chat, YouTube, etc.) facilitate daily the occurrence of minor and significant moral harms. Major moral harms happening online are hate speech, bullying, trolling, doxing, etc. Yet there are also some other harms challenging to categorise as major or minor: manipulation of others via truncating information or sharing misinformation, hateful speech disguised as inspirational (such as fit-spiration, which may be shaming specific body shapes), passive-aggressive speech, reporting innocent users to get them banned for expressing divergent opinions, trolling, etc. These harms are not exclusive to the online realm. Yet, social media platforms make it easier for regular users to have power over the unknown and remote people to a degree unimaginable in the pre-Internet times. Some people would never tell others to kill themselves in real life, yet they have no qualms about leaving such a comment to someone they have never met face-to-face. While many of these moral harms are intended to be hurtful in the case of users tactically deploying anonymity as a shield from moral or legal consequences, the issue that I focus on here lies in the grey zone of the unintended unacknowledged moral harms. This is one of the most salient features of the moral landscape online: people harm each other online without wanting to do so and not realising that they caused harm.[2] The ease of harming at a distance through online social platforms demands that we re-evaluate the principles of social media ethics to include these unintended and unseen harms. Going beyond classical moral theories about what people owe to one another, such an approach would need to look closer at the role played by online environments in shaping such behaviours. This approach would also need to point out the distinctive features of social environments online, such as the mediated and disembodied nature of the online interactions among users, the distance and time delays which introduce additional difficulties for the moral agents interacting online.

In this paper, I am concerned with establishing some basic principles for the design of online social environments aiming to minimise *the unintended harms* done by users to one another while interacting online. Such normative principles would need to consider the particulars of online social environments and how these shape interactions. I will argue that an ethics for user interactions on social media is a relational ethics and an ethics of mediation at the same time. I will look towards enactivism to help ground the basic principles for human interaction online. Enactivism, to be explained in the next section, is an approach to human cognition proposing that "at least in basic (perception- and action-related) cases, cognitive processes are not just in the head, but involve bodily and environmental factors" (Gallagher 2017, p. 1). Enactivism seems a promising approach for building moral principles since it focuses on the agent's interactions and the environment created through these interactions simultaneously.

The article is structured as follows. In the first section, I give examples of how unintended harms occur between social media users and how these harms can be traced to features of the users' interactions as mediated by online platforms; I conceptualise these unintended harms as a blunting of users' moral sensitivity. The second section shows how the concepts of participative sense-making and ecological niche construction provide new ways to deal with these specific online challenges. In the third section, I operationalise these concepts and advance three ethical principles for designing online environments that enhance users' awareness of the moral load of their interactions.

---

[2] In their book, *Evil Online*, Dean Cocking and Jeroen van den Hoven call this feature a *moral fog*, a term describing "conditions where normative competence is especially challenged" (Cocking and van den Hoven 2018 p. 86) and point out how evil-doers online often seem to be oblivious to the harm done or to their duties.

2 Moral Sensitivity and User Interactions on Social Media

Social media[3] is by no means a separate realm of experiences, rather a continuation of the users' regular lives mediated in a particular way (Isin and Ruppert 2015). What we do on social media often extends our offline lives: activism, pursuing a passion, gathering a community around a topic of concern, performing identity in front of others, etc. Many of the habits of online interaction were learned offline, such as manners of addressing others or knowing what to expect in a conversation, rules of engagement etc. Yet not all that happens on social media platforms can be attributed solely to the individual users' responsibility. Social networking platforms are interesting to study from an applied ethics perspective since these platforms facilitate certain behaviours through user interfaces designed with various digital affordances. Since various design factors can trigger desired behaviours, the interesting question for an ethicist is what to aim for when designing user interfaces. As mentioned in the beginning, I will look here specifically at behaviours whereby the user harms others without an intent to do so and is oblivious to the harms caused; such behaviours should count as failures of moral sensitivity.

Classical accounts of moral sensitivity depict it as the ability of a person to grasp that the situation at hand is of a moral nature and how this affects others (Jordan 2007, p. 324). Based on Rest's (1992) work on moral behaviour, moral psychologists have developed multi-grained models of moral sensitivity, breaking it into several dimensions such as: "interpreting others' reactions and feelings…, having empathy and role-taking ability…, understanding how one's actions can affect the welfare and expectations of both oneself and others…, and making inferences from others' behavior and responding appropriately to their reactions" (Jordan 2007, p. 326). In the case of the online interactions, I conceive moral sensitivity loss as having two dimensions: in the first instance, the user fails to grasp that her behaviour is not morally neutral—which does not necessarily mean that one fails to see only the harmful nature, the user may fail to register the praiseworthy behaviour as well; secondly, the user fails to acknowledge one's responsibility for it, by denying causality and invoking other circumstances to exonerate. In the existing literature, moral sensitivity is understood primarily as a relational capability, drawing on the user's existing capacities for emotional awareness (about one's states and of the others). However, this paper wants to draw attention to how moral sensitivity relies upon an environment facilitating emotional awareness. Such an environment would need to support the agent's moral sensitivity capabilities by having built-in features for unambiguous communication of emotions and shared sense-making.

Current online social environments pose specific challenges to relationality and affective expression that have consequences for hindering emotional awareness and, through it, blunting the user's moral sensitivity. There are several salient features of online user-to-user interactions that seem conducive to the hindering of moral sensitivity: the mediated nature of the interactions (we perceive others through the digital traces they leave online such as texts, images, videos, sounds), the bracketing of the user's bodies (we do not have perceptual access to the other's embodied presence), and the asynchronous communication (the initiating and reception of a message happen at different times, separated by unpredictable time intervals). If we were to limit ourselves to these three features, social media platforms would be nothing more than the digital equivalent of a postal box, a space "of relations between and among bodies acting through the Internet" (Isin and Ruppert 2015, p. 11). Yet, one additional feature makes social media interactions distinct from other long-distance interactions: the audience of a message is uncertain and unpredictable. Social media platforms have a built-in potential for turning any individual message into a broadcast, reaching innumerable receivers. This was the case with the impeachment tweet mentioned earlier, whereby a user with 200 constant followers reached 750.000 people. When we post something online, we cannot control the post's visibility unless we restrict it to certain friends; but even then, our friends may choose to share it further and thus give it more exposure than we initially intended.

The three features of interacting at a distance blunt moral sensitivity by hindering emotional awareness of the other's reactions, while the uncertainty of the audience size makes it difficult for users to assume moral responsibility for the effects of their actions. I will illustrate how these features work against our capabilities for moral sensitivity with a few examples. The asynchronous engagement creates a time delay between the emitting of the message and its reception by others. Often, we feel the urge to post something because we experience a strong emotion in the moment, yet the time delay between the responses makes it that the emotion has subsided when we get a reaction to our post. We may be out of sync with our previous emotions and the responses' emotional charge. The spatial distance and disembodied communication pose additional challenges. Even when a real-time chat allows for real-time online conversations, the exchange of messages may not fully engage our attention because of competing demands and interruptions from our

---

[3] Social media is the communicative online space constituted by various platforms for social networking such as Facebook, Twitter, Instagram, YouTube, etc. These platforms have two defining features: the content is posted by users, and users have unique identifiers (a nickname or their real nem) making it easy to follow, befriend them or simply distinguish their content from those posted by others (Boyd and Ellison 2007).

physical environment. The etiquette rules construe it as rudeness when people are preoccupied with their phones while others want to talk to them (Cahir and Lloyd 2015), yet it is unclear how much attention one should give to the remote interlocutors. When our attention is divided between the present situation and what is going on in an online chat, we are less likely to pick up on emotional cues, as we tend to read the lines of conversation hastily. In addition to having fewer clues about the interlocutor's mood without access to facial expressions or tone of voice, our hastiness causes frequent misunderstandings. I never see my friend's lips utter the words; I only see words on a screen. I read words instead of hearing them; I see images where perhaps I should be free to imagine myself the context; I interpret words and emojis as indications of affective states in the absence of the others' gestures or tone of voice. As online users, we are constantly interpreting from a limited set of clues and trying to get some feeling of the other behind the signs projected on digital screens, while most of the embodied gestures are invisible as we interact with the digital outcome of our gestures.

The unpredictability of the audience size means that users cannot estimate who their audience is and hence cannot tailor their speech to it, making the classical media-theoretical distinction between broadcast media and point-to-point communication (Watson 2008) irrelevant. Broadcast media such as radio or television target large audiences hence the messages are shaped impersonally. End-to-end communications such as letters, emails, instant messages are meant from one person to another, with the message tailored for that specific recipient. An unpredictable mix of both forms of communication occurs daily in user interactions on social media. I have shown elsewhere how the usage of social media mimics gossiping in the way it establishes networks of trust (Marin 2021). In addition to this, social media platforms are used to spread gossip, with commentators working as relays effectively. While evolutionary psychologists construe the role of gossip in making networks more reliable (Backer et al. 2016), from a moral standpoint, gossip is primarily harmful because it excludes the subject of gossip who cannot answer to the hurtful claims in an effective manner. While some work done on gossip in social epistemology has shown that gossip can facilitate overcoming epistemic injustice by giving a voice to those who are systematically oppressed and do not have access to authoritative communication (Alfano and Robinson 2017). However, because of the inherent asymmetry in gossip, the one spoken about cannot defend oneself and repair the damage, hence gossip can also create further epistemic injustices, depending on who does the gossiping. With social media, the morally problematic effects are aggravated because gossip can become visible to unintended audiences.

Given this unpredictability of the audience, some users simply assume that they broadcast all the time, while others assume that they speak to a select group of followers and nobody else can see their posts. Both strategies can backfire in various ways, but the latter approach is the most harmful since it ignores the affordances for propagating a message to many users.

3 Sense-Making in Online Interactions

Considering how interactions appear on social media as disembodied, asynchronous, and mediated, an enactive framework provides a useful vocabulary to characterise what happens in user interactions when misunderstandings occur. Given that the moral issue at stake is harming others without intent, this seems primarily a matter of misunderstandings between moral agents amplified by a blunting of one's moral sensitivity. Hence, we need to look first at how moral agents understand each other in day-to-day interactions; for this purpose, the enactive notion of sense-making has a lot to offer. Enactive cognition emerged in the 1990s as a theoretical framework for explaining human cognition. Enactivism distanced itself from classical representational views of the mind by proposing a dynamic systems approach that frames human cognition as the phenomenon happening when an agent makes sense of the world in an embodied way (Varela 1997). Enactive sense-making reframes cognition as making sense of a particular situation while interacting with it:

> Natural cognitive systems are simply not in the business of accessing their world in order to build accurate pictures of it. They actively participate in the generation of meaning in what matters to them; they enact a world. (Jaegher and Di Paolo 2007, p. 488)

Enactivist thinkers stress the importance of interactions that happen at pre-reflective levels, as we engage with affordances in our environment without representing or being necessarily aware of our intentions, desires or beliefs. Thus, when I enter a situation, I do not perceive it by analysing it from the outside, making a mental model in my head, and then acting; instead, I respond to the affordances presented to me by a situation by appropriately reacting to them. For example, the knowledge of how to open a door is not a mental model in my head of how door handles work, but it is in the embodied way in which I interact with door handles. If I was hard-pressed, I could generate a model and explain how doors work and how I can open them. Still, this symbolic knowledge that comes after the fact is not necessary for me to know how to open doors or orient myself in the world of affordances. Many activities require us to do symbolic thinking, to manipulate words and symbols without engaging in embodied interactions, yet the enactivist insight is that, for the social part of our interactions, we operate primarily at this pre-reflective embodied level. Hence this level needs to

be transposed fluently online if we want to continue to make sense of other people's intentions and inhabit the same social space as them.

Sense-making designates how an organism adapts to its environment and makes it part of one's form of life[4] (Di Paolo et al. 2018, p. 33). Sense-making in social interactions is *participatory sense-making*, designating "the situation in which the sense-making of two or more agents is mutually modulated as they engage in an interactive encounter" (Di Paolo et al. 2018, p. 139). A typical example of participatory sense-making is that of two people who meet on a narrow path and try to avoid each other by noticing what the other does, and they move their bodies in such a manner as to avoid collision while the other moves their body as well (Jaegher 2009). They both make sense of the situation and navigate skilfully through it, without any need to revert to mental representations of how the other thinks and what the other wants. In this situation, cognition is embedded in synchronised moves which work together towards a common aim—that of not bumping into each other—which both actors realise together. As Hanne de Jaegher put it, "taking interaction seriously involves focusing on how participants engage with one another, not on how each figures out the other" (Jaegher 2009, p. 535). In other words, cognition is in the skilled interaction and is realised in the encounter. However, there are no digital affordances for the encounter between users, the affordances designed on social media platforms are intended for the individual users seen as separate individuals. Currently, the unit of analysis in the designed interactions on social platforms are individual users' actions, not their interactions.

3.1 Missed Paths for Sense-Making on Social Media

On social media, we do not interact directly with others as bodies, instead, we engage digitally with each other's traces; I will call this phenomenon a *digital interaction* to distinguish it from face-to-face interactions. Digital interactions do not exclude embodied interactions since our categories of thought are already embodied, even when we interact online.[5] The language used to describe what we do online includes spatial references (up–down-left–right), bodily metaphors, and other body-related categories. Wittgenstein famously pointed out that "if a lion could talk, we could not

understand it" (Wittgenstein 2010) because presumably the lion's experience is not predicated in the same categories as we do, because of a radical difference in how we participate in the world through lion and human bodies. Understanding the unseen users online is grounds for the assumption that they are like us, embodied human beings experiencing the world through roughly the same lens as us. Hence the absence of having access to other's bodies while communicating online is not a problem for situated cognition *prima facie*, yet it becomes a problem when we try to construct shared meanings in the online realm.

Prima facie, it seems that participatory sense-making is occurring online since users follow tacit rules and communicate successfully, they create communities where linguistic meanings are shared, and furthermore, many users even learn to interact in new ways from other social media users. However, the particular ways users create meanings together while online pose difficulties for exercising one's moral sensitivity. Typical of online interactions, the case of unintended harms deserves further scrutiny. Of course, I can cause unintended harm to another also in the offline realm, when I step on someone's foot or when I say something nasty about someone and they are in hearing range. Yet in offline interactions, these unintended harms are easily identifiable as harms done to another. When I can see the pain on their face when I stepped on the foot, or when I see the anger on their faces if they heard me, then I have no doubts that I am the cause on another being's discontent. This is not to say that offline interactions are always fostering moral sensitivity since there are various degrees to which humans are receptive to others' mental states, and deciphering them is not always a straightforward task. Some people have moral insensitivities triggered by their social role: multiple case studies about engineering ethics concern failures of engineers to grasp the moral nature of the situation at hand because they do not look past their explicit professional responsibilities (Taebi 2021). Moral insensitivity is not unique to the online environment, but it is favoured by how interactions are fostered through the designed affordances.

If we look at the offline ("in real life") cases of moral participatory sense-making, we will notice that several features help raise awareness about possible harmful effects on others. In real life, the feedback we get from others is usually instantaneous, affective and directed at us as individuals. Suppose I gossip about a friend, and the friend happens to be around to hear it. In that case, I will see on her face an affective expression that allows me to discern that she has been hurt: this is direct feedback, caused by my gossip, expressed in no equivocal way, and I am thereby 'faced with' my responsibility for it. However, the moral agents need not be in the same space and time for their developing sensitivity to the harm they caused. Let us say I gossip about the same

---

[4] "Sense-making is the capacity of an autonomous system to adaptively regulate its operation and its relation to the environment depending on the virtual consequences for its own viability as a form of life. Being a sense-maker implies an ongoing (often imperfect and variable) tuning to the world and a readiness for action." (Di Paolo et al. 2018, p. 33).

[5] I am thankful to an anonymous reviewer for pointing out this distinction which I have further elaborated on here.

friend, but I know for a fact she is not around, yet a third person relays to her my gossip at a later time; after a while, I hear from someone else she is upset about what I said. While I do not have immediate access to her emotional states, I can infer why she was upset and trace back my responsibility to my initial utterance. This inference lies in my having a reliable input concerning the friend's affective states (an acquaintance I can trust) and because I am aware of the general rule that gossiping is hurtful when found out. By engaging in online gossip, I assumed the risk to be found out, but I didn't intend any harm to my friend, hoping that she would not find out. The unintended harm happened in this offline situation, yet I was aware of the harm once the gossip got out because I had access to the victim's affective states (albeit indirectly) and because my relation to the victim was clear (she was a friend that I had betrayed, our relation entailed duties which I infringed). Thus, the conditions needed for moral sensitivity, namely relationality, appropriate interactions and affective response, were fulfilled (Rest 1992).

Let us see next how the same conditions are thwarted in an online environment. Looking at gossip as an example, particularly at open gossip—a phenomenon typical for social media. Open gossip online entails talking negatively about a third party openly, via public channels (Facebook, Twitter, YouTube) and even tagging that person so that we make sure she will find out. When the targeted party does respond, her response is not enough to counteract the initial harm: the unpredictable visibility of who gets to see the original gossip and the response makes it that the two utterances travel online disconnected, with little influence on each other. Even when the initial gossipy post is deleted, other users can always make a screenshot and propagate it further on their own. Zeroing in on the moral implications of the open gossip online, it seems that the users involved in starting it and propagating it are missing specific barriers that would have made this behaviour less likely offline. Crockett (2017) observed that, because the possibility to encounter one's offenders in real life is minimal, users tend to engage in public displays of moral outrage online to the extent that would be unimaginable offline: "A paradigmatic example is the case of Justine Sacco, a woman who tweeted a comment about AIDS in Africa that many perceived to be racist. Within hours, she became the top trending topic on Twitter as millions of strangers around the world piled on the shaming bandwagon" (Crockett 2017). This has the unfortunate consequence that, even when affect is clearly expressed online, through words or emojis, it is hard to discern if the person expressing the affect is genuinely feeling it (Crockett 2017; Marin and Roeser 2020). Therefore, the condition of appropriate affective response is undermined; furthermore, the relationality condition for moral sensitivity is also relatively weak with open gossip, since people will express negative judgements about people they have

never met and with which they cannot enter into any direct relation.

When sharing misinformation to one's network of friends, a similar and often involuntary harm occurs. While the harmful effects of misinformation have been construed in terms of their societal damage, such as destabilising democracy and subverting the trust in public institutions, there is a distinctive harmful effect of misinformation sharing for the individual users; this effect can be construed as a failure in participative sense-making online. The misinforming effect of the impeachment tweet mentioned in the beginning can be depicted as an incident with cascading effects as a series of unfortunate yet systematic misunderstandings between all the users involved. When B. C. tweeted something that made him feel good, his intention was not to inform others—for this, he would have needed to do more research on the matter at hand, and he was not an expert. Instead, he expressed an affect in words and probably felt the need for this expression to be visible to others. Many misinformation streams online start with information either charged explicitly with emotions in the original text or shared by users with an emotional intent (Bakir and McStay 2018, p. 159). The impeachment tweet used here as an example is not qualitatively different from a genuinely informational tweet: both are utterances that other users take up to express some emotion about the news content. One could argue that the impeachment tweet is an example of successful participatory meaning-making: the community inflated its significance and made it visible; hence, a shared meaning emerged through the disparate acts of sharing and linking the tweet.[6] Yet, the meaning emerging from the storm of online interactions was incongruent with the *prima facie* meaning of the Tweet. The tweet's meaning was "if Trump is impeached, he will lose the presidential privileges", yet the shared meaning-making that emerged was more similar to "We would like Trump to be impeached, he deserves to lose his privileges". The difference between these two meanings is between a descriptive and a normative proposition. The descriptive one is false, yet people want it to be true to fulfil the normative one. The emerging sense-making is wishful thinking, although it may look like an agreement about facts—and this is misleading.

When users try to make sense of the information found on social networks, they are tasked with inferring from digital traces the emotions and meanings of others. The inferred meanings are not necessarily co-constructed but rather projections of what each user thinks other users may be thinking, based on their having first learned the language and some social norms of interactions, basically on habitus and patterns of linguistic interactions (Di Paolo et al. 2018, p.

---

6 I am grateful to an anonymous reviewer for making this observation.

193). This in itself should not be a problem since this is how language works—we do not learn to speak a new language with every new interlocutor, rather there is a negotiation between the old learned patterns of linguistic interactions and the newness introduced by the interlocutor, such as hijacking meanings or subverting them through irony. While the online space works for re-enacting linguistic patterns from the offline world, the potential for creating new meanings and social practices is hindered because we do not have access to the other users as face-to-face encounters. Imagine someone saying, "In my next utterance, I will be ironic". The utterance would be strange since this is not how irony works, making it explicit destroys its charm. What makes irony worthwhile socially is the risk that others may not get it and the pleasure resulting from the recognition –in the cases when they do get it. The irony is co-constructed by the participants in a dialogue, and without a receptive audience, the speaker has a hard time making an ironic statement.

Because of these limitations of online communications, most social media posts have a subtext that is difficult to read and are prone to misunderstandings. Even when someone answers my tweets with a comment, appearing to enter into a dialogue, they are not co-creating meanings with me but rather expressing post-factum interpretations. In the case of unintended harms caused online, the problem of perceiving the other's hurt expression is compounded by several issues: first, a disbelief in the face value of someone's utterances online. If someone is hurt, we are used to perceiving this by reading their facial expression, hearing it in their voices, or from their bodily posture; we can infer it from someone stating, "I am hurt", but this is not our usual way of finding out about harm.[7] Secondly, we cannot be sure that we caused someone else's harm since we do not have access to their immediate environment. We can always say it was something else that triggered them, they misunderstood us, they had issues, we didn't mean it like that. One of the most common online accusations among users is being "too sensitive" or "a drama queen". Thus, even if a user points at us as the leading cause of harm, we can still say it was a problem with how they interpreted it. And while this excuse also floats in the offline space, it is easier to use online since we are always interpreting other's utterances mediated by digital signs while our access to bodily cues is cut off completely; our interpretative work is quite exhausting online, continuous, and we are prone to make mistakes of interpretation more often there. Thus, the problem of participative meaning-making online appears as a problem whereby a user can always plausibly reject the meanings brought forth by another user. If we only engage with what we approve of from the other's utterances, hence if we pick and choose what to ignore, we are in danger of ignoring the interpretations of the others, their way of construing moral facts, and ultimately of dismissing their moral agency.

However, digital mediation does not pose an insurmountable challenge to experiencing moral sensitivity on social media. While enactivism does help conceptualise what is lacking from the online encounters, this lack is not the end of it all. On the contrary, if we take some inspiration from the philosophy of technology, mediation is a constant effect of technology usage. In particular, every technology aimed at communication tries to become as transparent as possible so that we forget that it is even there—with limited success since there is no such thing as a technology that does not alter the message it transmits (Krämer 2015, p. 27). However, as humans adapt themselves to the daily usage of a specific technology, they become oblivious to its mediation. The shock created by new technologies is partly due to the cognitive habits created around the old technology now obsolete. Thus, the printing press caused a shock because people were used to handwriting; radio and television shocked those used to newspapers and books, and now the Internet is still challenging our old habits acquired from the age of mass media. The success with which a technology manages to create new habits that we then later take for granted cannot be overstated: we tend to take as natural human behaviour what is highly mediated and adapted to technology. The Internet and social media communications did not replace face to face interactions, instead it replaced other mediated forms of communication such as letters, newspapers, phone calls. To compare face-to-face interactions with mediated ones (be those analogically or digitally mediated) is somewhat unfair. In this comparison, any technology will lose to the genuine encounters that are not mediated because patterns of interaction are easier to recognise, and new patterns can be forged in the shared space of instant face-to-face interactions. Provided that we embrace mediation and look beyond it to the quality of our encounters, enactivism offers a meaningful way to think about our mediated encounters by asking us to look at the environment created by our digital encounters as a whole. This will be the topic of the next section.

3.2 The Ecology of Norms in Online Environments

This section will highlight the interconnected nature of our online interactions and how an ecological approach to the online environment can help us flesh out some of the ethical problems arising. This approach was inspired by Luciano Floridi's concept of the digital environment as a sub-set of the overall informational environment or the infosphere (Floridi 2008). An ecological approach entails, as Floridi

---

[7] There are also cultural and race biases when perceiving other's pain or inferring any other emotions (van Grunsven 2020), we can be dismissive of other's expressions of pain simply because we do not belong to the same community and are not used to their modes of signifiying affect.

describes it, going beyond the agent-oriented classical ethics (i.e. evaluating the morality of the agents as individuals or of their atomic actions) and focusing, in addition, on the effects that these actions have on the digital environment itself. In an ecological vein, we are all connected through the medium that we generate through our actions, we not only consume existing information (found online), but through our actions, we generate and change this information. Hence our actions with information online will shape the environment and, through it, change how other users experience it. For example, when searching on Google for a current event, I can choose to engage with whatever appears as the first result or to click on the click-bait gossipy looking titles, or I can choose to scroll through results until I find what looks like a neutral source of information. My choice of engaging with the search results is not affecting only my informational environment, but also it is a signal for the algorithm that some results should go higher—since I engaged with them. Granted, my effect on the results of a search algorithm is negligible, yet the aggregated actions of thousands of users can, in effect, change what results appear to the top of the list. Floridi's focus was the informational agents as" connected informational organisms" (Floridi 2007) or "inforgs" that were working towards producing and disseminating information in the infosphere. I want to extend this by examining how inhabiting this online environment affects us as moral agents, hence our ecological interconnectedness as moral agents.

Enactivism is already an ecological approach, as it rejects the "fundamental assumption … that the individual agent is the correct unit of analysis for understanding mind, cognition and behaviour" (Froese and Di Paolo 2010, p. 43, cited in Herschbach 2012, p. 468). In distinction from Floridi's ecological ethics, the enactive ethics emphasises the activity of niche construction in which we are all dwelling: the environments in which we move are not passive set-ups that we find already constructed out there. Instead, we create the environment through our interactions, since the environment is perceived only in relation to us since what we do is "'laying down a path in walking,' i.e., of actions that change the world in ways that feed and structure those very actions, either now or at some future time" (Halpin et al. 2010, p. 2). This also holds for the online environment created through our messages, comments and reactions.

The online environment poses a specific problem for the ecology of norms and rules: these are not quite as graspable as the norms and rules in real life. A rule becomes actual once it fulfils specific criteria: "(i) regularity of behavior, (ii) a standard of criticism, (iii) a tendency to criticize for violations of the rule, and (iv) felt bindingness of the rule" (Hufendiek 2017, p. 4467). As Hufendiek explains it, all rules are "visible in social contexts in the form of re-occurring patterns of behavior and in people's sanctioning behavior when a rule is not followed"

(Hufendiek 2017, p. 4467). This is the case for social interactions happening in real life, where we can easily see if a rule is broken through the social feedback given by others. However, the behaviour patterns are not as easily discernible in the online environment and can be misinterpreted at will. For example, the movement Black Lives Matter (BLM) had several associated hashtags on social media and emojis carrying specific meanings for those endorsing the movement and those opposing it (Alfano et al. 2021). The meaning of using a purple heart emoji, for example, was known to the members of the BLM community, but it was still possible to have it used differently by an outsider who didn't know the rules of the game. Of course, after noticing for some time the patterns of use, one could deduct in what circumstances it is appropriate to use it or not, without needing to read an explicit rule— if enough members of the community used it correctly. Yet, because online communities are fluid and anyone can enter and contribute to them, the sanctions for misunderstanding the rules or incorrect rule-following are weak. In a worst-case scenario, a non-compliant member could be kicked out of the community (although this is hard to do given that the rule to be followed was not made explicit), but this would not stop them from misusing the rule in other communities. This points to a general problem with social media: the sanctions are weak and hard to enforce, making it easy for offenders to ignore social feedback. It is an environment where the normative force of the other's feedback has less weight, and it can lead to some users getting to pick and choose what rules to follow. In this context, repeatedly causing harm to others, with no real consequences or without raising concern for the offenders, becomes a frequent interaction pattern.

From an ecological enactive perspective, we constantly engage in *critical niche construction*, namely building ecological niches in our surroundings to live and flourish. In this respect, an enactive ethics will foreground the need to "live and let others live" while keeping in mind that, when online, we are not merely acting towards our self-interest, but that our actions affect others directly and indirectly via the environment. However, our moral sensitivity is not habituated to detect such cases where unseen others are affected and where an environment becomes toxic because of our actions. This requires that we supplement existing frameworks of digital ethics with several principles inspired by an enactive account to deal specifically with these moral challenges.

4 Three Enactive Ethical Principles for Social Media Interactions

In proposing some ethical principles for social media interactions, I am inspired by the work of Colombetti and Torrance (2009), who already showed that an enactive ethics is not competing with the other ethical frameworks

but complements them by drawing attention to two main aspects: the interaction between agents as an object of concern, and the affective component of all ethical processes of meaning-making. I am staying close to their insight that we need to overcome a methodological individualism that focuses on agents and their actions, what (Colombetti and Torrance 2009, p. 517) call "the individuocentric nature" of ethical inquiry, and instead focus on the interactions. I am adding to this approach the ecological concern for the environments created by such interactions and the specific focus on unintended harms on social media. As previously mentioned, we can harm others with intent when we interact online and offline, but the challenges to our moral sensitivity, specifically how we are unaware of harming others, deserve to be tackled through some different ethical principles. I propose below three guiding principles for an ethics of social media usage to make it possible for users to exercise their moral sensitivity online. The first two principles are relational, concerning what users owe each other, and the third principle is ecological, concerning what users owe to their digital environment. These principles are not exhaustive, and other duties and responsibilities enter into play when interacting on social media, but I see these as complementing existing ethical frameworks such as Floridi's digital ethics (2013) and value-sensitive design in ICT (van den Hoven 2007; Huldtgren 2015).

4.1 The Genuine Encounter Principle

Always look for the face of the other

This principle is inspired by van Grunsven's (2018) take on enactive ethics. Her focus on second-person embodied interactions leads to an account of moral obligation to the other as the obligation to create and respond to genuine interactions: "To be engaged in genuine interaction is to see the other, at least during the interaction itself, as someone to whom I am in some sense obligated to respond, and thus to know myself as someone who can take responsibility." (van Grunsven 2018, p. 154). Answering Colombetti and Torrance (2009) and enriching their account, van Grunsven (2018) argues that rejecting methodological individualism about responsibility does not entail rejecting moral responsibility at all, even individual ascriptions of it. We are still responsible for the relations we enter into and for our responsiveness to others. She depicts responsibility "as a stance or role that we as relational subjects can succeed or fail to inhabit with respect to the lives of others with whom we are bound up" (van Grunsven 2018, p. 154). This kind of ethical stance is inspired by feminist care ethics in which the unit of analysis is the quality of relations into which we enter with others. In addition to care ethics, van Grunsven's original point is that the second-person interactions create

autonomy at the same time as responsibility for the other, for the second-person relations in which we enter. As di Paolo has put it, since "an enactive approach places great importance on the autonomy of the individuals involved [in interaction], this approach to social cognition, while focusing on the interaction process, paradoxically also gives social agents an autonomy and role that has not been thematized before: that of participation in contrast to mere observation" (Di Paolo et al. 2010, p. 72).[8] This emphasis on participation puts the onus of responsibility on the social media user. The users are reconsidered from an enactive ethics perspective from merely passive consumers of content to fully responsible moral agents that understand that the digital traces they leave in social spaces are potentially the basis of relations with an unseen other. This first principle emphasises that, even when we do not see others face to face when interacting online, we are responsible as active moral agents for the kinds of relations we create through our digital actions.

This principle entails the constant asking of oneself: Are we attentive and responsive enough to others to acknowledge them as autonomous agents while interacting with them? This requirement may seem impossible for social media, where interactions are asynchronous and at a distance. However, this is where the first principle comes into play. In looking for the face of the other in every interaction, we are not necessarily looking for their physical expression but their symbolic expression of themselves. Online, we should treat the other as someone who can see our public messages and respond to them, hence never speaking about others, only with them directly. Here is an example: people speak of celebrities and public figures as if they are not present online. I may Tweet an angry remark about Donald Trump and hashtag it appropriately with his name, but, in doing so, I am not looking for any genuine encounter since I speak about Trump as if he is not present, as if he cannot answer. With the global interconnectedness afforded by the Internet, we should see each other as all being in the room simultaneously. In this example, I treat Trump as an object because I am not speaking with him but above him or about him (hence avoiding the second-person interaction crucial for genuine encounters). The more public a person is, the more we find it acceptable to speak about them and never to them, and the more we objectify them. Yet by following this first principle, we need to seek a genuine encounter with another by systematically avoiding speaking about another, only directly addressing them, as if we had their face in front of us.

4.2 The Handshake Principle

No user interaction is complete without a user response and an acknowledgement of this response by the initiator.

In the network engineering jargon, the transport layer security design[9] stipulates how two network devices are supposed to establish a secure session for their communication; the guiding principle for this layer is the handshake protocol. Without going into technical details, the gist of the protocol is to repeatedly send messages to the other device and then wait for them to respond and then answer to acknowledge their response. This three handshake protocol ensures that no misunderstanding can occur: every device knows for sure that the other device received its message. While this is optimal for network devices, misunderstandings should be expected in human communication, even in the face-to-face realm. In order to account for this shortcoming, the handshake principle stipulates that we at least repeat what the other has said and wait for the other to confirm or deny that. This principle establishes that one-way communication is never complete communication. When we post or comment online, we produce utterances aimed at others, but are we communicating or merely expressing ourselves? Nothing should be seen as communication in the absence of the receiver's feedback. For this to happen, we need to wait for the other's response and then show that we received it and how we interpret it. Not all social media platforms allow for this dual way feedback to happen. In the case of Twitter, every tweet is a small broadcast to a broad audience of unseen others, hence communication from one to another is never the intent. The B. C. tweet became viral because its initiator had no idea how others would receive it, and he got feedback when it was too late, and the tweet had gone viral. If the handshake principle were to be implemented on Twitter, no tweet would become viral or be featured as a trending topic without its initiator being made aware that this is happening. A design implementation of this principle could limit the number of retweets to 10 (or any other small number) after which the tweet's author would get a message stating that people are engaging with one's tweet and asking the author to read what others have said about it, and giving them the chance to modify it in any way. This principle would account for the unpredictability of the audience on social media platforms.

Many of the moral harms happening online are linked to impersonal communication: we post sentences addressed at nobody in particular, and then we wonder why we are

misunderstood. Without a particular receiver, our online posts and shares are the digital equivalent of someone speaking to oneself aloud on the street: it may be about self-expression, but it sure is not about communication nor common sense-making. In conjunction with the first one, the handshake principle ensures that no communication on social media has a generic, impersonal intent. Even something as generic as a tweet has some imagined audience aimed at by the tweet's author; by ensuring that an actual response from this audience is taken into account, the tweet's impersonal language will be adjusted to the context of understanding of a specific audience, thus minimising misunderstandings.

4.3 The Shared Symbolic Space Principle

we all depend on and are vulnerable to each other's actions because the online environment is the sum of all our actions

The third principle addresses what we owe to the digital online environment that we find on social media. Most natural environments are composed of niches constructed by the agent's interaction with the surroundings. One classic example in the enactive literature is that of the bacteria swimming in a sugary solution: the medium can be said to be sugary even before there were bacteria in it, but only the interaction of the bacteria gives the glucose its significance as a 'growth medium', rather than merely 'containing sugar'. As Varela put it, without the bacteria in there, the sugary environment would be just another chemical property, no more remarkable than its other properties: "Remove the bacteria as a unit, and all correlations between gradients and hydrodynamic properties become environmental chemical laws, evident to us as observers but devoid of any special significance" (Varela 1997, p. 79). For humans, environment construction occurs in the ways we interact with objects in our surroundings; for example, a patch of grass becomes a path after many people cross it repeatedly. However, users do not have that much power in carving out paths and roads to their heart's desire on social media platforms. The digital online environment is more resistant to shaping than the natural ones.

Teams of software engineers have designed our online environments. Such designed interfaces entail creating affordances for users to perform various actions such as clicking links or buttons, filling in text in boxes, uploading images, refreshing or scrolling pages, etc. The limited and pre-established list of possible actions does not mean that our online environment is immutable or indifferent to our actions. Even if we cannot directly modify the user interface's design, our actions create content and express rules of interactions visible to others. When members of a group interact, they carve

---

out an ecological niche which is discernible by the stable ways in which the members use the affordances to interact (Rietveld and Kiverstein 2014). This means that, whenever we are in a group, we should see ourselves as responsible for the moral climates created by our actions because of the inherent visibility of our actions. Each time we endorse or disapprove certain behaviours, we are setting an example whether we want it or not, we create paths that others may follow. The same holds for our behaviours in online communities: we are responsible for what we say, what we do and how we do it. The novelty of the enactive perspective here is that the responsibility is primarily ecological, concerning the kinds of environments we contribute to and make visible to others. As Arfini and colleagues put it, "online networks … are designed more to strengthen social connections rather than distribute accurate information" (Arfini et al. 2018, p. 379). This means that it is often the case that we inadvertently propagate our ignorance online instead of valuable information and that our cognitive niches online may be niches of ignorant cognition (Arfini 2019). In the case of online cognitive niches, we need to distinguish between niches that are critically open towards enriching the overall cognitive environment and mere echo chambers favouring only those assenting voices and conformist worldviews. A user enclosed in such a cognitive niche can aim to critically expand it responsibly by acknowledging that one's sense-making is unavoidably local and linked to other niche members and not destroying other cognitive niches. This is the gist of the third principle: every user should strive to enrich one's cognitive niche online while, at the same time, not destroying other niches.

Guided by these three principles, we can now better understand the moral harm of the impeachment tweet. First, we see that B. C. was not talking to anyone in particular when tweeting, he was emitting an utterance in the void, broadcasting without a clear audience. The huge amount of shares and likes he got could be considered an endorsement, yet B. C. did not try to understand what others got from this tweet because he did not check for shared meanings, nor did he attempt sense-making with his followers. Following the first two principles, the tweet was not an interaction with anyone, by any measure, rather an expression of affect thrown out there into the void. Looking at the third principle, we get a fuller picture of why B. C. should not to be blamed for what followed. Twitter is a medium that encourages quick tweets, gut reactions are aimed at no one in particular. The impeachment tweet aligned with how most users interact on Twitter: one-directional tweets, similar to soliloquies, not expecting any meaningful answers or critical replies. B. C. was following the norms of regular Twitter usage pre-established by millions of users before him, following the same paths carved out by others. However, if we understand that we are contributing to an environment of disconnected encounters with every action, we can start thinking more about the effects of our online actions. If every user would reconsider the paths they are making through their tweets and the kind of environments generated by their actions, then the tweet-sphere would slowly change as new modes of behaviour (such as slow reactions and reflective tweets) would become acceptable and ultimately mainstream.

We have developed digital habits of skilled information manipulating alongside our embodied day-to-day habits. Just as a professional gamer presses the buttons of the gaming console or artists take up a paintbrush without any deliberate reflection, one can also become a social media master. Such a master would skillfully like, comment, post and share with the intent of finding and sharing the information that can enrich others and, importantly, while responding to others' reactions and engaging with them in meaningful exchanges, trying to construct a shared understanding of the meaning at stake. Thus, while we do post content in a way that seems individualistic (mainly broadcasting our message to a world of unseen listeners), once we get to respond to others' responses to our posts, a possibility of shared meaning-making emerges. If we embrace it, we change our views or are at least open to the possibility that we were wrong in the first place. Critical niche construction in the online social realm starts with this possibility of changing oneself as a reaction to another's reaction. This is in line with the enactive view of information: not as a representation of data or mental content, instead of as "the making of a difference that makes a difference for some-body somewhere." (Thompson 2007, p. 57) As long as we publish information on social media without looking back, we are not exchanging it nor allowing it to make a difference for anyone, including ourselves. We will keep the other's agency and our own intact whenever we make something public as a proposal, a tentative making sense of something, committing to engage with other's reactions in a way that keeps us vulnerable to their understanding and reframing. Such an online environment would be richer and allow for the establishment of moral norms fostering the users' capabilities for moral sensitivity.

5 Conclusions

The three ethical principles proposed above aim to overcome the limitations of the online modes of interaction on social media platforms, namely interactions that are mediated, asynchronous, and ambiguous regarding the intended audience. An enactive take on social media ethics supplements missing components of classical moral frameworks by asking us to focus on the relations established through the interactions and on the environment created by the interactions. This does not entail that the units of analysis for classical ethics, namely the moral agent and the moral patient, are

now obsolete. One can still be responsible for inflicting harm on others while interacting on social media, such as those users who push others to kill themselves by leaving hundreds of comments of this sort under different names or the trolls that harass the vulnerable. These kinds of behaviours will count as immoral under any ethical framework. The main challenge in this paper was finding a way of morally evaluating those seemingly morally neutral behaviours which somehow become morally problematic when aggregated. There is still ongoing harm in social media environments even in the absence of any intent to harm, harms which cannot be traced to one agent or a collective group. In these particular situations, an ethics inspired by enactivism allows us to pay attention to the quality of our online interactions and the environment we generate for others, placing responsibility for us even for our most minor actions.

Declarations

References

Alfano M, Robinson B (2017) Gossip as a burdened virtue. Ethical Theory Moral Pract 20(3):473–487. https://doi.org/10.1007/s10677-017-9809-y

Alfano M, Reimann R, Quintana I, Cheong M, Klein C (2021) The affiliative use of emoji and hashtags in the Black lives matter movement: a twitter case study. https://doi.org/10.21203/rs.3.rs-741674/v1

Arfini S (2019) Situated ignorance: the distribution and extension of ignorance in cognitive niches. Synthese. https://doi.org/10.1007/s11229-019-02328-0

Arfini S, Bertolotti T, Magnani L (2018) The diffusion of ignorance in on-line communities. Int J Technoethics 9(1):37–50. https://doi.org/10.4018/IJT.2018010104

Backer CJ, Larson C, Fisher ML, McAndrew FT, Rudnicki K (2016) When strangers start to gossip: investigating the effect of gossip on cooperation in a prisoner's dilemma game. Evol Psychol Sci 2(4):268–277. https://doi.org/10.1007/s40806-016-0063-7

Bakir V, McStay A (2018) Fake news and the economy of emotions. Digit J 6:154–175. https://doi.org/10.1080/21670811.2017.1345645

Boyd DM, Ellison NB (2007) Social network sites: definition, history, and scholarship. J Comput-Mediat Commun 13(1):210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Cahir J, Lloyd J (2015) 'People just don't care': practices of text messaging in the presence of others. Media Cult Soc 37(5):703–719. https://doi.org/10.1177/0163443715577242

Cocking D, van den Hoven J (2018) Evil online. Blackwell public philosophy. Wiley, Hoboken

Colombetti G, Torrance S (2009) Emotion and ethics: an inter-(en)active approach. Phenomenol Cogn Sci 8(4):505–526

Crockett MJ (2017) Moral outrage in the digital age. Nat Hum Behav 1(11):769

de Jaegher H (2009) Social understanding through direct perception? Yes, by interacting. Conscious Cogn 18:535–542; discussion 543-550. https://doi.org/10.1016/j.concog.2008.10.007

de Jaegher H, Di Paolo E (2007) Participatory sense-making. Phenom Cogn Sci 6(4):485–507. https://doi.org/10.1007/s11097-007-9076-9

Di Paolo E, Rohde M, De Jaegher H (2010) Horizons for the enactive mind: values, social interaction, and play. In: Stewart J, Gapenne O, Di Paolo EA (eds) Enaction Toward a New Paradigm for Cognitive Science (pp. 33–88). Cambridge, MA: MIT Press

Di Paolo EA, Cuffari EC, de Jaegher H (2018) Linguistic bodies: the continuity between life and language. The MIT Press, Cambridge

Floridi L (2007) A look into the future impact of ICT on our lives. Inf Soc 23(1):59–64. https://doi.org/10.1080/01972240601059094

Floridi L (2013) The ethics of information. Oxford University Press, Oxford

Floridi L (2008) Foundations of information ethics. In: Himma KE, Tavani HT (eds) The handbook of information and computer ethics. Wiley, Hoboken; Wiley [distributor], Chichester, p 3–24

Froese T, Di Paolo E (2010) Modeling social interaction as perceptual crossing: an investigation into the dynamics of the interaction process. Connect Sci, 22(1), 43–68

Gallagher S (2017) Enactivist interventions: rethinking the mind. Oxford University Press, Oxford

Halpin H, Clark A, Wheeler M (2010) Towards a philosophy of the web: representation, enaction, collective intelligence. In: Proceedings of the WebSci10: Extending the Frontiers of Society On-Line. Raleigh, NC, US.

Herschbach M (2012) On the role of social interaction in social cognition: a mechanistic alternative to enactivism. Phenomenol Cogn Sci 11:467–486. https://doi.org/10.1007/s11097-011-9209-z

Hufendiek R (2017) Affordances and the normativity of emotions. Synthese 194(11):4455–4476. https://doi.org/10.1007/s11229-016-1144-7

Huldtgren A (2015) Design for values in ICTInformation and communication technologies. In: van den Hoven J, Vermaas PE, van de Poel I (eds) Handbook of ethics, values, and technological design: sources, theory, values and application domains. Springer Netherlands, Dordrecht, pp 739–767. https://doi.org/10.1007/978-94-007-6970-0_35

Isin EF, Ruppert ES (2015) Being digital citizens. Rowman & Littlefield International, London

Jordan J (2007) Taking the first step toward a moral action: a review of moral sensitivity measurement across domains. J Genet Psychol 168(3):323–359. https://doi.org/10.3200/GNTP.168.3.323-360

Krämer S (2015) Medium, messenger, transmission: An approach to media philosophy (Recursions: theories of media, materiality, and cultural techniques). Amsterdam University Press, Amsterdam

Marin L (2021) Sharing (mis) information on social networking sites. An exploration of the norms for distributing content authored by others. Ethics Inf Technol 23(3):363–372. https://doi.org/10.1007/s10676-021-09578-y

Marin L, Roeser S (2020) Emotions and digital well-being. The rationalistic bias of social media design in online deliberations. In: Burr C, Floridi L (eds) Ethics of digital well-being: a multidisciplinary approach. Springer, New York

Rest JR (1992) Development in judging moral issues. University of Minnesota Press, Minneapolis

Rietveld E, Kiverstein J (2014) A rich landscape of affordances. Ecol Psychol 26(4):325–352. https://doi.org/10.1080/10407413.2014.958035

Steinert S, Roeser S (2020) Emotions, values and technology: illuminating the blind spots. J Responsib Innov 7(3):298–319. https://doi.org/10.1080/23299460.2020.1738024

Taebi B (2021) Ethics and engineering: an introduction. Cambridge University Press, Cambridge

Thompson E (2007) Mind in life: biology, phenomenology and the sciences of mind. Harvard University Press, Cambridge

Van den Hoven J (2007) ICT and value sensitive design. In: The information society: innovation, legitimacy, ethics and democracy in honor of Professor Jacques Berleur SJ. Springer, Boston, pp 67–72

van Grunsven J (2018) Enactivism, second-person engagement and personal responsibility. Phenomenol Cogn Sci 17:131–156. https://doi.org/10.1007/s11097-017-9500-8

van Grunsven J (2020) Perceiving 'other' minds: autism, 4e cognition, and the idea of neurodiversity. J Conscious Stud 27(7–8):115–143

Varela FJ (1997) Patterns of life: intertwining identity and cognition. Brain Cogn 34(1):72–87

Watson J (2008) Media communication: an introduction to theory and process, 3rd edn. Palgrave Macmillan, New York

Wittgenstein L (2010) Philosophical investigations. Wiley, New York