



Shaping Social Media Minds: Scaffolding Empathy in Digitally Mediated Interactions?

Carmen Mossner¹ · Sven Walter¹

Accepted: 28 February 2024
© The Author(s) 2024

Abstract

Empathy is an integral aspect of human existence. Without at least a basic ability to access others' affective life, social interactions would be well-nigh impossible. Yet, recent studies seem to show that the means we have acquired to access others' emotional life no longer function well in what has become our everyday business – technologically mediated interactions in digital spaces. If this is correct, there are two important questions: (1) What makes empathy for frequent internet users so difficult? and (2) What can we do to alleviate the negative consequences? Correspondingly, the aim of this paper is twofold. First, we identify structural differences between offline and technologically mediated interactions that can explain why digital empathy is harder to achieve. Second, drawing on the literature on 'situated affectivity,' we consider the idea of modifying digital spaces in ways specifically designed to 'scaffold' empathy where our evolved mechanisms fail. Section 2 argues that empathy is requires *interpreting* the behavior of *embodied subjects*. Section 3 identifies three factors that are crucial for this interpretative endeavor: the empathizer's *affective repertoire*, their *perceptual input*, and their *background knowledge*. Section 4 argues that technologically mediated interactions differ from face-to-face interactions with regard to these factors in ways which render our evolved empathy mechanisms less effective in the digital world. Section 5 introduces the idea that situational factors can serve as 'empathic scaffolds,' i.e., as 'tools' that can 'shape' people's empathic reactions. Section 6 wraps up the main line of reasoning, responds to objections and invites further scholarship.

Keywords Digital empathy · Mind shaping · Nudging · Scaffolding · Situated affectivity · Social media

1 Introduction

Empathy is an integral aspect of human existence. Without it, our life would be rather bleak. Barely livable, in fact. Without at least a basic ability to access others' affective life, communication and interaction, caring, meaningful social relationships, and social or cultural norms—anything that requires *understanding* others—would be well-nigh impossible. But empathy is not uniquely human. The capacity to assess what others are feeling has evolved early during mammalian evolution: Non-human primates are capable of at least primordial empathy, as are elephants, dolphins, dogs, and rodents (Preston and de Waal 2002). This suggests that

empathy is an *adaptation*, i.e., a trait present in us because it was conducive to our ancestors' fitness. And indeed: Being able to understand others' emotional state allows us to provide the kind of comfort, advice, relief, assistance etc. that facilitates coping with environmental contingencies (Schulz 2017).

However, psychological mechanisms that used to be adaptive may cease to be so if the environment changes so rapidly that evolution cannot adapt them swiftly enough. The momentous cultural and technological developments throughout the past two millennia have changed our environment so radically that what evolutionary psychologists call our 'stone age minds' often cannot keep up. More and more studies claim to provide evidence for a noticeable decline in the capacity for empathy in an increasing number of people—quite generally (e.g., Konrath et al. 2011; Zarins and Konrath 2017), but specifically in frequent internet and social media users (e.g., Konrath 2013; Martingano et al. 2022; Small and Vorgan 2008; Twenge 2013; Turkle 2017),

✉ Sven Walter
mail@svenwalter.eu

¹ Institute of Cognitive Science, Osnabrück University, Wachsbleiche 27, D-49069 Osnabrück, Germany

leading to what has been called the ‘media-empathy paradox’ (Guan et al. 2019): Precisely the technologies that have been expressly designed to foster social connection seem to lead to a deterioration in people’s interpersonal capacities. Apparently, the means humans have, throughout the ages, acquired to access others’ emotional life no longer seem to function well in what has become our everyday business – technologically mediated social interactions in online spaces. With apparently tremendous consequences (e.g., Dregdge and Schreur 2020): The resulting decrease in empathy appears, among other things, to negatively impact the health (e.g., Abi-Jaoude et al. 2020) and life satisfaction (e.g., Usán Supervía et al. 2023) of frequent internet users, and to promote the spread of fake news (e.g., Vafeiadisa and Xiao 2021), online radicalization (e.g., Feddes et al. 2015) and hate speech (e.g., Hangartner et al. 2021).

If (and this is a large *if*; see Sect. 6) these diagnoses are by and large correct – i.e., *if* there is indeed a decrease in empathy in frequent online users and *if* that decrease is indeed something we should find worrisome – there are two philosophically and societally important questions: (1) *What makes empathy for frequent online users so difficult?* and (2) *What can be done to alleviate the negative consequences, both for them and others?* Taking the apparent empirical evidence for a worrisome decline in digital empathy at face value (but see Sect. 6), the aim of this paper is therefore twofold. Firstly, in order to address (1), we identify structural differences between offline and technologically mediated interactions that can contribute to an explanation of why digital empathy is harder to achieve. Secondly, in order to address (2), we draw on ideas from ‘situated affectivity’ research (e.g., Stephan et al. 2014; Walter and Stephan 2023) and consider the idea of modifying online spaces in ways specifically designed to foster empathy where and when our evolved mechanisms fail.

Section 2 argues that empathy is, at its core, a matter of *interpreting* the behavior of *embodied subjects*. Section 3 identifies three factors that are crucial for this interpretative endeavor: the empathizer’s *affective repertoire*, their *perceptual input*, and their *background knowledge*. Section 4 argues that technologically mediated and face-to-face interactions differ with regard to these factors in ways which often render our evolved empathy mechanisms less effective in the digital world. This answers question (1). Section 5 explores the idea that in such cases situational factors can serve as ‘empathic scaffolds,’ i.e., as ‘tools’ that can ‘shape’ people’s empathic reactions. This offers a tentative answer to question (2). Section 6 wraps things up, points out limitations, responds to objections and invites further scholarship.

2 Empathy: Interpreting Embodied Subjects

Empathy is usually described as a ‘multidimensional’ construct comprising various affective, cognitive, and conative facets (e.g., Lietz et al. 2011). Conceptually, it can be distinguished from several closely related phenomena such as mimicry, sympathy, perspective taking, or compassion (e.g., Maibom 2020, ch. 1).¹ Interesting and important as such nuanced differentiations might be, however, at a first approximation empathy might simply be regarded as the skill to affectively understand and adequately respond to others’ feelings. This is certainly too broad to be called an ‘account’ of empathy. But we do not want to arbitrate between the plethora of competing theories. Rather, we aim to offer an analysis of what can make digital empathy more difficult that can be valuable for as many accounts of empathy as possible. Therefore, distilling a fairly general ‘common core’ of extant accounts is precisely what is needed. This ‘common core,’ we venture, is that empathic observers have to (1) recognize the target as an *embodied subject* and (2) *interpret* the available behavioral and/or contextual cues.

First, if empathy is a matter of understanding (and eventually responding) to how the target is *feeling*, the claim that there is a kind of genuine empathy with fictional or otherwise inanimate objects is at least misleading: For if observers are aware of the inanimate nature of an object, they *know* that there is no emotional state they can ‘feel into.’ At best, they can project their own emotional state ‘onto’ the object. Such cases deserve consideration (e.g., Fuchs 2014; Safdari Sharabiani 2021), but they are a far cry from the kind of understanding we generally (try to) attain in social interactions: In order for us to even *wonder* what others’ affective perspective on the world is, we have to recognize (at least implicitly) that they *have* such a perspective, i.e., that they are not a mere ‘it,’ but a ‘you,’ not an object, but a *subject*. In particular, our recognition of others as a subject will typically include the (tacit) presupposition that the way they affectively interact with their environment is, among other things, a matter of their (lived) *body*. It is this assumption that the target is not only a *subject*, but an *embodied subject* (e.g., Gallagher 2008; Osler 2021; Osler and Zahavi 2022; Svenaeus 2021), that provides us with reasons for thinking that, given the similarities in bodily constitution and behavior, they experience and express emotions in similar ways,

¹ Additionally, the conceptual terrain is often muddled by distinguishing between ‘cognitive’ and ‘affective’ empathy, i.e., the capacity to *consider* the world from the target’s point of view and the capacity to *feel into* them. While this distinction may be pragmatically useful, for instance for health professionals, the empathic understanding characteristic of social interactions cannot be solely cognitive or affective – such empathy would be empty without affectivity, but blind without cognition.

so that we can (ideally; see Sect. 3) resort to our own experiences to understand how they are feeling.

Second, while the recognition that others are *embodied subjects* can trigger a basic empathic understanding that they are experiencing some affective state or other – that in contrast to mere objects there really *are* emotional shoes one might try to step into – that alone arguably does not reveal *what* affective state others are experiencing. In order to figure that out, empathic observers have to *interpret* their affective life. This claim requires some elaboration and, for reasons that will emerge soon, argument.

We use the term ‘interpretation’ in a very liberal, undemanding sense. As we understand it, interpretation need not be *conceptual*, *conscious*, or *cognitively onerous*. We explicitly want to sidestep the question whether it proceeds along the lines of simulation-theory (e.g., Goldman 2006), theory-theory (e.g., Carruthers and Smith 1996) or (most likely, we think) any other account (e.g., Fuchs and De Jaeger 2009; Gallagher 2012). The following sections only assume that *somehow or other*, observers have to ‘work out’ the target’s affective life from whatever cues are available in the given context. That many, especially traditional, accounts of empathy require some such interpretation is arguably uncontroversial. With regard to so-called ‘direct social perception accounts’ (e.g., Gallagher 2008; Krueger 2018; Osler 2021; Zahavi 2011), however, this claim will strike many as controversial.

According to direct social perception accounts, observers need not *infer* happiness or sadness from overt expressive cues, but can *directly perceive* them. Krueger (2009), for instance, has argued that given that an agent’s observable behavior is *constitutive* of their affective states, we can “very literally see, in a direct and noninferential way, various emotions and moods as they ripple and flow across the terrain of the body’s movement and gesture” (p. 683). But if I ‘very literally’ “see my partner’s sadness in her slumped shoulders, furrowed brow, and quiet speech” (Krueger 2018, p. 301), then, it might be said, no interpretation is required – and that might be taken to suggest that such accounts are immune to, or unlikely to benefit from, the considerations to come. The remainder of this section sets out our reasons for thinking that this concern is unwarranted.²

First, if push came to shove, we would not be convinced that direct social perception accounts are viable. The argument sketched above rests on the assumption that if x is constitutive of y , then by seeing x one (‘very literally’) sees y . And this is, at the very least, controversial (Walter 2018): A jigsaw tile is constitutive of the whole jigsaw, but by seeing the tile, one does not see the jigsaw; the CPU is constitutive

of a computer, but by seeing the CPU, one does not see the computer.³

More importantly, though, even if the part-whole inference of direct social perception accounts is viable, they at least seem to fail to paint the full picture. *Even if* by seeing the target’s smile the observer would literally see their happiness, empathic understanding would still require *some* kind of interpretation on top of mere perception. For just as Ralph can see Hubert without *knowing that* it is Hubert he is seeing, even if the observer would, by seeing the target’s smile, *see* the target’s happiness, they would not, thereby, *eo ipso know that* it is *happiness* they are seeing.⁴ But this knowledge is necessary for truly *understanding* the target’s affective state – after all, if the target were a super-actor, the observer would see the same smile, but no happiness. Advocates of direct social perception have objected that perception is so ‘smart’ that it can do “most of the work without the need of extra cognitive [...] processes” (Gallagher 2008, p. 538). As they see it, “*in the broad range of normal circumstances* there is already so much available in the person’s movements, gestures, facial expressions, and so on, as well as in the pragmatic or social context, that I can grasp everything I need for understanding” (ibid., p. 540; emphasis added), and they insist that super-actor and other cases where we are “misled by what we perceive” and have to actively decipher the target’s affective life are “relatively rare” (ibid.). However, even if we typically perceive others’ emotions without cognitive effort later in life, it takes infants days or weeks to make the link between smiling and happiness and much longer to associate less basic emotions such as shame, pride, envy, or jealousy with characteristic bodily, behavioral and contextual patterns. And for people with autism spectrum disorder, for instance, such links may remain forever imperceptible (e.g., Coelho et al. 2023). But their difficulties when it comes to dealing with other’s emotions are usually not merely attributed to deficiencies in *perception*. Whatever they lack (or better: in whichever

³ Advocates of direct social perception have objected that this fails to distinguish between a ‘member-collection relation’ (between, say, a wood and the individual trees) and a ‘component-integral object relation’ (e.g., Krueger 2018; Overgaard 2014). They claim that in the latter case, “certain components of ‘integral objects’ ... may be so essential, structurally and functionally, that seeing a component is in fact sufficient to see the object of which it is a part” (Krueger 2018, p. 312). Since the outward expressions of, say, anger, are such “a significant component” (Overgaard 2014, p. 140) of it, “seeing them could count as seeing the anger” (ibid., p. 139; subscripts removed). However, *prima facie* at least the CPU is an ‘essential structural and functional component’ of the computer and the provost is a ‘significant part’ of the university, and yet one does not see the computer just because one sees the CPU or the university just because one sees the provost.

⁴ In terms of Overgaard (2014): Seeing_n (non-epistemically seeing) anger is not *eo ipso* seeing_e (epistemically seeing) anger.

² We are indebted to an anonymous reviewer for encouraging us to spell out in more detail how the considerations subsequently made relate to direct social perception accounts.

respect they *differ* from neurotypical people), i.e., whatever has to happen for perception ‘proper’ to be enriched into ‘smart’ social perception, is what we would subsume under the term ‘interpretation.’ Fortunately, it is fine to leave this conceptual dispute unresolved here, for there is another reason why the considerations to come are relevant for direct social perception accounts, even if they *are* viable and talk about ‘interpretation’ is eschewed *tout court*.

As seen above, direct social perception accounts admit that they only cover “the broad range of normal circumstances” (Gallagher 2008, p. 538), suggesting that ‘non-normal circumstances’ are sufficiently rare to be ignored. As we will see, however, there are reasons for thinking that digital encounters are precisely *not* comparable to “normal everyday encounters” (ibid., p. 540). As a consequence, the more effortful cases that require active deciphering are becoming ever more pervasive. And advocates of direct social perception accounts are aware of this. Osler (2021), for instance, argues that embodied subjects communicate their affective states through their “expressive, lived body” (p. 3), that can, in contrast to their physical body, extend into interpersonal spaces, including digital ones, where they can, in principle, be directly perceived. But she agrees that the necessary perceptual process faces more problems in digital settings and therefore offers an analysis of these problems that overlaps with and is complemented by our own (see also Osler and Zahavi 2022). Given this, it seems that even ardent advocates of a direct social perception account can benefit from a better understanding of the structural differences between technologically mediated and face-to-face encounters, even if they remain at unease with the talk about ‘interpretation.’ We therefore take it that the following is indeed relevant for virtually *all* accounts of empathy on the table.

Section 3 takes a closer look at what exactly enables us to decipher others’ affective life in everyday encounters. Section 4 then identifies structural differences between offline and online interactions.

3 The Foundation of Empathy: Affective Repertoires, Perceptual Input, and Background Knowledge

As indicated in Sect. 2, we will not attempt to develop a full-fledged account of empathy. The goal of this section is to identify three factors that (so far; see Sect. 6) make empathic understanding possible in the first place or at least further its accuracy. While there is ample discussion about the decline of empathy in frequent internet and in particular social media users, the question of what exactly it is that makes digital empathy more difficult is still a matter of debate (e.g., Bortolan 2022; Konrath 2013; Osler 2021;

Osler and Zahavi 2022; Svenaeus 2021). This section paves the way for a discussion of this question in Sect. 4.

In order for the interpretive feats that have to be performed in empathic interactions to be accomplished, the empathizer, the target and the context of their encounter ideally have to fulfill three conditions.

First, observers must have an appropriate *affective repertoire*. In the course of our ‘affective biographies’ (von Maur 2021) we encounter a multitude of emotional situations. Our respective experiences and reactions are sedimented into our lived body in the form of both implicit and explicit affective memories.⁵ This affective repertoire determines, together with concrete situational factors, what we are able to understand (ibid.). When we see that others react to similar situations similarly, our affective repertoire is thus an invaluable source of information. At the same time, though, it *constrains* us: If we’ve never even been close to something like the other’s shoes, we cannot step into them just so. We may be able to gain some understanding through the target’s explanations (and possibly other resources; see below), but barring any comparable experiential memory, we cannot truly resonate with them. Consider, for instance, the restless-legs-syndrome, a neurosensory disorder causing decidedly unpleasant bodily sensations patients are typically unable to convey. Abstract knowledge can enable others to somehow grasp that their experiences are unpleasant and most likely also the rough extent or severity of these experiences, but accurately resonating with them proves to be an exceptionally arduous endeavor: Without an appropriate affective repertoire, observers might feel sympathy or compassion (see Sect. 1), but cannot relate the expressive cues to isomorphic experiences. In such cases, they at best have what Hume (1748, Sect. 2) called a “faint and dull” *idea*, a mere “copy” of the “more lively” *impressions* of the target.

Second, some *perceptual input* – from straightforward verbal reports to more opaque cues such as facial expressions, gestures, body posture, inflexion, diction – must be, first, *provided*, and, second, *perceived*, i.e., *accessed* and *processed*. The observer’s interpretation can change in arduousness and accuracy, depending on the richness, intensity, level of detail, accessibility, and processability of the perceptual input, which all can vary for different reasons. On the one hand, accurate empathy might be hard to achieve because the target provides no or only misleading input, say when they attempt to display a ‘false’ emotional state in order to provoke undeserved advice, relief, consolation etc. On the other hand, even if the target is honest and forthcoming and eloquently reports their feelings, accompanied by

⁵ Which is, again, also accepted by direct social perception accounts. As for instance Gallagher (2008) points out, “[m]emory works implicitly to inform my perception” (p. 537).

strong emotional expressions, unfavorable background conditions such as noise, bad lighting or an unusual context can affect the accessibility or processability of the input, again resulting in less accurate empathy. In addition, as Osler (2021) has pointed out, the intensity with which the target provides the input may itself hamper its accessibility or processability – shouting, for instance, despite the perceptually rich expression, can effectively *weaken* the amount of useful expressive cues conveyed.

Third, *background knowledge* – about the target, their affective state, or the context – can provide (additional) information that facilitates an interpretation. The more observers know about the target’s character, history, situation, hopes, commitments, expectations etc., the less ambiguous the input. Importantly, background knowledge cannot only supplement, but also to some extent be a *surrogate* for an insufficient affective repertoire or perceptual input.⁶ For instance, if the observer knows that the target has experienced a significant bereavement, they can empathize with them even in the absence of any further perceptual input. Similarly, if there is no suitable experiential memory in the observer’s affective repertoire, as in the case of restless-legs-syndrome, background knowledge – about, say, which familiar feeling the target’s feeling resembles – can allow for at least a basic understanding.

Given what has been said in this section, the increasing prevalence of digitally mediated interactions raises the question of whether, and if so, how, digital and face-to-face encounters differ (see also Osler and Zahavi 2022; Svenaeus 2021), in particular with regard to the *affective repertoire* of the observers, their *perceptual input*, and their *background knowledge*.

4 Empathy in Technologically Mediated Interactions: Perils and Pitfalls

Regardless of whether observers encounter the target face-to-face or through video-calls, Snapchat videos, personal text messages or impersonal status updates, we have argued, empathy requires that they *interpret* the target’s behavior, and this interpretation is based upon their affective repertoire, the extent to which perceptual input is provided, accessed, and processed and the background knowledge they can draw on. As we will see, all these factors are typically more error-prone in technologically mediated interactions, which contributes to the ongoing attempts (e.g., Osler and Zahavi 2022; Svenaeus 2021) to understand how non-mediated and mediated forms of sociality differ in ways

which make digital empathy often (see Sect. 6) harder to attain.⁷

4.1 Affective Repertoire

Frequent internet users often spend hours each day on social media. In particular, digital natives of the Gen Z and Gen α generation are increasingly growing up with such ‘screen time,’ not just passively consuming content but actively creating their own online social world, in which many interactions that still used to be face-to-face for the members of Gen Y at a comparable age are shifted to digital spaces.⁸ This influences their *affective repertoire* (e.g., Wood et al. 2016). The sparser the affective input during one’s development, the sparser one’s affective repertoire: For instance, a child raised in a Western climate of individualism and not accustomed to the affective intensity that accompanies interdependent affiliations in the Japanese culture will not be able to accurately empathize with the emotion of *amae* (roughly, a culturally ingrained indulgent dependency on others) that pervades the experience of Japanese children of the same age. And digital spaces have (as of yet) neither the same *depth* nor *diversity* than the non-digital world – a raised index finger emoji does not have the same affective vividness as one’s mother speaking to one’s conscience with a worried look and a caring voice, and the number of likes per post on Facebook does not distinguish between sincere maternal pride and shallow juvenile admiration. Compared to childhoods spent in nature, with peers, creative pretend play, and self-made non-technological toys, digital encounters are rather bleak and repetitive in their experiential diversity (Svenaeus 2021), not to mention the physical and mental health benefits, especially for fundamental executive functions, associated with ‘green time’ (Yogman et al. 2018). As adolescents have progressively less offline-experiences, there is thus a greater likelihood that there is less variety and depth in their experiences they can draw on (McNamee et al. 2021; Sriwilai and Charoensukmongkol 2016), with the consequence that they simply cannot understand or, for that matter, *feel* what the target is feeling (Uhls et al. 2014).

Moreover, *being capable* of experiencing the target’s emotion is to no avail if the experience, although generally

⁶ See also Svenaeus’ (2021) distinction between ‘perceptual’ and ‘imaginative’ empathy.

⁷ As Aagaard (2022) points out, digital interactions can be more cumbersome because of technological malfunctions that do not arise face-to-face, but also because of intrinsic factors that even affect digital interactions with fully functional technology. We will consider both kinds of difficulties without explicitly distinguishing between them.

⁸ Generation Y comprises the demographic cohort born between the early 1980s and late 1990s. Generation Z comprises those born in the mid-to-late 1990s and early 2010s; the ‘Gen Zers,’ colloquially also known as the ‘zoomers,’ are the first to have grown up with access to the internet and portable digital technology. Generation α is the latest cohort with the early 2010s as birth-years.

familiar, cannot be *called up* in the given context. Another potential problem is therefore that the observer's affective biography might be so different that it is impossible for them to feel what the target is feeling *in a given context*, even if the experience is *generally* familiar. Someone who hates nothing in the world more than the taste of Brussels sprouts will not be able to truly resonate with someone whose mouth waters at the mere sight of them; imagining their own favorite mouth-watering food might tell them 'what it is like' for the other to indulge in Brussels sprouts, but that alone does not make them *understand* the other's perspective, precisely because they cannot imagine how someone can have *this* experience when eating Brussels sprouts.⁹ Social media usage can exacerbate or make such fundamental interpersonal differences more likely by decreasing people's self-esteem, with the result that *social comparison* prevails over *emotional contagion*: With intact emotional contagion mechanisms, witnessing positive moments of others triggers a positive affective response, but when social comparison dominates, positive experiences reported in posts or personal messages *negatively* affect the viewer (Gomez et al. 2022; de Vries et al. 2018). Yet, if the observer is unable to feel happy precisely *because* the target is happy, empathy is impossible.

Still another problem is that digital spaces can extenuate or suppress the *activation* of affective states. First, online interactions hold the danger of 'emotional exhaustion.' Social media users are confronted with emotional news of the same kind – for instance the death of someone's parent, partner, or pet – much more often than people that (used to) lead only 'offline lives.' Such recurring emotional input, partly or mostly from strangers and distant acquaintances, that usually does not trigger any emotional involvement beyond a few emoticons *desensitizes*, lessening the likelihood that a truly affective state is activated – the more who die, the less we care. Second, apart from an overload of empathy-demanding input, digital spaces also hold the danger of a general 'emotional numbness,' given that the emotions communicated there are often 'disembodied' (Fuchs 2014), i.e., expressed through written words and/or emoticons only, without facial expressions, bodily postures etc. To the extent that bodily expressions are an integral component of emotions, emotions are no longer the same if their expression is significantly different.

The difficulties discussed so far arise primarily in interactions between people with *different online biographies*,

such as, typically, the Gen Zers and their digital immigrant parents or grandparents,¹⁰ but they probably do not so much affect interactions between those accustomed to similar digital realities. Even such interactions are prone to a paucity in empathy, though.

4.2 Perceptual Input

As others have also pointed out (e.g., Aagaard 2022; Osler 2021; Svenaeus 2021) digital and 'real life' interactions differ with respect to the *amount* and the *quality* of the *perceptual input* the target *provides* and the observer can *access* and *process*.

First, digital communication channels affect which input the target *provides* to begin with. Which personal thoughts and attitudes are shared and the manner in which they are expressed depends upon whether the interaction takes place online or offline. It is easier to feign misleading emotions online and correspondingly more difficult to decipher the target's true affective state (see Sect. 4.3). Moreover, digital spaces negatively affect people's feelings of restraint, leading to an 'online disinhibition' (Suler 2004) that makes them say things they would not dare say face-to-face. As a result, online communications can be more toxic, resulting again in a desensitization or even "dehumanization" (Harel et al. 2020) that makes empathy more difficult (but see Sect. 6).

Second, online spaces offer only a limited number of *modalities* in which perceptual input can be provided. In face-to-face encounters, one not only hears and sees the other, but can, for instance, also smell and touch them, which is (so far) impossible during online encounters. And yet, both smell and social touch demonstrably increase the quality of empathy.

Third, the communication channels offered by digital spaces are *quantitatively less rich* because observers can access only a limited part of the target's expressive field. In simple text messages or posts, they see only the other's written words or emoticons; in voice messages, they hear their spoken words, but still don't see them; in Instagram posts, they see them, but only in still images; in Snapchat videos, they see and hear them in dynamic action, but cannot interact with them etc. Even in live video-calls only a small fraction of the information that would be available offline is readily accessible (see Aagaard 2022). And yet, again, all these factors – seeing the target's face, hearing their voice, perceiving their gestures and postures, in particular dynamically and not just statically, or the context,

⁹ Would empathy allow for the context being totally blanked out, someone could be said to empathize with, say, the joy and satisfaction a psychopath is feeling when slaughtering a victim by imagining what they themselves are feeling when finishing a marathon – which is arguably not what we would ordinarily be prepared to say. We are indebted to an anonymous reviewer for asking us to address this point in more detail.

¹⁰ Importantly, while digital natives are more susceptible to the difficulties described given that they are more likely to be frequent social media users, the real issue is the communicative structure of digital spaces and what they (fail to) afford in interactions – and that affects everyone of any age.

and conversationally interacting with them – demonstrably increase the quality of empathy.

Fourth, input provided online is prone to be *qualitatively less rich*. Virtually everyone has experienced the hassles of bad audio quality, feedback loops, frozen cameras and connection time-outs. Nothing along these lines happens regularly face-to-face and it renders digital empathy more arduous and more likely inaccurate.

Fifth, another substantial advantage of face-to-face interactions is the possibility of *immediate feedback*. Such interactions enable observers to constantly monitor and adapt their empathy, for instance by asking clarifying questions or disclosing doubts, making it more likely that inaccuracies can be detected and corrected early on. In contrast, many technologically mediated forms of communication are not guaranteed to provide such immediate feedback, given that the other can respond with a delay of hours, days, or even more. Even the closest digital relative to face-to-face interactions, video-calls, can suffer from smaller time delays due to, say, transmission lags; and even barring any technological breakdowns, spontaneous and smooth turn-taking is made more difficult online (Aagaard 2022). But again, receiving immediate rather than delayed feedback demonstrably increases the quality of empathy.

4.3 Knowledge

Lastly, the anonymity of digital spaces also affects the observers' *background knowledge*, in particular about their conversational partners.

First, social media increase the number of contacts with people about which one has little personal and contextual information. This makes it harder, for example, to judge whether one's words may be hurtful or misunderstood, or what emotional value 'really' lies behind the target's stories, thereby affecting the accuracy of empathy. The more background knowledge is available, the easier it is to feel empathy (e.g., Behbahani and El-Nasr 2011). This is corroborated by the fact that *perceived similarity* and *perspective taking* are positively correlated, even cross-culturally (e.g., Heinkes and Louis 2009). To the extent that anonymity impedes the possibility of discerning commonalities such as cultural and political stances, physical appearance, attitudes, desires etc. that makes the observer perceive the target as a similar other, *authentically* stepping into the target's shoes becomes, once more, exceedingly difficult.

Second, besides the problems caused by the sheer number of social contacts, the information divulged on the internet is often filtered, embellished and concentrated on only few (positive) aspects of life, such as photoshopped Instagram stories. This makes it more difficult to obtain, literally, a true 'picture' of the target. Online platforms also make it

easier to actively deceive others, so that observers not only possess less knowledge about the target, but can also be less sure about its veracity, which can also decrease the accuracy of empathy.

Importantly, since the function of background knowledge is precisely to compensate for 'gaps' in the available perceptual input and/or the affective repertoire (see Sect. 3), the lack of knowledge aggravates the problems identified before.

To sum up, the considerations above illustrate how widely online spaces differ from non-digital ones and contribute further to an explanation for the apparent decrease in digital empathy: Through their effects on the observer's affective repertoire, the perceptual input and the background knowledge, technologically mediated interactions in digital spaces can make it (1) harder to experience empathy at all, (2) more likely that empathy is inaccurate, and (3) more difficult to assess how accurate whatever empathic understanding is eventually achieved actually is. This answers the first question raised in the introduction. Which brings us to the second: What can be done about the decrease in digital empathy that apparently results from the problems just described?

5 Changing Social Media Minds: Nudging Empathy with Empathic Scaffolds?

Even if the way online interactions currently work makes empathy often harder to achieve, digital empathy might still be fostered in some way or other. After all, wide – albeit not unanimous (e.g., Preston and de Waal 2002) – consensus has it that empathy is not a 'fixed' trait, but a skill. If so, then empathy, like any other skill, can be improved. In this section, we consider three techniques for potentially increasing empathy in online interactions: *behavioral*, *reflective*, and *automatic interventions*. We argue that the latter two are preferable to the first, and we tentatively suggest that among these two 'scaffolding' approaches, automatic ones might be preferable to reflective ones – although, admittedly, most of the pertinent research is still (not even) in its infancy.

5.1 Behavioral Interventions

Behavioral interventions make people engage in some sort of behavior that focusses their attention on the need for or motivation to seek empathic understanding and engage in empathic behavior (e.g., Weisz and Cikara 2021; Weisz et al. 2021). For instance, instructing subjects to engage in peer role plays, re-enact others' experiences, meditate or practice perspective-taking, for instance by writing essays about the target or by means of immersive virtual environments, can

positively affect their empathy. Unfortunately, behavioral interventions arguably won't be able to combat the apparent decrease in empathy among social media minds.

First, being tailor-made to specific *contexts* and *observer-target groups*, behavioral interventions do not *generalize* to novel situations or encounters outside of the original setting. While there is nothing wrong with such limitations per se, the concerns raised in Sect. 4 are so multifarious that more versatile techniques seem to be more promising.

Second, even if extant behavioral interventions could be amended so as to be applicable to a broader range of contexts or people, they typically require extensive preparation or training, oftentimes spanning weeks or months. Given their incessive interest in instant gratification (Wilmer and Chein 2016) and their reduced impulse control (Reed 2023), such long-term projects are unlikely to be pursued on their own precisely by those who seem to be most vulnerable, viz., the 'online community builders' of Gen Z and Gen α .¹¹ Moreover, there is no guarantee that behavioral interventions will be as effective for these cohorts, given the apparently pronounced cognitive and emotional changes they have undergone and still undergo (Uncapher and Wagner 2018). Techniques that are not totally alien to the daily routines of those who most likely need them therefore seem to be more promising.

Third, while extant research suggests that behavioral interventions can induce *temporary* changes in empathy, there is virtually no evidence that they can also prompt *long-term changes* (Behler and Berry 2022). In the absence of long-term effects, one strategy is to *repeat* interventions. However, in light of the enormous number of online encounters, it is hardly practical to play role-playing games, write essays, meditate, or immerse in virtual environments whenever one might face social interactions. Techniques that either have long-term effects or can effortlessly be coupled with online interactions therefore seem to be preferable.

The foregoing suggests that especially digital natives, but also frequent internet users in general, might benefit from techniques that can be seamlessly embedded into their daily online routines rather than sticking out as disruptions. Incorporating digital technologies into existing communication channels therefore appears to be an obvious option.

Let us illustrate the basic idea. The software *Project Us* (Rojas et al. 2022) is designed to mitigate the negative consequences of sparse or indistinct perceptual input (see Sect. 4.2) in online communications and to facilitate the interlocutors' empathy by relieving them from the task of having to actively decipher the other's affective state.

Project Us can be coupled with any common video conferencing system and uses machine learning algorithms to analyze the interlocutors' tone of voice and facial expressions and shares real-time feedback in the form of a simple binary color transition when it detects a negative emotional valence in one of the conversation partners. Users report that this 'affective traffic light' increases their awareness of others' emotions, as well as their own expressions and how they may be perceived, and makes them adapt their behavior accordingly. Similarly, *StoryChat* (Yen et al. 2023) is a tool for live streaming chatrooms that uses comic strip like visualizations to picture how a protagonist might experience the atmosphere of the chat, again relieving users from the task of having to detect emotionally relevant content in the flood of information they are exposed to. Users reported that "they felt the same way [as] the main character" (ibid., p. 10), with the result that their empathy and their level of irritation towards negative comments increased (ibid., p. 11).

Project Us and *StoryChat* are what one might call 'empathic scaffolds.' As situated approaches to cognition have argued, we can save precious cognitive resources by actively structuring our environment so as to solve routine problems or open up entirely new domains of competence *with least internal effort* (e.g., Clark 2008). Proponents of situated affectivity have argued that this also holds for our affective life (e.g., Stephan et al. 2014; Walter and Stephan 2023): We can modulate, sustain, enrich, expand or even make possible specific experiences by drawing on natural, social, or technological resources. This 'piggybacking' on external structures is known as 'scaffolding': We use the environment as a 'scaffold' when we actively structure it in such a way that we deliberately alter the challenges we face, our ability to cope with them, or the way we cope with them to our advantage. This is what *Project Us* and *StoryChat* do: The cognitive and affective work required to decipher emotions and detect and decode emotionally relevant comments are offloaded onto the software, with the goal of making errors less likely and freeing resources. Just as an abacus serves as a cognitive scaffold, and music, clothes, or therapists can serve as affective scaffolds (e.g., Coninx and Stephan 2021), *Project Us* and *StoryChat* serve as 'empathic scaffolds.' In the remainder of the paper, we consider two potentially different kinds of such empathic scaffolds.

5.2 Reflective Interventions

According to so-called 'dual process theories,' the human mind has two modes of decision making: 'System 1' is fast, automatic and unconscious; 'System 2' is slow, reflective and conscious (Kahneman 2011). So-called 'nudges' (Thaler and Sunstein 2008) target either System 1 or System 2.

¹¹ For example, a full 99% of the 12-17-year-olds in the UK are online on average three hours a day, with the majority of this time spent on social media platforms, which 98% of the 12-17 year olds and already 48% of the 3-4-year-olds use (Ofcom 2023).

A *nudge* is anything that influences the likelihood that people choose a particular option without forcing, incentivizing or rationally convincing them to do so, simply by altering the choice environment. One important distinction that has emerged in the burgeoning nudging literature is that between *reflective* and *automatic* nudges (Hansen and Jespersen 2013).

Reflective nudges aim to pull people out of the well-trodden mindless paths of System 1 in the hope of making them reach better decisions with System 2. For instance, asking people to consciously reconsider what they have written when an AI detects offensive or otherwise unwanted content can be strikingly effective: Twitter users reformulated or even abandoned offensive tweets after having been asked to review them before submitting and were subsequently less likely to draft offensive tweets (Katsaros et al. 2022). Along similar lines, Wang et al. (2014) designed a Facebook-plugin that prompts a reflective decision by displaying the profile pictures of five random contacts from the user's friend list as they are typing a post, telling the user that 'These people and [X; depending on the user's number of friends] more can see this' and giving them 20 seconds to edit the post afterwards. AI can also be used to suggest ways of rephrasing (or even automatically rephrase) offensive posts while preserving as much of their (non-offensive) meaning as possible (e.g., Tran et al. 2020).

Although reflective interventions are implemented by third parties and not actively recruited by the users themselves (see Sect. 5.3), they also do provide 'scaffolds' in the sense that they actively structure the environment with the goal of changing the way in which people cope with empathic challenges to their advantage. And they have some potential to foster digital empathy or empathic behavior: Unlike behavioral interventions, they are not limited to specific audiences and contexts, but available on any platform that offers some form of chat, comment or post function, and even if they do not have lasting effects, they can effortlessly be used time and again. However, they have problems of their own.

Reflective interventions work by activating the cognitive skills of System 2 required for self-monitoring, self-control, attention etc. But this is precisely what frequent internet users seem to be increasingly bad at (see Sect. 5.1). Interventions that pose fewer cognitive and motivational demands therefore seem to be preferable. Moreover, by disrupting System 1, reflective interventions are precisely that: *disruptions*. They will therefore arguably be experienced as intrusive, annoying, or patronizing, with the result that they fail to work or even 'backfire' (Hummel and Maedche 2019). Less intrusive interventions therefore seem to be preferable. *Automatic* interventions fulfill both requirements.

5.3 Automatic Interventions

As efficient cognitive systems try to reduce their cognitive load, System 2 kicks in only when System 1 cannot handle a situation. And System 1 can handle a lot: Our social life is to a striking extent a matter of processes that are non-consciously triggered and guide action to completion without our deliberate intervention. Much of the time, we operate on automatic pilot, relying on 'fast and frugal' heuristics that make our decisions fast and easy. These heuristics, however, also make us susceptible to *cognitive biases*. According to the *availability heuristic*, for instance, something is more important or better the more present it is in our mind. This allows us to quickly and effortlessly decide, say, which of two comparable products to buy. Yet, it also leads to an *availability bias*: We overestimate the importance or quality of something simply because it is more recent, vivid, or emotional. *Automatic* interventions target such biases. Unlike reflective interventions, they do not disrupt System 1, but tweak the circumstances in such a way that the unconscious biases of System 1 make a certain behavior more likely. The result is what Walter and Stephan (2023) have called 'mind shaping': Scaffolds are not only resources for affecting *one's own mind*, but can also serve as 'tools' for *shaping others' minds* by varying inconspicuous situational factors that influence their behavior beyond their conscious control.

Mind shaping of this kind has the potential to enhance digital empathy (but see Sect. 6). Something so seemingly marginal as the smell of chocolate chip cookies, freshly roasted coffee, or all-purpose cleaner or the hair-color or body weight of other people appears to have effects on pro-social behavior (Walter and Stephan 2023). In fact, such environmental and situational cues may even be *better predictors* of empathic understanding and empathic behavior than the observer's stable character traits (Darley and Batson 1973). This opens up the possibility of exploiting automatic interventions as 'empathic scaffolds' in order to unconsciously shape social media minds. Here are some examples from current research to illustrate the basic idea.

The *status-quo bias* makes us go along with the path of least resistance. As a consequence, *changing default settings* can affect our behavior. For instance, when the online multi-player game *League of Legends* changed its settings in such a way that players had to explicitly activate the chat function, rather than the all-chat being the default, toxic conversations decreased and positive exchanges increased (Kiritchenko et al. 2021). The same bias might also be exploited to combat the anonymity characteristic of online platforms that hampers empathy by masking that one is interacting with 'real' people. Making the use of real photos or realistic avatars (Ekdahl and Osler 2023) in social media

profiles the default option, so that one has to explicitly opt-out if one wants to remain hidden behind the digital veil of anonymous speech, might already make a difference. Not only could it prevent people from attacking others if their behavior could be traced to them, it would also give both offenders and victims literally a ‘face,’ making similarities more visible, thereby furthering empathy (see Sect. 4.3). It need not even be faces: All that matters is that a victim becomes *the* victim (Small and Loewenstein 2003), i.e., that the other is torn out of anonymity and made an identifiable *subject* (see Sect. 2). For instance, changing a typical social media interface in such a way that the comment box below a post does not say ‘Write a comment’ but ‘Write to ...’ followed by the user’s first name, has the potential to increase social transparency and accountability and prompt a greater internal motivation to intervene when the user is bullied (Taylor et al. 2019).

The *confirmation bias* refers to our tendency to focus on information that matches our beliefs. Together with the *availability bias*, the clearly visible number of ‘Likes’ on many social media platforms reinforces users in their views, including bullying and hate speech, by drawing attention effortlessly and exclusively to ‘friendly’ voices, while those who disagree have no equally straightforward means of expressing their dissent. Adding a ‘Dislike-button’ lessens such reinforcement by making critical alternative views more readily available, thereby decreasing the one-sidedness of the perceptual input provided online (Lee et al. 2022).

The reflective interventions discussed in Sect. 5.2 that ask users to reconfirm their tweets, messages etc. are also tapping into the *availability bias* by bringing the inappropriateness of their behavior to the foreground of their minds. Automatic nudges might accomplish the same while minimizing intrusiveness (see Sect. 5.2). For instance, Agapie et al. (2013) used a colored aureole around the query text box to make users formulate more elaborate queries. The aureole is red when the query box is empty, but as information is added, the aureole starts to fade, becoming blue when the input is perceived as enough to retrieve reliable search results. Using an AI to indicate potentially offensive language in a similar way might yield better results than the direct confrontation with verbal feedback, yet again reducing the cognitive load required to decipher the target’s affective state that may be the consequence of one’s own (intended) expressive behavior. This can be especially valuable during online interactions that lack the immediate feedback characteristic of face-to-face interactions (see Sect. 4.2).

6 Conclusion: Objections and Further Issues

We have argued that our traditional empathy mechanisms rely on some sort of interpretation or other (Sect. 2). The accuracy of such interpretation is a matter of the affective repertoire, the perceptual input and the background knowledge (Sect. 3). The fact that technologically mediated interactions are more problematic with regard to these factors (Sect. 4) can help explain the apparent decrease in empathy in frequent internet users. Supposing that such decrease is indeed problematic (see below), the question is what can be done about this. We have suggested that one option might be to develop ‘empathic scaffolds,’ in particular technological ‘nudges’ that actively shape social media minds (Sect. 5). This obviously raises many important questions that are beyond the scope of this paper.¹² We briefly touch upon some of them in our concluding remarks.

One important concern is the so far exclusively negative portrayal of the internet’s role with regard to empathy.¹³ The observations above were made, as announced in Sect. 1, against the background of the fact that more and more studies seem to demonstrate a lamentable decline in empathy in frequent internet and in particular social media users. As a result of this, it might have sounded as if the internet were per se ‘bad,’ producing only ‘narcissists,’ emotionally blind and cognitively crippled ‘social phobics.’ This is expressly neither our view nor true. First, while some studies are indeed pessimistic, others have actually documented an *increase* in empathy through social media use (e.g., Vossen and Valkenburg 2016), and even those who have been highly skeptical have recently come to reckon with the possibility that “the downward trend witnessed before 2009 might not have continued” (Konrath et al. 2023, p. 2). As so often, many empirical findings fall somewhere in between the extremes (e.g., James et al. 2017), calling for more research that distinguishes more carefully, for instance, between passive vs. active social media use (e.g., Verduyn et al. 2017). Nevertheless, as long it is still a live option that digital communication channels *do* have detrimental effects on empathic abilities, specifically of those who frequently use them from an early age on, the question what could be done to counter them remains important – and this is the question we have

¹² Among them the well-known questions regarding the ethics of nudging. However, apart from admitting that they apply to our considerations no less than to any other sort of nudging, this is not the place to go into them. Another intriguing and potentially fruitful issue that is beyond the scope of this paper is the advantage of technological nudges that, thanks to big data and machine learning methods, they can be personalized (Mills 2022), making them potentially more effective than the nudges hitherto investigated (Schmauder et al. 2023).

¹³ We are, once more, indebted to an anonymous reviewer for asking us to also address the positive sides of technologically mediated communication channels.

been concerned with above. Second, there can be hardly any doubt that digital encounters are – at least sometimes for at least some people – *superior* to face-to-face encounters: The internet can provide a safe space in which individuals who, for various reasons, experience rejection and discrimination by, or do not feel like they fit into, society, can cultivate relatively rich forms of empathy, find support, connection, mutual understanding, and a sense of empowerment within various online communities. Nevertheless, as long as there are also those who do *not* benefit from the structure of technologically mediated interactions, the question what could be done to counter them remains, again, important.

As a matter of fact, the two sides of the coin just mentioned might be related. The analysis offered in Sect. 4 of why online interactions are more problematic for some (neurotypical) individuals may also shed light on why they can be so valuable for others. If empathy is affected by, among other things, suitable background knowledge, a perceived similarity, the quality and quantity of the input, the immediacy of the feedback etc. (see Sect. 3), this not only explains why digital empathy might be harder to achieve for many, but also why others might benefit precisely from the special character of online spaces: The members of a closed LGBTQ+ forum might share more common background knowledge than they share with offline friends and family, fostering a perceived similarity that simplifies truly ‘feeling into’ the others; addicts might benefit from the ‘online disinhibition effect’ because the anonymity of the online self-help community allows them to speak candidly – to say what they would not dare say in ‘real’ life – enabling them to provide the input others need to emphasize with them; socially anxious or autistic people might use the extra time offered by the possibility of a delayed response to process information in the way they need to make an empathic response; introverted individuals or individuals with hearing impairments or speech disorders who might feel overwhelmed by the rapid pace and unfiltered massive input of in-person conversations might be better able to concentrate on the important things in digital spaces, again facilitating their understanding of others. Viewed from such a perspective, the analysis offered above might prove fruitful even beyond the (limited) use to which it was put in this paper.

Another obvious concern is that the scaffolds discussed in Sect. 5 do not, in fact, foster *empathy* at all, but instead replace it with a shallow surrogate, with a sort of ‘*as-if*-empathy’ that merely makes non-empathic people behave in superficially more pro-social ways. This concern seems to be corroborated by the fact that in Sect. 3 we identified factors that are necessary for empathy and at least some of the technologies discussed in Sect. 5 do nothing to change the interlocutors’ affective repertoire, the perceptual input or the background knowledge, instead seeking new ways

of fostering empathy that can dispense with what was necessary for face-to-face mechanisms of empathy to work. There are two ways to deal with this concern.

On the one hand, if one thinks that the observable behavior and internal reactions triggered by the sort of ‘empathic scaffolds’ discussed above have so little in common with ‘genuine’ empathy that the term ‘empathy’ is unwarranted, one might have to concede that such technologies might not stop the apparent empathy decay. At least three thoughts are still worth considering in this case, though. First, there might be a glimmer of hope, for getting used to behave *as if* one were empathic might be an important first step towards actually *being* empathic. Second, if the empathy deficit is lamentable enough, a bird in the hand might be worth two in the bush: Barring any other remedies, fostering *as-if*-empathy might be better than nothing. Third, if there is a decrease in ‘genuine’ empathy, the important question is: Is that indeed a problem, all things considered? Can we hope to cope with the challenges of our radically altered environment if we continue to chase the sort of empathy that results from mechanisms that have evolved under completely different circumstances? Is greater ‘genuine’ empathy really what we need (Mezzenzana and Peluso 2023)?

On the other hand, however, the objection under consideration rests on an intuition that will hardly resonate with those who accept the paradigm of situated cognition and affectivity. To claim that someone who adds two numbers by using pen and paper is not ‘genuinely’ doing math or that someone who keeps track of their appointments by using a notebook is not ‘genuinely’ remembering manifests precisely the sort of “biochauvinistic prejudice” (Clark 2008, p. 77) advocates of situated approaches have been opposing all along. In their eyes, the mechanisms underlying our cognitive and affective capacities can be implemented in various ways, some purely internal, but some also, and often more efficiently, by processes spanning brain, body, and environment, including technological devices. And just as doing sums with one’s fingers is not mere *as-if-calculation* but *another way of calculating*, the sort of empathy fostered by the technologies discussed in Sect. 5 might be ‘genuine,’ although different from the sort of empathy that results from our traditional mechanisms. This is admittedly at odds with the widespread tendency to think of empathy, just as of emotions such as love or shame, as something like a ‘natural kind’ that can be given a (real or nominal) definition in terms of individually necessary and jointly sufficient conditions. But just as the view that emotions have boundaries carved in nature has come under pressure (Barrett 2006), it is far from clear that empathy is indeed a homogeneous kind (Smith 2017) and not rather a multidimensional construct the manifestation of which is malleable and dependent,

among other things, on cultural factors such as prevailing norms, group standards etc. (Jami et al. 2023).¹⁴

While much of the foregoing is sketchy, and many questions need to be addressed, we hope that our thoughts resonate at least with those who have cast their eyes upon the issue of digital empathy and have walked away not entirely convinced.

Acknowledgements We thank the editors and reviewers for their time, work, and support. This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number GRK- 2185/1 (DFG Research Training Group Situated Cognition). We gratefully acknowledge the support of the Research Training Group Situated Cognition (RTG SC at <https://situatedcognition.com/>) and the comments and suggestions made by the members of both the Reading Club “Affectivity” of the Institute of Cognitive Science and the “Oberseminar” of the Institute of Philosophy of Osnabrück University, in particular Achim Stephan.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing interests The authors declare no competing interests.

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

References

- Aagaard J (2022) On the dynamics of Zoom fatigue. *Converg* 286:1878–1891. <https://doi.org/10.1177/13548565221099711>
- Abi-Jaoude E, Naylor K, Pignatiello A (2020) Smartphones, social media use and youth mental health. *Can Med Assoc J* 192(6):E136–E141. <https://doi.org/10.1503/cmaj.190434>
- Agapie E, Golovchinsky G, Qvarfordt P (2013) Leading people to longer queries. *SIGCHI Conf Hum Factors Comput Syst* 2367–2376. <https://doi.org/10.1145/2470654.2481418>
- Barrett LF (2006) Are emotions natural kinds? *Perspect Psychol Sci* 1(1):28–58. <https://doi.org/10.1111/j.1745-6916.2006.00003.x>
- Behbahani PA, El-Nasr MS (2011) The effect of privacy on social presence in location-based mobile games. In: Anacleto J, Fels S, Graham N, Kapralos B, El-Nasr MS, Stanley K (eds) *Entertainment computing – ICEC*. Springer, Berlin, pp 307–318
- Behler AMC, Berry DR (2022) Closing the empathy gap: a narrative review of the measurement and reduction of parochial empathy. *Soc Personal Psychol Compass* 16(9):e12701. <https://doi.org/10.1111/spe3.12701>
- Bortolan A (2022) Online emotions: a framework. *Inq*. <https://doi.org/10.1080/0020174X.2022.2044378>
- Carruthers P, Smith PK (1996) *Theories of theories of mind*. Cambridge University Press, Cambridge
- Clark A (2008) *Supersizing the mind*. Oxford University Press, Oxford
- Coelho S, Vendrell Ferran I, Stephan A (2023) Emotional abilities and art experience in autism spectrum disorder. *Phenomenol Cogn Sci*. <https://doi.org/10.1007/s11097-023-09917-y>
- Coninx S, Stephan A (2021) A taxonomy of environmentally scaffolded affectivity. *Dan Yearb Phil* 54:38–64. <https://doi.org/10.1163/24689300-bja10019>
- Darley J, Batson C (1973) ‘From Jerusalem to Jericho’: a study of situational and dispositional variables in helping behavior. *Pers Soc Psychol* 27(1):100–108. <https://doi.org/10.1037/h0034449>
- de Vries DA, Möller AM, Wieringa MS, Eigenraam AW, Hamelink K (2018) Social comparison as the thief of joy: emotional consequences of viewing strangers’ Instagram posts. *Media Psychol* 21(2):222–245. <https://doi.org/10.1080/15213269.2016.1267647>
- Dregdge R, Schreur L (2020) Social media use and offline interpersonal outcomes during youth: a systematic literature review. *Mass Commun Soc* 23(6):885–911. <https://doi.org/10.1080/15205436.2020.1810277>
- Ekdahl D, Osler L (2023) Expressive avatars: vitality in virtual worlds. *Phil Technol* 36(2):1–28. <https://doi.org/10.1007/s13347-023-00628-5>
- Feddes AR, Mann L, Doosje B (2015) Increasing self-esteem and empathy to prevent violent radicalization: a longitudinal quantitative evaluation of a resilience training focused on adolescents with a dual identity. *J Appl Soc Psychol* 45(7):400–411. <https://doi.org/10.1111/jasp.12307>
- Fuchs T (2014) The virtual other: Empathy in the age of virtuality. *J Conscious Stud* 21(5–6):152–173
- Fuchs T, De Jaegher H (2009) Enactive intersubjectivity: participatory sensemaking and mutual incorporation. *Phenomenol Cogn Sci* 8(4):465–486. <https://doi.org/10.1007/s11097-009-9136-4>
- Gallagher S (2008) Direct perception in the intersubjective context. *Conscious Cogn* 17(2):535–543. <https://doi.org/10.1016/j.concog.2008.03.003>
- Gallagher S (2012) Empathy, simulation, and narrative. *Sci Context* 25(3):355–381. <https://doi.org/10.1017/s0269889712000117>
- Goldman A (2006) *Simulating minds*. Oxford University Press, Oxford
- Gomez M, Klare D, Ceballos N, Dailey S, Kaiser S, Howard K (2022) Do you dare to compare? The key characteristics of social media users who frequently make online upward social comparisons. *Int J Hum Comput Interact* 38(10):938–948. <https://doi.org/10.1080/10447318.2021.1976510>
- Guan SSA, Hain S, Cabrera J, Rodarte A (2019) Social media use and empathy: a mini meta-analysis. *Soc Netw* 8:147–157. <https://doi.org/10.4236/sn.2019.84010>
- Hangartner D, Gennaro G, Alasiri S, Bährich N, Bornhoff A, Boucher J, Demirci BB, Derksen L, Hall A, Jochum M, Murias Munoz M, Richter M, Vogel F, Wittwer S, Wüthrich F, Gilardi F, Donnay K (2021) Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proc Natl Acad Sci* 118(50):e2116310118. <https://doi.org/10.1073/pnas.2116310118>
- Hansen PG, Jespersen AM (2013) Nudge and the manipulation of choice: a framework for the responsible use of the nudge approach to behaviour change in public policy. *Eur J Risk Regul* 4(1):3–28. <https://doi.org/10.1017/S1867299X00002762>
- Harel TO, Jameson JK, Maoz I (2020) The normalization of hatred: identity, affective polarization, and dehumanization on Facebook

¹⁴ If this line of reasoning is on the right track, then, as indicated above, the factors identified in Sect. 3 are necessary for our traditional empathy mechanisms to work, but not for empathy per se.

- in the context of intractable political conflict. *Soc Media + Soc* 6(2):1–10. <https://doi.org/10.1177/2056305120913983>
- Heinke MS, Louis WR (2009) Cultural background and individualistic-collectivistic values in relation to similarity, perspective taking, and empathy. *J Appl Soc Psychol* 39(11):2570–2590. <https://doi.org/10.1111/j.1559-1816.2009.00538.x>
- Hume D (1748) *An enquiry concerning human understanding*. Millar, London
- Hummel D, Maedche A (2019) How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *J Behav Exp Econ* 80:47–58. <https://doi.org/10.1016/j.socec.2019.03.005>
- James C, Davis K, Charamaraman L, Konrath SH, Slovak P, Weinstein E, Yarosh L (2017) Digital life and youth well-being, social connectedness, empathy, and narcissism. *Pediatrics* 140(2):S71–S75. <https://doi.org/10.1542/peds.2016-1758F>
- Jami PY, Walker DI, Mansouri B (2023) Interaction of empathy and culture: a review. *Curr Psychol* <https://doi.org/10.1007/s12144-023-04422-6>
- Kahneman D (2011) *Thinking, fast and slow*. Penguin, New York
- Katsaros M, Yang K, Fratomico L (2022) Reconsidering tweets: intervening during tweet creation decreases offensive content. Sixteenth Int AAAI Conf web Social Media 477–487. <https://doi.org/10.1609/icwsm.v16i1.19308>
- Kiritchenko S, Nejadgholi I, Fraser KC (2021) Confronting abusive language online: a survey from the ethical and human rights perspective. *J Artif Intell Res* 71:431–478. <https://doi.org/10.1613/jair.1.12590>
- Konrath SH (2013) The empathy paradox: increasing disconnection in the age of increasing connection. In: Luppincini R (ed) *Handbook of research on technoself: identity in a technological society*, vol 1. ICI Global, Hershey, pp 204–228
- Konrath SH, O'Brien EH, Hsing C (2011) Changes in dispositional empathy in American college students over time: a meta-analysis. *Pers Soc Psychol Rev* 15(2):180–198. <https://doi.org/10.1177/1088868310377395>
- Konrath SH, Martingano AJ, Davis M, Breithaupt F (2023) Empathy trends in American youth between 1979 and 2018: an update. *Soc Psychol Pers Sci*. <https://doi.org/10.1177/19485506231218360>
- Krueger JW (2009) Empathy and the extended mind. *Zygon* 44(3):675–698
- Krueger JW (2018) Direct social perception. In: Newen A, de Bruin L, Gallagher S (eds) *The Oxford handbook of 4E-cognition*. Oxford University Press, Oxford, pp 301–320
- Lee SM, Thomer AK, Lampe C (2022) The use of negative interface cues to change perceptions of online retributive harassment. *ACM on human-computer interaction*. 334:1–23. <https://doi.org/10.1145/3555226>
- Lietz CA, Gerdes KE, Sun F, Geiger JM, Wagaman MA, Segal EA (2011) The empathy assessment index (eai): A confirmatory factor analysis of a multidimensional model of empathy. *J Soc Soc Work Res* 2:104–124 <https://doi.org/5243/jsswr.2011.6>
- Maibom H (2020) *Empathy*. Routledge, London
- Martingano AJ, Konrath S, Zarins S, Okaomee AA (2022) Empathy, narcissism, alexithymia, and social media use. *Psych Pop Media* 11(4):413–422. <https://doi.org/10.1037/ppm0000419>
- McNamee P, Mendolia S, Yerokhin O (2021) Social media use and emotional and behavioural outcomes in adolescence: evidence from British longitudinal data. *Econ Hum Biol* 41:100992. <https://doi.org/10.1016/j.ehb.2021.100992>
- Mezzenzana F, Peluso D (eds) (2023) *Conversations on empathy: interdisciplinary perspectives on imagination and radical othering*. Routledge, London
- Mills S (2022) Personalized nudging. *Behav Public Policy* 6(1):150–159. <https://doi.org/10.1017/bpp.2020.7>
- Ofcom (2023) *Children and parents: media use and attitudes report 2023*. London. https://www.ofcom.org.uk/_data/assets/pdf_file/0027/255852/childrens-media-use-and-attitudes-report-2023.pdf
- Osler L (2021) Taking empathy online. *Inq* 1–28 <https://doi.org/10.1080/0020174x.2021.1899045>
- Osler L, Zahavi D (2022) Sociality and embodiment: online communication during and after Covid-19. *Found Sci*. <https://doi.org/10.1007/s10699-022-09861-1>
- Overgaard S (2014) McNeill on embodied perception theory. *Phil Q* 64(254):135–143. <https://doi.org/10.1093/pq/pqt020>
- Preston SD, de Waal FBM (2002) The communication of emotions and the possibility of empathy in animals. In: Post SG, Underwood LG, Schloss JP, Hurlbut WB (eds) *Altruism & altruistic love*. Oxford University Press, Oxford, pp 284–308
- Reed P (2023) Impact of social media use on executive function. *Comput Hum Behav*. <https://doi.org/10.1016/j.chb.2022.107598>. 141
- Rojas C, Zuccarelli E, Chin A, Patekar G, Esquivel D, Maes P (2022) Towards enhancing empathy through emotion augmented remote communication. 2022 CHI Conf Hum Factors Comput Syst 454(1–9). <https://doi.org/10.1145/3491101.3519797>
- Safdari Sharabiani A (2021) Genuine empathy with inanimate objects. *Phenomenol Cogn Sci* 20:831–846. <https://doi.org/10.1007/s11097-020-09715-w>
- Schmauder C, Karpus J, Moll M, Bahrami B, Deroy O (2023) Algorithmic nudging: the need for an interdisciplinary oversight. *Topoi* 42:799–807. <https://doi.org/10.1007/s11245-023-09907-4>
- Schulz AW (2017) The evolution of empathy. In: Maibom H (ed) *The Routledge handbook of philosophy of empathy*. Routledge, London, pp 64–74
- Small DA, Loewenstein G (2003) Helping a victim or helping the victim: Altruism and identifiability. *J Risk Uncertain* 26:5–16. <https://doi.org/10.1023/A:1022299422219>
- Small G, Vorgan G (2008) *iBrain: surviving the technological alteration of the modern mind*. Random House, New York
- Smith J (2017) What is empathy for? *Synthese* 194(3):709–722. <https://doi.org/10.1007/s11229-015-0771-8>
- Sriwilai K, Charoensukmongkol P (2016) Face it, don't Facebook it: impacts of social media addiction on mindfulness, coping strategies and the consequence on emotional exhaustion. *Stress Health* 32(4):427–434. <https://doi.org/10.1002/smi.2637>
- Stephan A, Walter S, Wilitzky W (2014) Emotions beyond brain and body. *Phil Psychol* 27(1):65–81. <https://doi.org/10.1080/09515089.2013.828376>
- Suler J (2004) The online disinhibition effect. *Cyberpsychol Behav* 7(3):321–326. <https://doi.org/10.1089/1094931041291295>
- Svenaues F (2021) Empathy and togetherness online compared to IRL: a phenomenological account. *J Phen Psych* 52(1):78–95. <https://doi.org/10.1163/15691624-12341384>
- Taylor SH, DiFranzo D, Choi YH, Sannon S, Bazarova NN (2019) Accountability and empathy by design: Encouraging bystander intervention to cyberbullying on social media. *ACM on human-computer interaction*. 118, 1–26 <https://doi.org/10.1145/3359220>
- Thaler RH, Sunstein CR (2008) *Nudge: improving decisions about health, wealth and happiness*. Penguin, London
- Tran M, Zhang Y, Soleymani M (2020) Towards a friendly online community: an unsupervised style transfer framework for profanity reduction. *Int Conf Comput Linguistics* 2107–2114. <https://doi.org/10.48550/arXiv.2011.00403>
- Turkle S (2017) *² Alone together*. New York, Basic Books
- Twenge JM (2013) The evidence for generation me and against generation we. *Emerg Adulthood* 1(1):11–16. <https://doi.org/10.1177/2167696812466548>
- Uhls YT, Michikyan M, Morris J, Garcia D, Small GW, Zgourou E, Greenfield PM (2014) Five days at outdoor education camp without screens improves preteen skills with nonverbal emotion

- cues. *Comput Hum Behav* 39:387–392. <https://doi.org/10.1016/j.chb.2014.05.036>
- Uncapher MR, Wagner AD (2018) Minds and brains of media multitaskers: current findings and future directions. *Proc Natl Acad Sci* 115(40):9889–9896. <https://doi.org/10.1073/pnas.1611612115>
- Usán Supervia P, Salavera Bordás C, Quílez Robres A, Lozano Blasco R, Latorre Coscolluela C (2023) Empathy, self-esteem and satisfaction with life in adolescent. *Child Youth Serv Rev* 144. <https://doi.org/10.1016/j.chilyouth.2022.106755>
- Vafeiadisa M, Xiao A (2021) Fake news: how emotions, involvement, need for cognition and rebuttal evidence (story vs. informational) influence consumer reactions toward a targeted organization. *Publ Relat Rev* 47(4):102088. <https://doi.org/10.1016/j.pubrev.2021.102088>
- Verduyn P, Ybarra O, Réisibois M, Jonides J, Kross E (2017) Do social network sites enhance or undermine subjective well-being? A critical review. *Soc Issues Policy Rev* 11(1):274–302. <https://doi.org/10.1111/sipr.12033>
- von Maur I (2021) Taking situatedness seriously. Embedding affective intentionality in forms of living. *Front Psychol* 12:599939. <https://doi.org/10.3389/fpsyg.2021.599939>
- Vossen HGM, Valkenburg PM (2016) Do social media foster or curtail adolescents' empathy? A longitudinal study. *Comput Hum Behav* 63:118–124. <https://doi.org/10.1016/j.chb.2016.05.040>
- Walter S (2018) Cognition, action, and self-control from the 4E Perspective. In: Newen A, de Bruin L, Gallagher S (eds) *The Oxford handbook of 4E-cognition*. Oxford University Press, Oxford, pp 321–332
- Walter S, Stephan A (2023) Situated Affectivity and Mind Shaping: Lessons from Social Psychology *Emotion Review* 15(1):3–16. <https://doi.org/10.1177/17540739221112419>
- Wang Y, Leon PG, Acquisti A, Faith Cranor L, Forget A, Sadeh N (2014) A field trial of privacy nudges for facebook. SIGCHI conference on human factors in computing systems. 2367–2376 <https://doi.org/10.1145/2556288.2557413>
- Weisz E, Cikara M (2021) Strategic regulation of empathy. *Trends Cogn Sci* 25(3):213–227. <https://doi.org/10.1016/j.tics.2020.12.002>
- Weisz E, Ong DC, Carlson RW, Zaki J (2021) Building empathy through motivation-based interventions. *Emot* 21(5):990–999. <https://doi.org/10.1037/emo0000929>
- Wilmer HH, Chein JM (2016) Mobile technology habits: patterns of association among device usage, intertemporal preference, impulse control, and reward sensitivity. *Psychon Bull Rev* 23:1607–1614. <https://doi.org/10.3758/s13423-016-1011-z>
- Wood MA, Bukowski WM, Lis E (2016) The digital self: how social media serves as a setting that shapes youth's emotional experiences. *Adolesc Res Rev* 1:163–173. <https://doi.org/10.1007/s40894-015-0014-8>
- Yen R, Feng L, Mehra B, Pang CC, Hu S, Zhicong L (2023) StoryChat: Designing a narrative-based viewer participation tool for live streaming chatrooms. *CHI Conf Hum Factors Comput Syst* 795:1–18. <https://doi.org/10.1145/3544548.3580912>
- Yogman M, Garner A, Hutchinson J, Hirsh-Pasek K, Michnick Golinkoff R (2018) The power of play: a pediatric role in enhancing development in young children. *Pediatrics* 142(3):e20182058. <https://doi.org/10.1542/peds.2018-2058>
- Zahavi D (2011) Empathy and direct social perception: A phenomenological proposal. *Rev Phil Psychol* 2(3):541–558. <https://doi.org/10.1007/s13164-011-0070-3>
- Zarins S, Konrath S (2017) Changes over time in compassion-related variables in the United States. In: Seppälä EM, Simon-Thomas E, Brown SL, Worline MC, Cameron CD, Doty JR (eds) *The Oxford handbook of compassion science*. Oxford University Press, Oxford/New York, pp 331–352

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