Mindshaping jerks: the folk psychology of online incivility and moral disengagement

Giacomo Figa' Talamanca

S1006192

Supervisor: Prof. Marc V.P. Slors

12th of May 2021

Thesis for obtaining a "Master of Arts" degree in Philosophy

Radboud University Nijmegen

I hereby declare and assure that I, Giacomo Figa' Talamanca, have drafted this thesis independently, that no other sources and/or means other than those mentioned have been used and that the passages of which the text content or meaning originates in other works - including electronic media - have been identified and the sources clearly stated.

Place: Nijmegen

Date: 12th of May 2021

Mindshaping Jerks: the Folk Psychology of Online Incivility and Moral Disengagement

In this paper, I will propose a systematic philosophical approach to understand and describe instances of hostile interaction in online environments. I will propose a recent approach in philosophy of mind that understands social cognition as fundamentally relying on socially shared norms as particularly resourceful for understanding hostile online interaction. Specifically, the lack of a determined normative context online makes behavior difficult to interpret, and users tend to apply their own normative standards for (in)appropriate behavior, while those standards might not be followed by other people. This leads to a lack of common ground and to reciprocal aggression. I will conclude by arguing that digital media literacy policies should favor virtuous and healthy engagement through the establishment of shared norms online.

Keywords: Online hostility - Folk psychology - Mindshaping - Context collapse - Online virtues - Reactive attitudes

Introduction

Incivility, toxicity, polarization and disinhibition are *du jour* concepts in contemporary research and political debate surrounding social media. Online polarization, understood as the way people with different or contrasting (political) views see each other as two sharply distinct and irreducible groups on social media, is a widely researched issue (Hart, Chinn, and Soroka, 2020; Iyengar and Massey, 2019; Maher, Igou, and van Tilburg, 2018; Nisbet, Cooper, and Garrett, 2015; Talisse, 2019). Social networking sites (henceforth SNSs) enable people to be exposed to and engage with a variety of information sources, as well as interacting with people holding beliefs and views different from their own (see e.g. Flaxman, Goel, and Rao 2016; Vaccari et al. 2016; Beam et al. 2018: Dubois and Blank 2018). However, they also enable a variety of aggressive and toxic behaviors, including flaming (Moor et al. 2010; Cho and Kwon 2015; Rost et al. 2016); cyberbullying (Langos 2012; Thomas, Connor and Scott 2014; Kowalski et al. 2014); trolling (Phillips 2015; Bail et al. 2020; Aro 2016); dehumanization during political discussion (Pacilli et al 2016; D'Errico and

Paciello 2017; Harel et al. 2020); and a perceived general lack of civility[1] in interaction (Antoci et al. 2016; Rains et al. 2017; Masullo Chen et al. 2019).

Despite the generalized understanding among academic and public contexts that online interaction is more disinhibited and toxic, the social and cognitive conditions that enable hostility and moral disengagement are somewhat understudied. Research on online (toxic) disinhibition as such is not new (Suler 2004; Lapidot-Lefler and Barak 2012). However, what is missing is an understanding of hostile interaction as a practice, and of how certain design features of SNSs such features affect the way their users make sense of others' behavior online and interact with one another.

In this paper, I will propose to examine online incivility and moral disengagement through the lens of philosophical theories of social cognition that highlight the role of the social context in our understanding of other minds. Various philosophers have highlighted the importance of the embodied and embedded context in social interaction (Gallagher and Hutto 2007; Gallagher 2008; Hutto 2008; Gallagher and Varga 2014) and its function as a frame that makes other agents predictable by creating expectations of conformity to established conventions and norms (McGeer 2007, 2015, 2020; Zawidzki 2008, 2013, 2018). In online environments, not only other agents are not embodied (they are not physically present), but there is no undisputed normative context determining what is (in)appropriate to say or do. I will argue that online incivility and moral disengagement arise due to the lack of shared communicative and interactional norms, which is due to some structural design features of these platforms. The absence of a shared context that can frame and constrain social interaction in online settings transform users' practical commitments, as well as the understanding of others' commitments, leading to a distortion of social cognition. This distortion, caused by an absence of shared contextual norms that would normally constitute a common ground between agents, leads to mutual hostility.

The argument will proceed as follows. In section 1, I will highlight two salient features of online interaction, that is the tight relation of users into their online profile and activities (which I will call cyborgification): and context collapse, that is the way generated content might be consumed by people who were not intended for consuming it. I will stress that these features are design choices, features of mainstream online platforms, such as Facebook and Twitter, that are intended by their programmers. In section 2, I will introduce

[1] To avoid paternalistic ways of approaching the problem of online incivility (Masullo Chen et al. 2019), I will follow Antoci et al. (2016) and define online incivility as: "[A] manner of offensive interaction that can range from aggressive commenting in threads, incensed discussions and rude critiques, to outrageous claims, hate speech and harassment" (1).

some philosophical reflections on the importance of context in social cognition. I will give special focus on the theory of mindshaping, which argues that social norms and conventions not only frame the way we understand other minds, but can also constrain others' behavior (and make them feel compelled to conform to such expectations) through the imposition of an expectancy effect. In section 3, I will analyze online hostility and aggression through the lens of the philosophical reflections of section 2. I will argue that the underdetermination of the context of interaction, coupled with the extreme focus of the platform's design to the individual user's content, beliefs and values, constitutes an inherent impediment for functional interaction due to the technologically-derived difficulty for interpreting each other's behavior. In other words, the design choices of the platform hinder the fundamental socio-cognitive capacity to understand other people's minds, which in turn leads to frustrated and aggressive interaction. In section 4, I will argue that policies and proposals that intend to make online interaction less violent and uncivil need to take into account the role of socially shared norms, and that establishing such norms is key for making online interaction civic and functional.

1. Agents and Context Online: Setting the Stage

Digital platforms, including SNSs, afford a wide range of activities. On sites such as Facebook and Twitter, users can find and interact with other people; they can produce content and express their thoughts; they can express their preferences and interests by both following the content of other Facebook pages and by creating some themselves; they can organize events that may take place both online and offline; and they can consume multimedial information such as the news. The users' newsfeed, where the platform provides consumable content, is designed to maximize a comfortable experience for users, enabling a smooth integration of the digital platform into users' practices, possibly making the technology involved "dissolve into the background" of user's activities on the platform.

In general, there is a consensus among philosophers that technological artifacts embed moral and political values and actively shape human experience, intentionality, decision-making and information processing (Ihde 1990; Clark 2003; Verbeek 2011; Miller 2021). Digital environments such as SNSs are no exception. They are complex and immersive ecosystems that enable a wide array of social, cognitive and epistemic practices (Smart et al. 2018a, 2018b). While the information users consume there always refers to people, objects and institutions that exist outside of the digital world (Taddeo 2015; Frömming et al. 2017), the information users interact with is what these digital platforms are made of. As Arfini et al (2017) put it: “[T]here is no gap between information and matter” (382) - for the information generated by users and webpages and provided by the platform’s algorithms is the only matter users directly interact with.

This sketch of SNSs as digital environments serves to contextualize the two features of social networking sites that I wish to highlight for characterizing the relevant characteristics of social interaction on SNSs. Both these features of SNSs are especially relevant for understanding social cognition and interaction online because they originate from what might be called a general informational scarcity. In face-to-face settings, social interaction is naturally embodied and embedded. In online settings, not only the expressivity and the influence of embodied feedback that normally prompts the direction of your response is missing; as I will argue in length below, the context that would normally mediate behavior and mutual understanding is also undetermined and disputed. On the one hand, users of SNSs are configured as exclusively constituted by information; on the other, in many digital platforms, the context of the interaction is somewhat undetermined. I will now examine these two features of SNSs, their ambiguous status and their relevance for online interaction.

1.1 Cyborgs Online: Users as Information

The first and probably most prominent feature of interaction in social networking sites, in contrast to face-to-face interaction, is, well, the lack of faces. In social networking sites, during real-time interaction, embodied cues are missing: users might be considered disembodied[2], because there is no physical proximity between them. This striking difference between face-to-face and computer-mediated interaction is somewhat compensated by the multifunctional interface of SNSs, which enable users to generate and share multimedial information about themselves, others and what they think.

Following Bertolotti et al. (2017), I will call this configuration of SNSs users as a cyborgification. The word cyborg has been popularized in philosophy of mind in recent years (Clark 2003; Verbeek 2007, 2011). The term is used to indicate how many artifacts, when tightly coupled[3] with agents, do not simply mediate the agent’s relation with the world, but

[2] I wish to add emphasis, here, on the “*might*”. Online harm and aggression can have real, offline consequences, despite the fact that, to those who (willingly or unwillingly) perpetrate the harm, this might not be evident.

[3] The instance of cyborgification proposed by Bertolotti et al. is inspired by, but not equal to, what Clark calls cognitive extension - when an artifact is integrated to a cognitive system to such a degree that it counts as a de facto part of that cognitive system (Clark and Chalmers 1998; Clark 2008). In general, I do not find to be the

can come to co-constitute them: the intentionality, experience and decision-making and cognitive processing of the human subject is radically transformed by these technological implementations.[4] What makes SNSs users "cyborgified" is their presence and configuration within digital platforms: on SNSs, users are made of the information they provide and the content they produce - that is, the personal information; their stated preferences; their friendship network; the multimedial content they post on their own newsfeed; and the comments to posts that other users and pages produce.

Agents on SNSs are fused with and perceived as their generated content, which appears as a set of manageable and interactable items in themselves, while also being informed about the material lives of the individuals that use them (Waite and Bourke 2013). Furthermore, in order to exist as SNS users and to interact with other people and information, people rely on the algorithms that select and provide possibly relevant content on the user's newsfeed. The sets of information SNSs users interact with are those automatically deemed relevant by the platform's algorithms. SNSs users are "diffused" cyborgs both because they are constituted by the information they generate (both as produced content or as the digital traces they leave, like preferences or "likes") and because these interactions are enabled by the programming of the platform. This inseparability of offline-originated user activity and the platform where the user acts and generates content enable users to access, communicate and share multimedial information; expressing, shaping and sharing its own narrative, values and thoughts in an autonomous way; and interact with members of her social network and expand it - practices that can be performed online only insofar the user acts upon her online profile (Taddeo 2015).

While the multifunctionality of SNSs for their users does allow for a relatively wide range of actions and interactions, the cyborgification of SNSs users compensates in not-neutral ways for the lack of embodied cues. When the expressivity and feedback that embodied presence provides is missing, social cognition needs to rely on what is available in the platform: the user's generated content. The first thing users see when interacting with other people is the content others generate, *before* the others themselves. This means that on SNSs we first see the other's beliefs (or, at least, their expressions of beliefs), and then we see

case that Web-based systems provide cognitive extension, and that such systems (including SNSs) transform and guide our cognitive processing in more nuanced ways (see Heersmink and Sutton 2020).

[4] Verbeek (2007, 2011), considers examples of "cyborgification" when pacemakers are installed to support a patient's heartbeat; or pharmaceutical technologies such as antidepressants help change people's moods; or prosthetic limbs are implanted. While these instances of cyborgification involve the physical body of the human agent, the meshing of technology and human agency in the case of SNSs needs to be understood as occurring within the digital environment.

the others. The lack of embodied cues[5] as a cognitive resource for understanding what's in the other person's mind causes a relocation of the user's attention to what is immediately available, the content itself, while the underlying motivations and offline background of the generated content is left ambiguous. From a first person-perspective, the design of SNSs such as Facebook and Twitter prompts users to define themselves, first of all, on the base of their preferences, desires and beliefs. In contrast, from a third-person perspective, people and their content are merged with the newsfeed: they become, in other words, objectified, a mere part of the informational environment.[6]

1.2 Context Collapse: the Indeterminacy of Audiences and Expectations

The problem underlying context collapse was already synthesized well by Kiesler et al. (1984), when they wrote that "Communicators must imagine their audience, for at a terminal it almost seems as though the computer itself is the audience" (p. 1125). As opposed to face-to-face interaction, when communicating with somebody online Internet users need, in a sense, to imagine their intended audience. On the one hand, the recipient of their communication is not tangible, but within or "on the other side of" the computer; on the other hand, there are many cases where the message may reach audiences that are not those intended by the user. In other words, in online environments such as SNS, different people from different social contexts end up consuming different information, which may or may not be intended for them specifically, within the same space. The online network of a user's *friends* or *followers* might include friends, family members, colleagues from work, or people she did not know offline who share her interests and beliefs. Therefore, when a SNS user produces content, she might have some of these people in mind as recipients, but the content might be seen by other people, who are not the intended recipient of the message.

[5] One might make a case that (at least some) emojis can be considered as a further way to supply for the lack of embodied cues. I will not examine this possibility here, and I do not believe that emojis, while somewhat less ambiguous than "mere" words, can fully supply for embodied expressivity: not only they are, at the end of the day, still computer-generated symbols, but they cannot supply for the direct feedback that characterizes face-to-face interaction.

[6] One might be tempted to compare the interaction with "mere information" that characterizes SNSs with message correspondence via mail. However, in the latter case there is awareness on the side of the recipient that the message is "just" a message, which is cognitively separated from the agent who wrote it. Furthermore, the sender of the mail is usually already known, or at least provides background information regarding herself and her motivations in the letter. This is also true for most interactions through email. In contrast, due to the environmental nature of SNSs, users within them are seen only as the information they provide, which can in many cases appear "out of context" - as a comment under another user's post.

This phenomenon is known as context collapse, "[T]he flattening out of multiple distinct audiences in one's social network, such that people from different contexts become part of a singular group of message recipients" (Vitak 2012, p. 451). This is an intrinsic feature of many Web-based technologies, a feature that users of those technologies need to adapt to, deal with or even make use of in a variety of ways. It is a particularly researched mechanism in SNSs, primarily in its implications to users' privacy management (Vitak 2012; Marwick and boyd 2014) and news sharing and exposure (Beam et al. 2018; Kim and Ihm 2020).

When it comes to social interaction as such, the collapse of different offline contexts into one has particularly relevant implications. Marwick and boyd (2011) observed that many American Twitter users keep in mind that the content they produce might be consumed by unintended recipients. This leads to, the authors point out, the users imagining the audience for the content: users need to moderate the content they post by keeping in mind who they think is going to end up reading it.

This does not necessarily mean that SNS users imagine themselves as actors at the center of attention and others online as a public (Szabla and Blommaert 2020). What it entails is that the various expectations, more or less acknowledged standards for appropriate and inappropriate behavior that underlie social interaction in different face-to-face settings become undetermined. The audience of the content you produce sets normative standards for how to behave and what to say. In face-to-face interaction, you would behave differently in front of strangers and with friends or family members: depending on the social context and on the people you are interacting with, there will be certain things that you are expected or supposed to (not) do or say. Online all of these audiences, as consumers of the content you produce, are reduced to one: therefore, standards of what is appropriate to say and do become equivocal.[7] To imagine the audience, in other words, is not simply to imagine the people who consume your content: it is to imagine the behavioral standards of user-generated content, to imagine the norms that underlie interaction in face-to-face settings and, at a more practical level, what other people online expect you to say.

[7] There are also some in-platform means to reduce or determine better the intended recipient(s) or a post (e.g. tagging on Facebook or @ing on Twitter); however, most of these built-in strategies do not avoid context collapse completely, insofar as the content is posted. The exception would be to send the content to your intended recipient through a private message. On the other hand, some users may create multiple social media profiles in order to avoid the conflation of different social contexts (Costa 2018). However, while such a strategy may cushion the impact of context collapse, it is questionable whether it is avoided completely - generated content might end up consumed by unexpected or unintended receivers anyway.

To summarize, user cyborgification and context collapse are two characteristics of SNSs that radically transform social cognition and interaction and are both results of the informational scarcity typical of computer-mediated communication. They are both examples of what Floridi (2017) calls "copy-pasting", or de- and re-coupling. In the case of user cyborgification, there is a de-coupling in a phenomenological sense, because the content generated by users is perceived as a separate entity from the user in his physically situated context; and a re-coupling, because SNSs users, within the digital environment, become constituted by the content they generate and the information they provide. A similar dynamic occurs for context collapse: pieces of behavior[8] that would normally have a specific meaning in a specific and determined context offline are "out there" and may be consumed and interpreted by people outside of the intended context. Not only users need to imagine their audience and its appropriate behavioral standards, but also consumers of the generated content might apply their own idea of what is appropriate to say and do, because the actual social context of SNSs is undetermined and its norms are disputed. Or, as Phillips and Milner (2017) put it: "Online, it is often easier to separate people from their embodied experiences, or to mistake the part for the whole - or to never even see the whole, and therefore never understand the context from which a particular collection of pixels has been unmoored (89).

Before moving one, there is one further thing that needs to be brought up. Cyborgification and context collapse can be considered, at a first glance, as inescapable aspects of computer-mediated communication. However, on SNSs such as Facebook and Twitter the primary activity consists in information consumption and sharing within one's own social network, which can also be expanded. Because this is the main function of these SNSs, the determination of the user's identity through information, as well as the blurring boundaries between the privacy of one's own close social network and the publicity of other SNS users, are design choices, features that are implemented by the platform's programmers. Platforms such as Facebook and Twitter are, to put it shortly, places where to express yourself and connect with all our friends, family and other people from many different backgrounds. The configuration of one's own presence within the platform is designed in relation to a collapsed and ultimately undetermined context. This is reflected by Mark Zuckerberg's statement:

[8] That is, in this case, generated content.

> “You have one identity … The days of you having a different image for your work friends or co-workers and for the other people you know are probably coming to an end pretty quickly ... Having two identities for yourself is an example of a lack of integrity.” (Quoted in Kirkpatrick 2010: 199)

In the course of this section, I highlighted two primary features of SNSs that drive social interaction online, affecting the perception of other people and the context where interaction occurs. However, this is not yet sufficient grounds to claim that these elements affect not only interaction with, but also the understanding of other people, their intentions, emotions and beliefs. I will now turn to a set of philosophical theories that highlight the importance of the embodied and embedded context for mental state attribution.

2. Mindshaping and Social Cues: the Importance of Context for Understanding other Minds

In philosophy of mind, folk psychology[9] is understood as the capacity to attribute mental states such as beliefs, intentions and desires. This capacity for understanding other people’s minds is primarily understood as a form of mindreading. Because we do not have a direct access to other people’s minds in the same way we do to our own, understanding what others are thinking consists in inferring and attributing mental states. Social cognition is therefore understood as a capacity to interpret and predict the behavior of other people in the course of social interaction. This capacity is generally understood as reliant on the possession of a theory regarding what mental states are and how they interrelate with actions (an account known as “theory theory”; see e.g., Nichols and Stich 2003), or on the possibility of modeling others’ mental states on our own (an account known as “simulation theory”; see e.g. Goldman 2006; Gallese 2009). Despite their differences, both these approaches see social cognition as individualistic and primarily epistemic. They rely on the internal capacity of singular agents to infer and know others’ mental states, a capacity that is generally understood as third-personal, or spectatorial.

This understanding of social cognition as a form of mindreading has been challenged from various standpoints in the last twenty years. Many philosophers criticized this

[9] This notion was elaborated presupposing an understanding of social cognition as presupposing the possession of a theory regarding how mental states work (Sellars 1959). However, even the philosophers who criticize this understanding of social cognition use the term folk psychology in their own explanation of the process (McGeer 2015). For this reason, I will use the terms social cognition and folk psychology interchangeably in the course of this paper.

spectatorial understanding of social cognition by highlighting the expressive role of embodied cues (Gallagher 2001, 2008; Zahavi 2007, 2011) and the typically engaged and participated nature of social interaction (Gallagher and Hutto 2007; Hutto 2008; Andrews 2009, 2012). Others criticized, in contrast, the primarily epistemic aspect of mindreading. Instead, these philosophers the importance of the social context in moderating and constraining behavioral expectations, theorizing folk psychology as a way of regulating behavior and not just predicting it (Mameli 2001; McGeer 2007, 2015, 2020; Zawidzki 2008, 2013, 2018; Gallagher and Varga 2014; Fernandez-Castro 2020). I will call this latter group of theories the "mindshaping" or regulative view of folk psychology, as they understand social cognition as reliant on and constrained by socially shared norms. I will now turn to examine the normative influence of the embodied and embedded context of social interaction, focusing especially (but not exclusively) on the second group of theories.[10] Understanding this role of the context will prove crucial in the understanding of social cognition on SNSs, where the context becomes underdetermined.

Before proceeding to this analysis, it is worth highlighting one specific advantage of these theories over traditional accounts of social cognition. Specifically, theory theory and simulation theory have problems in accounting for the fact that a piece of behavior can express one among many mental states. For instance, laughter can mean enjoyment, dismay, nervousness or contempt among others. In other words, it is problematic for mindreading accounts to provide an understanding of the relevant information in inferring what the other person is thinking, both what contextual elements are relevant in mental state attribution and what is the mental state underlying a piece of behavior. Accounts of folk psychology that rely entirely on an "in-the-head process" (be it the possession of a theory or the projection of oneself into the other's situation) have problems in explaining how can people reliably single out the appropriate characterization of the other person's behavior (and of the situation where the behavior occurs) to interpret the behavior accurately. One piece of behavior, taken by itself, can have different meanings under different circumstances. For this reason, any successful account of folk psychology needs to explain how agents can grasp with relative ease the relevant characterization of another person's situation and behavior in order to attribute her the correct mental state. This is called *the frame problem* (Haselager 1997;

[10] In this paper, I will not deal specifically with the different philosophical assumptions and traditions these different accounts build on. For instance, while Zawidzki and McGeer's accounts accept a degree of compatibility with traditional accounts of mindreading, Andrews' account is more critical of traditional understandings of social cognition. An in-depth analysis of these differences goes beyond the scope of this paper, where I will simply highlight their similarities and points of contact.

Wilkerson 2001; Bermudez 2003; Spaulding 2010). I will now turn to examine how these recent accounts of folk psychology can help account for this problem.

2.1 The normativity of the context: framing social cognition

As I anticipated above, traditional accounts of folk psychology see the understanding of other people as a form of mindreading: as we don't have access to other people's minds as we have to our own, understanding others relies on our epistemic capacities to infer what it is they are thinking. Folk psychology is understood as a primarily epistemic practice, reliant on the individual's own capacity for reasoning and comprehension. However, what such an approach leaves out is the importance of the sociocultural context where the "folk psychologist" (i.e. the person) lives and interacts with other people. What is known as the mindshaping approach, in contrast, argues that the context of the growth and interaction of the agent plays an active role in the formation of her cognitive capacities, and that it has a normative influence on the understanding of other's behavior.

McGeer (2015) gets this point across by considering the game of chess. To be a competent chess-player, understanding what your opponent is thinking is crucial. Understanding and predicting your opponent's next move requires learning the rules of the game and their correct application. This understanding is a know-how, a practical skill that can be developed and be subject to other players' feedback. Importantly, knowing how to play chess is an embodied skill: to apply the rules of the game one needs to develop other sensorimotor and visualization skills to engage with other players, the chessboard, the chess pieces and the patterns they afford. Understanding what other players are thinking and how they are strategizing is possible because both the actions of other players and your own competence at playing chess need to conform to the rules of the game. The normative infrastructure of chess makes available the beliefs, desires and intentions of the players in their chess-playing behavior, if one has enough expertise. Furthermore, if the players' behavior does not conform to the rules of chess, it will be difficult to make sense of it and, most importantly, it will be subject to correction in virtue of the norms dictated by the game.

McGeer (see also Zawidzki 2013, 2019) argues that social cognition works in the same way: to attribute beliefs, intentions and mental states in general to other people, the context where the interaction takes place has a normative role. Just as the rules of chess, there are norms underlying the different contexts of social interaction that mediate the expectations and interpretations of other's behavior. In every social context where interaction takes place, there are norms and conventions of appropriate and inappropriate behavior: human beings are raised in situated social environments characterized by (spatially and temporally variable) shared practices, norms and values. In other words, the context has a normative influence not only on how people behave, but on the interpretation of people's behavior. Whereas traditional accounts of mindreading focus on "inner" mechanisms (either the possession of a theory or the capacity to model others' mental states), mindshaping prompts to take into account "outer" dynamics, the social norms and conventions that reside, one could say, outside of one's own head, and in the context of the interaction. This shift of focus from the subject's mind alone to her interaction with preexisting shared social structure, complements traditional accounts of social cognition (Peters 2019, McGeer 2020) by providing a solution to the fame problem. Social practices, conventions and norms which characterize different contexts and communities help *framing* other people's behavior. Conformity (or subversion) to those norms makes that behavior readable, and such behavior is subjected to evaluation, interpretation and corrigibility to others in reference to those norms (McGeer 2007; Zawidzki 2013; de Bruin 2016; Zeppi and Bloekpoel 2017).

With context-dependent norms, here, I mean a wide array of social constructs and concepts. The kind of behavioral expectations dictated by the context include not only social norms and conventions, but even socially-dictated constructs such as stereotypes, virtues and character traits help framing mental states attribution. Mameli (2001) mentions as an example the results of experiments conducted by Condry and Condry (1976). In this experiment, a group of experimental subjects was asked to interpret the behavior of an infant who interacted with some toys, and who cried when opening a jack-in-a-box. Those who were told the infant was male replied that the infant cried due to anger, while those who were told the infant was female replied she did so due to fear. In this case, gender - a social construct that can vary across different cultural settings - creates an expectation in the folk psychologist that mediates her interpretation of the target's behavior. This concept, which does not refer to the state of mind of the target, constitutes a reference point for the folk psychologist's interpretation of the target's behavior. Social norms, conventions, values, stereotypes and identifiable character traits all narrow down the possible interpretations of a possibly ambiguous piece of behavior and constrain its possible interpretations by providing a mutually acknowledged and followed frame of reference.[11]

[11] It is relevant to note that the normative structures underlying interaction and framing social cognition (e.g. gender stereotypes) do not necessarily make the mental state attribution more *accurate*. They simply narrow down the possible number of interpretations by providing a frame.

The most important claim of the regulative view of folk psychology, then, is that shared and context-dependent norms frame and guide mental state attribution and, consequently, social cognition. This relevance of the situated context can overcome what I called the frame problem, because the contextual norms underlying interaction narrow down the interpretation of mental states. However, the importance of the embodied and embedded norms underlying interaction have a further implication. The example of stereotypes, virtues, moral and character traits, is meant to highlight not just their variability across time and cultures, but what can be called their *ideological* component. The normative assessment of other's behavior depends on the collective consideration of what standards of (in)appropriate or (im)moral behavior are. The perceived validity of a gender stereotype, for instance, relies on a collective agreement of how people of different genders behave; and while constituting a common ground that binds the members of the community's behavior and understanding, can be not only arbitrary but even expressive of forms of injustice and oppression towards the target (Haslanger 2020).

2.2 Expectation, enforcement and community

So far, I portrayed the role of context in social cognition as passive: shared social and cultural norms and structures work as an interpretative frame for enhancing mental state attribution. However, for some proponents of the mindshaping view, social norms, stereotypes and character traits do not simply work as a passive frame, but can also play an active role in mental state attribution.

In this view, certain social practices, like pedagogy, imitation and norm enforcement, do not simply involve mental state attribution: they enforce specific behavioral patterns in their targets, so that they conform to social and behavioral norms - so that, thanks to conformity to the social context, the target can make herself readable to folk psychologists. Teaching people how to behave in certain contexts or reprimanding a piece of behavior as inappropriate are a way of instilling specific behavioral patterns. To stick with the example of chess proposed by McGeer, to become a competent chess player, one needs to learn - or be taught - the rules of the game. Once the player learns the rules, not only she will be under external pressure to conform to them by other players: she will feel compelled to conform to them, because it is necessary to follow them to play the game at all. In the same way, during a university seminar, one student may feel compelled to raise her hand before asking a question.

Not only her interrupting without previous notice may hinder the seminar as a practice, but she learned to signal her intention to talk throughout her previous education. The normativity of the context in interaction and social cognition not only helps interpreting behavior, but can make people feel compelled to adhere to it for practical reasons - both for ensuring interaction to "go smoothly" and to conform to the expectations of the context.

Certain practices such as teaching and norm-enforcement, in other words, create an expectancy effect in their target: if one is taught what is the appropriate behavior in a certain situation, the target will be (and feel) expected to behave in the manner she is taught. That being said, there are cases when not just specific practices, but mental state attribution itself can create an expectancy effect. Mameli (2001) and Westra (2020) consider an experiment conducted by Miller et al. (1975) in this light. In the experiment, one group of school children was told they ought to be tidy, while a second group was told that they already were tidy. Both groups of students were equally tidy when the experiment commenced. After a few weeks, the researchers found that the children who were attributed the character trait of tidiness ended up conforming to such attribution, even though they were not especially tidy in the first place. The same did not stand for the children who were told they had to become tidier. The attribution of a virtue (tidiness) created an expectancy effect in the target (the children), which made them feel compelled to conform to it.

Zawidzki (2008, 2013, 2018) argues that this kind of mindshaping mechanism - that is, the creation of behavioral expectations and conformity in agents - is the most significant aspect of human sociality and cooperativity. While previous theories focusing on the evolution of human cognition saw cooperation and collective action as enabled by complex cognitive abilities (see e.g. Humphrey 1980; Tooby and Cosmides 1995; Sperber and Wilson 2002), Zawidzki argues that it is the need for cooperation among peers that enabled the development of sophisticated cognitive skills (see also Sterelny 2012; Geurts 2019, 2020; Fenici and Zawidzki 2020). Human beings need to establish a collective, shared normative framework for successful interaction: these norms constrain action and enable agents not just to interpret other's behavior, but to make oneself interpretable. The normativity of the shared and situated context, then, not only supports interpretability of other's behavior, but actively constrains behavior to be readable. Not only human beings integrate these constraints through teaching, imitation and norm enforcement, but even the attribution of a mental state or a character trait can lead to conformity to such attribution.

2.3 Community and rationality: context sensitivity and the role of reactive attitudes

So far, I highlighted not only how the behavioral expectations dictated by specific contexts enable mindreading, but also how certain practices not only enforce conformity to those norms but make their targets feel compelled to conform to them. I wish to then mention a relevant aspect of the way we enforce context-dependent expectations and we feel compelled to conform to them. Reactive attitudes, and the human sensitivity to reactive attitudes, pervade social interaction in general and illustrate some important aspects of contextually-bound social cognition. Namely, they highlight the sensitivity to normative assessments brought about by our peers, and how the integration of those norms leads to the evaluation of others as rational and moral agents.

McGeer (2015, 2018, 2020) takes the notion of reactive attitudes from Strawson (1981), intending with this notion "“the non-detached attitudes and reactions of people directly involved in transactions with each other … of such things as gratitude, resentment, forgiveness, love, and hurt feelings” (5). Strawson points out that these emotional states - reacting to others’ behavior through appraisal, disapproval, resentment and so on, they hold them to certain normative standards regarding (morally) appropriate and inappropriate behavior. When we assume these attitudes, we either expect our target to understand and conform to normative standards, or we are illustrating to her what those standards are and, eventually, why she should conform to them. For these reasons, McGeer argues, reactive attitudes play a key role in mental state attribution. If folk psychology is a normative practice; and the kind of assessment present in reactive attitudes implies a normative evaluation of behavior; then reactive attitudes can be considered as a way of structuring and directing the target’s behavior to a correct standard. Conformity to context-dependent norms matters for successful and smooth social interaction to take place, and not only the felt need to correct inappropriate behavior, but the sensitivity to these corrections are fundamental to understand others’ behavior and to make oneself understandable. The conformity to these behavioral standards, which are shared within a community, not only makes behavior readable and interpretable, but makes people accountable for that behavior depending on its (non-)conformity.

The connection of the notion of reactive attitudes with social cognition has a two-folded implication. On the one hand, the capacity of reading other’s behavior and being readable is deeply entangled with normative and moral evaluation of behavior. Social cognition is, as I stated in 2.1, embroiled in context-dependent norms and ideologies that can vary across populations.

On the other hand, the prominent presence of reactive attitudes highlights what might be called a natural sensitivity of human beings to normative evaluation and feedback. While the norms underlying social interaction and mental state attribution are context-dependent, the general sensitivity to norm infringement and the need for conformity to a social structure can be considered as general features of human beings. To assess the cognitive and moral capability of others - and to feel potentially subject to such evaluation - demonstrates the engagement with the social context and its normative structures, not just in terms of situatedness but also participation in a community of peers. If the competence of mental state attribution is comparable to learning and playing a game like chess, reactive attitudes can be considered the norm enforcement players assert to those who do not follow the rules. The possession of this competence is not "merely" epistemic, but intertwined with socially situated values shared within the community. To understand the moral and social norms that characterize the given community is integral to be treated as sensitive to reasons and actions that are grounded in those norms. The reactive attitudes that we adopt when others' behavior does not conform to our socially-grounded expectations are constitutive of mindshaping, for they are the means the other mind is "shaped".

As remarked by Haslanger (2020) the cultural and normative framework constituted by not just social norms but habits, conventions and descriptive norms provide tools for successful cooperation and coordination. Not simply joint action, but our mental life and our understanding of others' mental lives are determined within our own cultural milieu. To be sensitive to these values and norms is intertwined with the possession of the intelligent capacities necessary for understanding other people and interacting with them successfully. Given the importance of cooperation and mutuality among human beings, the skill required for appropriate mental state attribution implies the possession of an essentially interpersonal capacity. In light of this, reactive attitudes can be considered as the means of scaffolding people's behavior into not simply adequacy but mere comprehensibility.

3. Mindshaping jerks: the folk psychology of online hostility and moral disengagement

Let us recap the main points of the previous sections before proceeding to the bulk of my argument. In section 1, I argued that:

1) In SNSs, users are constituted and perceived as information;
2) In SNSs, the context of interaction is undetermined and left to users' imagination;

In section 2, I argue that:

3) Social cognition in face-to-face settings is framed by the situated context, which moderates behavioral expectations and makes people and their minds readable;
4) Occasionally, people will be and feel obliged to conform to the expectations dictated by the normative context;
5) The social context, in face-to-face interaction, is composed by social (and moral) norms, conventions, values and shared beliefs agents deem appropriate, and individuals' competence in following and mutually understanding those norms is the presupposition for social cognition.

The contrast between 1)-2) and 3)-5) not only highlights the most radical differences between offline interaction and interaction on mainstream social media such as Facebook and Twitter[12]. This contrast is key to understanding the cognitive dynamics that enable toxic and aggressive interaction to arise in digital environments, which emerge from a radical transformation of the context where interaction occurs. The design of mainstream SNSs leads, I argue, to a lack of a shared normative context, which is a consequence of the user-centeredness of these platforms made evident by cyborgification and context collapse.

I will now turn to a careful analysis of how cyborgification and context collapse cause a general epistemic uncertainty in SNSs users regarding the interpretation of other people's motivations and underlying mental states. I will then describe how, due to context collapse, SNSs users need to refer to their own standards for appropriate or inappropriate behavior to assess other's behavior and their underlying mental states, and how the lack of common ground can easily lead to dehumanizing practices.

[12] I will focus on mainstream social media such as Facebook and Twitter because they are strongly characterized by context collapse, more so than more specific social media platforms. I will return to this point in section 4.2.

3.1 From Lack of Frame to Individualized Context

In face-to-face interaction, the situated, embodied and embedded context is fundamental for mutual understanding and joint action. However, what happens when everyday mindreading occurs on SNSs? As I explained in section 1, the use of SNSs is characterized by two aspects relevant for social interaction. The first is cyborgification, the merging of a human agent with the technology in question. This integration has a two-folded implication. On the one hand, the user-centeredness of the platforms' design provides users with a sense of ownership and control, as the newsfeed is her own. On the other hand, cyborgification implies that other users are not just indistinguishable from, but defined by their activity on the platform, by their connections and generated content. The second aspect that transforms social interaction and cognition is context collapse, the indetermination of social contexts and backgrounds into one. The different people that are part of social contexts which, in face-to-face settings, would be spatially and temporally separated, become merged and integral part of the platform's ecology.

The merging of different people (cyborgification) and of the contextual expectations that those people embody (context collapse) are extremely significant in the transformation of social interaction in online settings. I propose that these two design features cause users to experience one of the traditional criticisms of traditional accounts of folk psychology, that is the frame problem. The frame problem of folk psychology, as I stated in section 2, problematizes traditional mindreading accounts of social cognition by highlighting how the same piece of behavior can be given different interpretations in terms of beliefs, desires, intentions and emotional states. The regulative view of folk psychology provides an answer to this problem by highlighting the normative role of the social context, which constrains people's behavior and, consequently, informs and narrows down the possible interpretations of their behavior. The social context, whose normativity is ensured by the community who inhabit it - by people who also follow, enforce and illustrate those norms when needed - is fundamental in mediating behavioral expectations and making the reasons and mental states underlying behavior less equivocal.

Online context collapse, as discussed in section 1.2, implies a general indeterminacy of the context of the interaction, where the boundaries of private and public space, between members of one's own social network and outsiders, and between separated offline contexts become blurred. To put it differently: SNSs do not have an undisputedly defined, shared context. If the social context in face-to-face interaction provides a frame and a common ground, then on SNSs social interaction and cognition is *groundless*. One piece of behavior - i.e. one post or comment - is not framed by mutually shared norms or embodied expressivity as it would be in ordinary face-to-face interaction, and its motivations and background are not accessible to the interpreter. Undisputedly and mutually shared norms for interaction and communication are absent in SNSs such as Facebook or Twitter, due to context collapse. While in face-to-face interaction there are contextual, shared norms that mediate people's behavior and their interpretation of behavior, online these norms are absent and make behavior (i.e. generated content) much more ambiguous to understand. In other words, in mainstream SNSs users incur into the frame problem of folk psychology: it is intrinsically more complex to interpret the meaning of a piece of behavior, which may have divergent underlying motivations and be produced under circumstances inaccessible to the interpreter. The absence of a shared context entails the absence of a reference frame for interpreting behavior.

That being said, the lack of an undisputed frame does not (obviously) deny the capacity to interpret behavior entirely. Rather, this re-presentation of the frame problem - that is, the intrinsic difficulty to frame the beliefs, desires, intentions, motivations or emotional states underlying an agent's behavior - has a significant impact on the way SNSs users understand behavior. Users find themselves in a contradictory scenario: they are in a user-centered *and yet* interpersonal space, their access to the platform and its configuration is their own, and yet other users and information sources are not clearly determined. Information consumption, social interaction and mental state attribution have intrinsically ambiguous premises, and the agents that carry out these practices - the users of the platform - must somehow solve this inherent tension.

If face-to-face interaction and mental state attribution requires the presence of a social context with normative value that can frame and make people's behavior understandable; and if online such a frame is ultimately lacking due to a fundamental ambiguity intrinsic to the platform's design; then users need to implement a reference frame of their own. The ambiguity and informational scarcity that characterize SNSs leads users to have to make do with what they have. Firstly, and most intuitively, the object of the user's (the folk psychologist's) cognition is transformed. In face-to-face settings, an agent (the target of the folk psychologist's mental state attribution) and the behavior that agent produces (the interpretandum, if you will) are generally distinguishable from one another. Cyborgification online has the implication of blurring this distinction between agent and behavior, because the content produced by the agent and the agent herself are perceived as one[13]. However, cyborgification and context collapse have a further implication for social cognition and interaction, one that affects the folk psychologist herself and not simply the targets of her mindreading practices. On mainstream SNSs, users supplement the absence of shared norms by applying their own, offline norms for appropriate and inappropriate behavior.

The application of one's own, imagined norms of appropriate and inappropriate behavior might seem, in itself, unjustified. After all, while SNSs such as Facebook and Twitter are extremely user-centered, they are also fundamentally interpersonal spaces, where users interact with people they may or may not know. However, the lack of justification in this behavior does not entail its irrationality. As Rini (2017) points out, because the norms of communication of social media are disputed and there is no common understanding of why people share and generate content (as I elaborated so far), users of social media find themselves in a generalized epistemic uncertainty. The motivations, intentions, beliefs or desires underlying sharing and generating content are not immediately available and lack a shared interpretative frame. For Rini, taking a partisan stance towards seemingly immoral content - and specifically, sticking to one's own normative frame of reference, one's own values and standards for (in)appropriate behavior - can be considered a rational choice. Exactly because there is no undisputed reference frame, SNSs users may choose to refer to their own. Taking a partisan stance, by reiterating one's own values and standards as appropriate and informational behavior (generated content) that contradicts them as inappropriate, means to make explicit one's own frame of reference.

To understand the rationality - or at least the "reasonability" - of applying one's own frame to assess and interpret others' behavior online, it is important to note the role of the design in this choice. The user-centeredness of mainstream SNSs like Facebook and Twitter plays a significant role in framing others' behavior. On Facebook's homepage, the platform invites the user to generate content with the sentence "What's on your mind?". The platform is designed in such a way that users are invited to express their thoughts, to share what they want, to express their own values and personality. Nelson (2018) expresses this idea best when she writes that "Social media allows us to persistently emphasize who we are and set aside the question of what we are altogether" (178).

Because SNSs are so extremely focused on the user, on what she thinks and what she values, one might say that to apply one's own standards for assessing appropriate or

[13] One might argue that the producer of the content is simply unseen, instead of merged with the generated content. This interpretation does not alter my general argument, as the producer of the content is still the target of the mental state attribution.

inappropriate behavior is rational: rather, it is *actively encouraged* by the platforms' design itself. Because the design of the SNSs promotes the expression of one's own values, other people's generated content will be interpreted by default as a form of assertion or endorsement. This occurs not only because, as Rini points out, there are no undisputed communicative norms to determine why people post or share content. The problem is that mainstream SNSs such as Facebook are designed to draw out each individual user's opinions and thoughts. The extreme user-friendliness, or user- centeredness of mainstream SNSs, is designed to draw out the expressivity of individuals using the platform. However, the extreme focus on the user-centeredness of the platform clashes with the collapsed and undetermined context that characterizes interaction in these websites. The content of a user's feed is (at least promoted as) tailored for her: the feed is her own, with content that either for engagement by members of her online social network or by algorithmic suggestion is proposed for her own consumption. However, this way of personalized framing of others' generated content, which is a result of a design choice - context collapse - encourages users to apply their own normative frame to interpret this content. The content that seemingly contravenes this frame will not simply be perceived out of context, but unjustified. I will now turn to analyzing the consequences of the perceived out-of-context-ness for users' cognition and reactions.

3.2 Mindshaping Jerks: Online Incivility as Groundless Reactive Attitudes

Online incivility and toxic disinhibition are a widely shared and investigated concern among academics. On the one hand, the lack of embodied cues and feedback is generally thought to play a role in toxic disinhibition and aggression. On the other hand, polarization is a more generalized (and not exclusively academic) concern regarding the way people debate online: there seems to be a generalized belief about SNSs, that people tend to be more radicalized in online debates, especially when it comes to divisive societal issues.

While there is a general awareness of this issue, my discussion of the implications of context collapse and cyborgification in social cognition highlight a further, social and cognitive aspect of toxic interaction. This aspect can be summarized as follows: because there is no undisputed norms of communication and interaction online, and because, *by design*, users need to apply their own normative framework to interpret others' behavior, users will feel compelled to counter and react negatively to behavior that (seemingly) contradicts their own normative framework. Not only, as I explained in section 2, are people tendentially sensitive to enforce these normative standards when somebody infringes them - in actuality, the SNSs' interface invites users to express those standards. The normative frame that underlies users' behavior and their interpretation of behavior, however, can vary greatly across users and their situated, offline context. For this reason, users will tend to reciprocally enforce their own normative frames on each other, to take on reactive attitudes towards each other's behavior when they consider it immoral, unreasonable or unjustified, while there is no shared frame where such reactive attitudes can be considered themselves justified or reasonable.

What emerged from the analysis of mindshaping and of the technologically-mediated frame of experience provided by the platform is the contextual sensitivity to social and moral norms. This sensitivity is an incredibly important characteristic of human beings, as it is a key enabler of the high degree of cooperativity that characterizes our species. It is this sensitivity that prompts us to assume reactive attitudes towards those who infringe the contextual norms of our interaction. In the technological settings of SNSs, however, not only there are no undisputedly shared norms among agents: the hyper-individualized framing of interaction and information consumption of mainstream SNSs has the consequence of individualizing agency, and to cause other agents to feel as mere part of the digital environment. It is for these reasons that, in front of the collapsed context, users are prompted to apply their own normative standards of appropriate behavior.[14]

The absence of shared values of interaction, and the application of each user's normative framework to make sense of each other's behavior, potentially leads to what I call the practice of *mindshaping jerks*. With this term, I refer to the transforming of people's expectations and evaluation of others by treating them as immoral agents, by enforcing onto them one's own standards of appropriate or inappropriate behavior when those standards are not perceived as such by the targets of the enforcement.

Recall that mindshaping, which is enabled by our natural sensitivity to norms and is constituted by our reactive attitudes, consists in a normative practice with the intent or effect of prompting the target to conform to socially-grounded expectations. Online, an individual user's expectations for others' behavior is grounded in their own, offline social context. Following Schwitzgebel (2019), I define a *jerk* as someone who "culpably fails to appreciate

[14] One may argue that the design of mainstream SNSs is actually the exploitation of this sensitivity, and that the service providers actually aim to cause user engagement with the platform by taking advantage of this sensitivity. This point could be further elaborated by connecting it with the practice and finalities of profiling through engagement (Hildebrandt 2008,) This is a very relevant point that would deserve much deeper research on its own, and that cannot be explored appropriately here.

the perspectives of others around him, treating them as tools to be manipulated or fools to be dealt with rather than as moral and epistemic peers" (4-5) and I assume that one may behave like a jerk while not deliberately or explicitly intending to harm others (10-12). This definition applies well to the case at hand. If SNSs prompt users to apply their own normative frame to otherwise decontextualized behavior (section 3.2); and if this technologically- encouraged way of seeing others' behavior is effective because of the natural sensitivity to assess others' moral competence based on some given normative framework (sections 2.2-2.3); then the online behavior of people who do not conform to our standards will be instinctively assessed as morally defective, and the person who produced that behavior will be considered morally and epistemically inferior.

From an individual folk psychologist's point of view, the absence of a unified frame, which agents' behavior can refer to in order to be readable at all, has the implication of (mis)representing behavior that follows different and unseen normative frames as morally and epistemically incompetent, and will lead users to treat the others as such. Our reactive attitudes towards those pieces of behavior, our attempts to "mindshape" others' behavior into conformity are not simply unjustified and ineffective due to the absence of a shared frame. The other person will herself feel called out without a justification, and consider your behavior immoral and unjustified; and, consequently, she will herself react to your behavior by assessing it as immoral and assessing you as the morally and epistemically incompetent agent. The application of each other's frames entails a lack of common ground, and interaction becomes dysfunctional and, oftentimes, violent. Online aggression does not simply occur due to issues that are by themselves polarizing, but because the absence of a shared frame means that a much more socio-cognitive cognition that enables any form of mutual understanding and joint action is gone.

Put differently, one could understand the kind of interaction described above as a reciprocal attempt at correcting the other person's behavior to one's own normative frame. The adoption of reactive attitudes towards the target can be understood as an attempt at shaping the other person's mind into one's normative standards, which are not necessarily shared by the online other. While the initial adoption of (negatively charged) reactive attitudes towards others may rise due to a general informational scarcity, its (at least tentatively) regulatory aim may be seen as somewhat pedagogical: it consists in an attempt to correct others' behavior and state of mind to normative standards that *we* follow, but *they* do not necessarily do. This lack of a common ground constitutes an impediment for functional interaction and mutual understanding. However, due to the context sensitivity of human beings, users feel compelled to react to behavior that (seemingly) does not follow norms of appropriate behavior. The incongruence of the interaction caused by the lack of common ground leads not simply to the projection of one's own normative frame to other people's behavior, but to feeling the need to correct that behavior to one's own standards. These attempts at evaluating and appraising others' behavior will be inefficient due to a lack of an actually shared frame, which will result in a frustrating interaction overall and, tendentially, to a reduced capacity to actually understand the actions and state of mind of the online other.

This explanation of how polarization and hostility rise in online environments is justified by the discussion, in section 2, on the importance of the socially situated context in interaction and mental state attribution, and by the impact of cyborgification and context collapse in user activity on SNSs. While I believe that many empirical studies on online incivility and hostile interaction may somewhat implicitly presuppose such a picture, I believe that making this aspect explicit can enrich our understanding of online toxic behavior.

In the various research on incivility and dehumanization in SNSs there is some degree of awareness regarding the technological aspects of these digital platforms that enable or facilitate aggression and the perception of "the other side" as morally distant. The lack of embodied presence that comes with computer-mediation reduces the possibility of empathy among agents involved in the interaction, without fear of bodily harm or undisputedly effective sanctioning. Other people's speech acts and generated content is also perceived as truncated, in contrast to the spatiotemporal continuum typical of face-to-face interaction. However, the examination of moral disengagement in online interaction in light of the regulative view of folk psychology - i.e., by considering the way specific design features of mainstream SNSs supplant the absence of a shared context with a hyper-individualized perspective - shows that the influence of these platforms runs much deeper than a mere removal of embodied presence. By valuing the individual's own values and preference before all else, the platform's very design leads to the individual's application of one's own normative context, naturally distorting not just the moral judgment of others' behavior and what we consider (our own) moral behavior, but the very possibility of understanding behavior or what the people behaving that way are thinking, desiring and valuing at all. The lack of a determined normative context for interaction, which is the result of the platform's underlying ideology, prevents the very possibility of an accurate understanding of other people's intentions, beliefs and general state of mind.

4. Online Virtues and the Importance of Contextual Norms

The diffusion of epistemically problematic content on SNS, as well as the higher degree of polarized interaction, are uncontroversially considered a problem both in academia and the general public.[15] Given the generalized concern regarding polarization online, one of the main lines of solutions to this complex issue consists in the ideation and application of adequate media literacy, whereby media literacy is understood as the acquisition of "the critical knowledge and the analytical tools that will empower media consumers to function as autonomous and functional citizens" (Khan 2008, 15). The development of this kind of capacity is meant to enable media consumers - or, in this case, SNSs users - to consume and interact with the media and other citizens in functional ways, conscientiously and healthily. Given the success and diffusion of SNSs and the attention given to polarization, aggression and generally toxic interaction in these digital environments, the acquisition and practice of these tools is considered key in order to avoid these issues.

Some authors propose that these kinds of media policy should promote virtues in Internet (and SNSs) users. For instance, Heersmink (2016, 2018) proposes that digital media literacy policies should support the development of epistemically responsible behavior. Specifically, he proposes that Internet users should develop intellectual virtues (intended as acquired or learned character traits) such as intellectual humility (i.e. the acknowledgement of one's own cognitive limitations), carefulness (i.e. trying to avoid intellectual errors and mistakes) and thoroughness (i.e. the disposition to probe for a deeper understanding of the information at hand, its origins and implications). Similarly (and more closely related to this paper's concern), Worden (2019) argues that, for interaction and debate on social media to be successful and not be morally disengaged, there is need for developing civic friendship, intended as a form of mutual respect for other (and potentially unknown) SNS users. Here, civility is enabled by mutually shared and agreed upon knowledge regarding each other's social and political community. Without such shared knowledge, Worden argues, any form of cooperation (including political debate) will be inevitably dysfunctional if not outright unachievable, as the members of the community will lack a sense of mutual understanding. Civil social media-based debate can be properly maintained only insofar SNSs users exercise of inclusivity regarding the value of other viewpoints, self-control in the way and tone of expression, discretion in choosing the debate to partake in, and audience sensitivity (i.e.

[15] See, however, Peters (2020), who warns that claims about the persistence of online polarization may inadvertently cause users to expect a polarized environment and act accordingly.

recognizing the moral and political character of both the people actively taking part in the debate and of those who may simply monitor the conversation without jumping in).

I believe that the approaches to online incivility and more general epistemological problems related to social media use that focus on the importance of exercising virtues have some relevant merits. However, this kind of approach, taken at face value, runs into two problems. The first issue regards the general feasibility of user-centered media literacy policies. On the one hand, the enactment and eventual success of a policy supporting media-related education that supports intellectual and civic virtues would take time. This can be particularly problematic if one assumes that digital technologies may develop and transform fast enough to make those policies too obsolete. On the other hand, while, as Heersmink (2016) points out, all citizens should have equal digital literacy skills or at least an equal chance to develop them, an effective policy would have to effectively deal with various levels of what is called "the digital divide", i.e. social inequalities in access and use of informational resources (see e.g. Notten et al. 2009; van Deursen and van Dijk 2019). In general, media literacy policies would have to account for the differences caused by different economic, cognitive and sociocultural resources, not only in physical (access to the Internet) and material access (the kind of device used), but also in digital skills and in the capacity to translate Internet usage into favorable offline outcomes.

The second problem is somewhat more complex, and regards not just the effectiveness of user-centered digital literacy policies, but also whether focusing only on the user is the most appropriate way to tackle the problem of unhealthy online interaction and information consumption. As Brown and Hennis (2019) point out, the outsourcing of responsibility to users only for aggression and abusive behavior online can be seen as a way for service providers to not take responsibility in moderating content. The consideration of digital platforms (and their moderators) as value-neutral when it comes to toxic behavior not only fails to properly safeguard its victims, but fails to consider the way the platforms' design choices (or lack thereof) enable or influence such behaviors.

In light of my previous discussion of social cognition online, the role of the platform's design in toxic interaction runs extremely deep. If the capacity to understand each other's minds is fundamentally enabled by the presence and joint adequacy to socially shared contextual norms, and if on platform such as Facebook and Twitter these norms are left undetermined by context collapse - which is a design choice - then user-centered literacy policies cannot tackle the issue of uncivil and toxic interaction properly. The configuration of mainstream SNSs, which prompts users to express their own thoughts and values and on framing other people's behavior through their own normative standards, distorts what would otherwise be an essential component of everyday social cognition. In other words, any policy decision that intends to tackle the problem of aggression and hostile interaction online needs to take into account the features of the platform's design that actively shape user experience. Interventions for making social interaction online healthier and functional need to consider the social and cognitive implications of SNSs' design choices; and while the rendering of users into cyborgs is somewhat inevitable[16], intervention on the collapsed context of online interaction represents a more viable path for positive change.

A stronger determination of the context of social interaction, so that the audience and the norms of appropriate behavior are not left to users' imagination, can play a key role in steering users' behavior. As Gunn (2020) argues, forms of responsible agency intended to promote productive (or, in our case, civil) communication needs to be sensible to the context that mediates our conception of one another as competent agents and communicators. In this sense, focusing on personal behavioral and belief-formation regulation needs to be grounded to a joint commitment of all participants of the interaction to mutually acknowledged ends. For the promotion of civil behavior as the one proposed by Heersmink and Worden there must be some mutually binding norms that can drive interaction in a healthy way. Similarly, Miller and Record (unpublished) propose that setting up explicit norms around responsibility for the content and tone of generated content (either by communities of users or by the service providers themselves) can help a healthy management of toxic content. Furthermore, platforms might offer users more options to contextualize their posts by making their reasons and feelings around the production and sharing of content more salient.

The establishment of contextual norms is essential for the implementation of policies that are aimed at the development of online virtues, as it provides a way to properly take into account that some structural design choices of digital platforms have a relevant influence in shaping behavior. The way users are meshed with the platform's configuration, and the (lack of) determination of the audience are constitutive of the way they interact with information and others online. If the entire ideology behind the design of a platform like Facebook is self-affirmation and expression above all else, these design choices not only will drive users' cognitive and affective processes online: due to the specific meshing of user and technology in the collapsed context, the configuration of this platform can be considered constitutive of their moral and cognitive character (Skorburg 2019). And to develop epistemic and civic

[16] I.e. I am not aware of any way to accurately replicate the embodiment of everyday face-to-face interaction in computer-mediated communication.

virtues, there must be a stronger focus not just on individual users, but on the entire sociocultural and cognitive ecology of the platform they use (Phillips and Milner 2021), to give them the possibility to put those virtues into practice within a community that values those behavioral patterns.

The discussion of mindshaping highlighted the importance of the contextual, communally shared norms that underlie social interaction in face-to-face settings for interpreting behavior, a context that in mainstream SNSs is by definition collapsed. The moderation of behavior through context-dependent expectations and obligations that those norms provide do not simply determine the values of a community, but make the behavior of that community's members understandable to begin with. To focus on possible norms of interaction and communication online that can be jointly accepted and followed can therefore be a viable path to make us better and healthier people in the online world, by properly tackling a problem that lies at the intersection of technology, cognition and sociality.

5. Conclusion

In this paper, I proposed to examine the problem of online hostility and aggression through the theory of social cognition known as mindshaping. This theory accounts for a key role of contextual and socially shared norms for the understandability of each other's behavior. By describing some structural design features of mainstream social networking sites, I argued that the design of social media such as Facebook and Twitter overfocus on the user's self-expression and leave the context of social interaction and information consumption severely under- determined. This underdetermination of the social context, I argued, leads users to interpret each other's behavior through their own normative frame, leading to a severe reduction of this fundamental socio-cognitive aspect of social interaction, and to a distorted understanding of others and to hostility and aggression. For this reason, I concluded that it is necessary to instantiate a sense of community and mutually shared norms through a clearer determination of the context of interaction to properly tackle the problem of hostile interaction. While the precise way of establishing contextual norms online needs much further exploration, I believe it can be a fruitful path that literacy policies and digital designers can follow to improve the functionality of online interaction.

Bibliography

Andrews, K. (2009). Understanding Norms Without a Theory of Mind. *Inquiry* 52, 433-448. https://doi.org/10.1080/00201740903302584

Andrews, K. (2012). *Do apes read minds? Toward a new folk psychology*. Cambridge, MA: MIT Press.

Antoci A, Delfino A, Paglieri F, Panebianco F, Sabatini F (2016). Civility vs. Incivility in Online Social Interactions: An Evolutionary Approach. *PLOS ONE* 11(11): e0164286. https://doi.org/10.1371/journal.pone.0164286

Aro, J. (2016). The cyberspace war: propaganda and trolling as warfare tools. *EUROPEAN VIEW* 15: 121-132. DOI:10.1007/s12290-016-0395-5

Aydin, C., González Woge, M. & Verbeek, P. (2019). Technological Environmentality: Conceptualizing Technology as a Mediating Milieu. *Philosophy & Technology* 32: 321–338. DOI:10.1007/s13347-018-0309-3

Bail, C.A., Guay, B, Maloney, E., Combs, A., Hillygus, S.S., Merhout, F., Freelon, D. and Volfosky, A. (2020). Assessing the Russian Internet Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. *PNAS* 117(1): 243-350. DOI:10.1073/pnas.1906420116

Beam, M. A., Child, J. T., Hutchens, M. J., & Hmielowski, J. D. (2018). Context collapse and privacy management: Diversity in Facebook friends increases online news reading and sharing. *New Media & Society* 20(7): 2296–2314. https://doi.org/10.1177/1461444817714790

Bermúdez, J. (2003). The Domain of Folk Psychology. *Royal Institute of Philosophy Supplement* 53: 25-48. doi:10.1017/S1358246100008250

Bertolotti T, Arfini S, Magnani L. (2017). Of Cyborgs and Brutes: Technology-Inherited Violence and Ignorance. *Philosophies*: 2(1):1-14. https://doi.org/10.3390/philosophies2010001

Brown, J.J. Jr., and Hennis, G. (2019). Hateware and the Outsourcing of Responsibility. In Reyman, J., & Sparby, E. (Eds.) *Digital Ethics: Rhetoric and Responsibility in Online Aggression* (1st ed.) (18-31). Routledge.

Cho, D., and Kwon, K.H. (2015). The impacts of identity verification and disclosure of social cues on flaming in online user comments. *Computers in Human Behavior* 52(A): 363-372. https://doi.org/10.1016/j.chb.2015.04.046.

Clark, Andy (2003). *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press.

Costa, E. (2018). Affordances-in-practice: An ethnographic critique of social media logic and context collapse. *New Media & Society* 20(10): 3641–3656. https://doi.org/10.1177/1461444818756290

D'Errico, F., & Paciello, M. (2018). Online moral disengagement and hostile emotions in discussions on hosting immigrants. *Internet Research* 28: 1313–1335. https://doi.org/10.1108/IntR-03-2017-0119.

Dubois, E., Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society* 21(5): 729-745. https://doi.org/10.1080/1369118X.2018.1428656

Fenici, M., and Zawidzki, T. (2020). The origins of mindreading: how interpretive socio-cognitive practices get off the ground. *Synthese*: 1-23. DOI: 10.1007/s11229-020-02577-4

Fernandez Castro, V. (2020). Regulation, Normativity and Folk Psychology. *Topoi* 39: 57–67. https://doi.org/10.1007/ s11245-017-9511-7

Flaxman, S., Goel, S., and Rao J.M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quart.* 80(S1): 298–320. https://doi.org/10.1093/poq/nfw006

Floridi, L. (2017). Digital's Cleaving Power and Its Consequences. *Philos. Technol.* 30: 123–129. https://doi.org/10.1007/s13347-017-0259-1

Gallagher, S. (2001). The practice of mind: Theory, simulation or primary interaction? *Journal of Consciousness Studies* 8 (5-7): 83-108. DOI: 10.1.1.710.5008

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition* 17(2): 535-543. https://doi.org/10.1016/j.concog.2008.03.003

Gallagher, S., and Hutto, D. D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), *Converging evidence in language and communication research (CELCR): Vol. 12. The shared mind: Perspectives on intersubjectivity* (17–38). John Benjamins Publishing Company.

Gallagher, S., Varga, S. (2014). Social Constraints on the Direct Perception of Emotions and Intentions. *Topoi* 33: 185–199. https://doi.org/10.1007/s11245-013-9203-x

Gallese, V.M.D. (2009) Mirror Neurons, Embodied Simulation, and the Neural Basis of Social Identification. *Psychoanalytic Dialogues* 19(5): 519-536. DOI: 10.1080/10481880903231910

Geurts, B. (2019). Communication as commitment sharing: Speech acts, implicatures, common ground. *Theoretical Linguistics* 45(1–2): 1–30. https://doi.org/10.1515/tl-2019-0001

Goldman, A. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press USA.

Gunn, H. K. (2020). How Should We Build Epistemic Community? *Journal of Speculative Philosophy* 34 (4): 561-581. https://doi.org/10.5325/jspecphil.34.4.0561

Harel, T. O., Jameson, J. K., and Maoz, I. (2020). The Normalization of Hatred: Identity, Affective Polarization, and Dehumanization on Facebook in the Context of Intractable Political Conflict. *Social Media + Society*. DOI: 10.1177/2056305120913983

Hart, P. S., Chinn, S., & Soroka, S. (2020). Politicization and polarization in COVID-19 news coverage. *Science Communication* 42(5): 679-697. https://doi.org/10.1177/1075547020950735

Haselager, W. F. G. (1997). *Cognitive science and folk psychology: the right frame of mind*. Sage Publications.

Haslanger, S. (2020). Cognition as a Social Skill. *Tandf: Australasian Philosophical Review* 3(1): 5-25. https://doi.org/10.1080/24740500.2019.1705229

Heersmink, R. (2016). The internet, cognitive enhancement, and the values of cognition. *Minds and Machines* 26(4): 389-407. DOI:10.1007/s11023-016-9404-3

Heersmink, R. (2018). A virtue epistemology of the Internet: Search engines, intellectual virtues, and education. *Social Epistemology* 32(1): 1–12. https://doi.org/10.1080/02691728.2017.1383530

Humphrey, N. (1980). Nature's psychologists. In: B.D. Josephson and V.S. Ramachandran (eds.), *Consciousness and the physical world* (57-80). Oxford: Pergamon Press.

Hutto, D. D. (2008). *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons*. Bradford.

Ihde, D. (1990). *Technology and the lifeworld*. Bloomington/Minneapolis: Indiana University Press.

Iyengar, S., & Massey, D. S. (2019). Scientific communication in a post-truth society. *Proceedings of the National Academy of Sciences of the United States of America, 116*(16), 7656-7661. https://doi.org/10.1073/pnas.1805868115

Khan, A. W. (2008). Media Education, a Crucial issue in the Building of an Inclusive Knowledge Society. In Carlsson U., Tayie S., Jacquinot-Delaunay G. & Perez Tornero J.M. (Eds.), *Empowerment through Media Education. An Intercultural Dialogue* (15-18). Gothenburg: International Clearinghouse on Children, Youth and Media, in co-operation with UNESCO, Dar Graphit and the Mentor Association.

Kiesler, S., Siegel, J., and McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist* 39, 1123–1134. https://doi.org/10.1037/0003-066X.39.10.1123

Kim, E.M., and Ihm, J. (2020). Online News Sharing in the Face of Mixed Audiences: Context Collapse, Homophily, and Types of Social Media. *Journal of Broadcasting & Electronic Media* 64(5): 756-776. DOI: 10.1080/08838151.2020.1835429

Kiri Gunn, H. (2020). How Should We Build Epistemic Community? *Journal of Speculative Philosophy* 34 (4): 561-581. https://doi.org/10.5325/jspecphil.34.4.0561

Kirkpatrick D. (2010). *The Facebook Effect: The Inside Story of the Company that is Connecting the World*. New York: Simon and Schuster.

Kowalski, R. M., Giumetti, G., Schroeder, A., & Lattanner, M. (2014). Bullying in the digital age: a critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin* 140: 1073- 1137. DOI: 10.1037/a0035618

Langos C. (2012). Cyberbullying: the challenge to define. *Cyberpsychology, behavior and social networking* 15(6): 285–289. https://doi.org/10.1089/cyber.2011.0588

Lapidot-Lefler, N., and Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior* 28(2): 434–443. https://doi.org/10.1016/j.chb.2011.10.014

Maher, P., Igou, E., and van Tilburg, W. (2018). Brexit, Trump, and the polarizing effect of disillusionment. *Social Psychological & Personality Science* 9(2): 205-213. https://doi.org/10.1177/1948550617750737

Mameli, M. (2001). Mindreading, mindshaping, and evolution. *Biology and Philosophy* 16(5): 597–628. https://doi.org/10.1023/A:1012203830990

Marwick, A. E., & boyd, danah. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society* 13(1): 114–133. DOI:10.1177/1461444810365313

Marwick, A. E., & boyd, danah. (2014). Networked privacy: How teenagers negotiate context in social media. *New Media & Society* 16(7): 1051–1067. https://doi.org/10.1177/1461444814543995

Masullo Chen, G., Muddiman, A., Wilner, T., Pariser, E., and Stroud, N. J. (2019). We Should Not Get Rid of Incivility Online. *Social Media + Society*. https://doi.org/10.1177/2056305119862641

McGeer, V. (2007). The regulative dimension of folk psychology. In D. D. Hutto & M. Ratcliffe (Eds.), *Folk psychology re-assessed* (137–156). New York: Springer.

McGeer, V. (2015). Mind-making practices: the social infrastructure of self-knowing agency and responsibility. *Philosophical Explorations* 18: 259 - 281. https://doi.org/10.1080/13869795.2015.1032331

McGeer, V. (2019). Scaffolding agency: A proleptic account of the reactive attitudes. *Eur J Philos.* 27: 301–323. DOI:10.1111/ejop.12408

McGeer, V. (2020). Enculturating folk psychologists. *Synthese*: 1-25. https://doi.org/10.1007/s11229- 020-02760-7

Miller, B. (2021). Is Technology Value-Neutral? *Science, Technology, and Human Values* 46 (1): 53-80. https://doi.org/10.1177/0162243919900965

Miller, B., and Record, I. (unpublished). People, Posts, and Platforms. Reducing the spread of online toxicity by contextualizing content and setting norms.

Miller, R. L., Brickman, P., & Bolen, D. (1975). Attribution versus persuasion as a means for modifying behavior. *Journal of Personality and Social Psychology* 31(3): 430–441. DOI: 10.1037/h0076539

Moor, P.J., Heuvelman, A., and Verleur, R. (2010). Flaming on YouTube. *Comput. Hum. Behav.* 26(6): 1536–1546. https://doi.org/10.1016/j.chb.2010.05.023

Nelson, L. (2019). *Social Media and Morality. Losing Our Self Control.* Cambridge: Cambridge University Press.

Nichols, S., and Stich, S. P. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford University Press.

Nisbet, E. C., Cooper, K. E., and Garrett, R. K. (2015). The partisan brain: How dissonant science messages lead conservatives and liberals to (dis)trust science. *The Annals of the American Academy of Political and Social Science,* 658(1): 36-66. https://doi.org/10.1177/0002716214555474

Notten, N., Peter, J., Kraaykamp, G., and Valkenburg, P.M. (2009). Research Note: Digital Divide Across Borders—A Cross-National Study of Adolescents' Use of Digital Technologies. *European Sociological Review* 25(5): 551–560. https://doi-org.ru.idm.oclc.org/10.1093/esr/jcn071

Pacilli, M. G., Roccato, M., Pagliaro, S., & Russo, S. (2016). From political opponents to enemies? The role of perceived moral distance in the animalistic dehumanization of the political outgroup. *Group Processes & Intergroup Relations*, 19(3), 360–373. DOI: 10.1177/1368430215590490

Peters, U. (2019). The complementarity of mindshaping and mindreading. *Phenomenology and the Cognitive Sciences* 18(3): 533–549. https://doi.org/10.1007/s11097-018-9584-9

Peters, U. (2021). How (Many) Descriptive Claims About Political Polarization Exacerbate Polarization. *Journal of Social and Political Psychology* 9(1): 24-36. https://doi.org/10.5964/jspp.5543

Phillips, W. (2015). *This Is Why We Can't Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture.* The MIT Press.

Phillips, W., and Milner, R. (2017). *The Ambivalent Internet. Mischief, Oddity, and Antagonism Online*. Polity Press

Phillips, W., and Milner, R. (2021). *You Are Here. A Field Guide for Navigating Polarized Speech, Conspiracy Theories, and Our Polluted Media Landscape*. The MIT Press.

Rains, S.A., Kenski, K., Coe, K., and Harwood, J. (2017). Incivility and Political Identity on the Internet: Intergroup Factors as Predictors of Incivility in Discussions of News Online, *Journal of Computer-Mediated Communication* 22(4): 163-178. https://doi.org/10.1111/jcc4.12191

Rini, R. (2017). Fake News and Partisan Epistemology. Kennedy Institute of Ethics Journal 27 (S2): 43-64. DOI:10.1353/ken.2017.0025

Rost, K., Stahel, L., & Frey, B. S. (2016). Digital Social Norm Enforcement: Online Firestorms in Social Media. *PloS one* 11(6): e0155923. https://doi.org/10.1371/journal.pone.0155923

Sanfilippo, M., Yang, S. and Fichman, P. (2017), Trolling here, there, and everywhere: Perceptions of trolling behaviors in context. Journal of the Association for Information Science and Technology, 68: 2313-2327. DOI:10.1002/asi.23902

Schwitzgebel, E. (2019). *A theory of jerks and other philosophical misadventures.* New York: MIT Press.

Sellars, W. (1956): "Empiricism and the Philosophy of Mind". In H. Feigl and M. Scriven (eds.), *Minnesota Studies in the Philosophy of Science* (vol. 1) (253-329). Minneapolis: University of Minnesota Press.

Smart, P., Clowes, R., & Heersmink, R. (2017a). Minds online: The interface between Web science, cognitive science, and the philosophy of mind. *Foundations and Trends in Web Science*. 6(1–2), 1–234. DOI: 10.1561/1800000026

Smart, P., Heersmink, R., & Clowes, R. (2017b). The cognitive ecology of the Internet. In S. Cowley & F. Valle ́e-Tourangeau (Eds.), *Cognition beyond the brain: Computation, interactivity and human artifice* (251–282). Dordrecht: Springer.

Skorburg, J. A. (2019). Where are virtues? *Philosophical Studies* 176 (9):2331-2349. DOI: 10.1007/s11098-018-1128-1

Spaulding, S. (2010). Embodied cognition and mindreading. *Mind and Language* 25(1): 119–40.

Sterelny, K. (2012). *The evolved apprentice*. Cambridge: The MIT Press.

Sperber, D. and Wilson, D. (2002). Pragmatics, modularity, and mind-reading. *Mind and Language*, 17(1–2), 3–23. https://doi.org/10.1111/1468-0017.00186

Suler, J. (2004). The Online Disinhibition Effect. *Cyberpsychology & Behavior* 7(3): 321-326. DOI: 10.1089/1094931041291295

Sutton, J. (2013). Skill and Collaboration in the Evolution of Human Cognition. *Biological Theory* 8: 28-36. https://doi.org/10.1007/s13752-013-0097-z

Szabla, M. & Blommaert, J. (2020). Does context really collapse in social media interaction? *Applied Linguistics Review* 11(2): 251-279. https://doi-org.ru.idm.oclc.org /10.1515/applirev-2017-0119

Taddeo, M. (2015) The Struggle Between Liberties and Authorities in the Information Age. Sci Eng Ethics 21, 1125–1138. https://doi.org/10.1007/s11948-014-9586-0

Talisse, R. (2019). *Overdoing democracy: Why we must put politics in its place*. Oxford, United Kingdom: Oxford University Press.

Thomas, H.J., Connor, J.P. & Scott, J.G. (2015) Integrating Traditional Bullying and Cyberbullying: Challenges of Definition and Measurement in Adolescents – a Review. *Educ Psychol Rev* 27: 135–152. https://doi.org/10.1007/ s10648-014-9261-7

Tooby, J. and Cosmides, L. (1995). The language of the eyes as an evolved language of mind. Foreword to: S. Baron-Cohen (ed.), *Mindblindness: an essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Vaccari, C., Valeriani, A., Barberá, P., Jost, J. T., Nagler, J., & Tucker, J. A. (2016). Of Echo Chambers and Contrarian Clubs: Exposure to Political Disagreement Among German and Italian Users of Twitter. *Social Media + Society.* https://doi.org/10.1177/2056305116664221

van Deursen, A. J., & van Dijk, J. A. (2019). The first-level digital divide shifts from inequalities in physical access to inequalities in material access. *New Media & Society*, *21*(2), 354–375. https://doi.org/10.1177/ 1461444818797082

Verbeek, P. (2008) Cyborg intentionality: Rethinking the phenomenology of human–technology relations. Phenom Cogn Sci 7, 387–395. DOI:10.1007/s11097-008-9099-x

Verbeek, Peter-Paul (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. University of Chicago Press.

Vitak, J. (2012) The Impact of Context Collapse and Privacy on Social Network Site Disclosures. *Journal of Broadcasting & Electronic Media* 56(4): 451-470. DOI:10.1080/08838151.2012.732140

Westra, E. (2020) Folk personality psychology: mindreading and mindshaping in trait attribution. *Synthese*. https://doi.org/10.1007/s11229-020-02566-7

Wilkerson, W.S. (2001). Simulation, theory, and the frame problem: The interpretive moment. *Philosophical Psychology* 14(2): 141-153 DOI:10.1080/09515080120051535

Worden, Kirsten J. (2019). Disengagement in the Digital Age: A Virtue Ethical Approach to Epistemic Sorting on Social Media. *Moral Philosophy and Politics* 6(2): 235-259. DOI:10.1515/mopp-2018-0066

Zahavi, D. (2007) Expression and Empathy. In Hutto, D.D. & Ratcliffe, M. (eds.) *Folk Psychology Re-Assessed* (25-47). Springer.

Zahavi, D. (2011) Empathy and Direct Social Perception: A Phenomenological Proposal. *Review Philosophical Psychology* 2: 541-558. https://doi.org/10.1007/s13164-011-0070-3

Zawidzki, T. W. (2008). The function of folk psychology: mind reading or mind shaping? *Philosophical Explorations* 11(3): 193–210. https://doi.org/10.1080/13869790802239235

Zawidzki, T. W. (2013). *Mindshaping: A new framework for understanding human social cognition*. Cambridge, MA: MIT Press.

Zawidzki, T. (2018) Mindshaping. In Newen, A., de Bruin, L., and Gallagher, S. (Eds.) *The Oxford Handbook of 4E Cognition*. Oxford University Press. DOI: 10.1093/oxfordhb/9780198735410.013.39

Antisociality Online: Understanding Social Cognition on Social Networking Sites

Key words: Social Cognition – Social Media – Online Incivility – Active Externalism – Postphenomenology – Online Polarization – Social Norms

1. Description of the Theme and Aim of the Research Project

Social networking sites (henceforth SNSs) have become an integral part of social interaction in contemporary society. Despite the variety of studies on psychology and sociality on SNSs, there is no unitary theoretical framework to account for the psychological and social transformations brought about by SNSs.

Among these transformations, social cognition online is arguably under-researched. Online polarization, aggression, cyber-bullying and trolling are well-known and researched issues involving digital technologies. However, not much research addresses the more general problem of social cognition online - of how users understand the thoughts and behavior of other people in online settings. Nonetheless, understanding toxic social interaction in digital environments requires a theory of how social cognition in general occurs online.

There is a wide philosophical debate regarding the nature of social cognition. In SNSs, social cognition grows in complexity: understanding others changes for a lack of embodied cues and for certain features of SNSs that become integrated in everyday practices without acknowledgement or scrutiny. These features of SNSs shape users' cognitive processes and dictate the rules and means of social interaction.

This research project aims to provide a theory of social cognition in digital environments that explains how SNSs integrate with and selectively employ human cognitive capacities. Understanding how the structure of digital technologies mediates interpersonal relationships and radically integrates cognitive processes constitutes an important research path to understand how social cognition is affected online and how it can turn toxic.

2. Background and Research Questions

I will refer to three major developments in contemporary philosophy in order to provide a perspicuous description of the research question and suggest appropriately precise answers at the end of the project. Each of these research frameworks, despite their different backgrounds, have a bearing for a full understanding of the dynamics of social cognition in digital environments; and while the first two frameworks do not investigate social cognition as such, they do have substantial implications for understanding social cognition online and the overall research project.

2.1 Technological Mediation

The philosophical theory of technological mediation originates in Heidegger's writings on technology and is majorly represented in the works of Don Ihde (1990) and Peter-Paul Verbeek (2011). This approach focuses on the role of technologies in the relation between human beings and their world: technologies are not neutral, but define human beings and constitute their way of life and their relationships with the world.

This theory has been applied to a wide variety of technologies, from FMRI to nanotechnologies to smart cities. Digital environments and SNSs also figure among the technologies under the scope of mediation theory (Nelson 2018). The first applications of mediation theory to SNSs provide some promising foundations to put into clearer focus how these technologies mediate our understanding of other people.

2.2 Active Externalism in Digital Environments

Active externalism in philosophy of mind, while also involving technology use, comes from a different philosophical tradition than mediation theory. This theory argues that extracranial items, under the right circumstances, can co-constitute cognitive processes and transform them over time (Clark & Chalmers 1998; Menary 2010). Clark (2003, 2008) argues that human beings are "natural-born cyborgs": the evolutionary trajectory of our species favored a natural reliance on external structures that ground our cognitive processes. This is consistent with a recent development in evolutionary theory known as niche-construction theory (Laland & O'Brien 2011; Laland, Matthews & Feldman 2016; Sterelny 2010; Sterelny 2013) emphasizing how organisms can manipulate their own environment and influence their evolutionary drive.

Elements that can enable cognitive extension include not only artifacts but also socio-cultural practices and institutions (Gallagher 2013; Slors 2019, 2020). SNSs represent an interesting case for cognitive extension: the dissolution of biological boundaries characterizing extended cognition would take unprecedented form in digital environments (Smart, Clowes & Heersmink 2017, 2018) and the high degree of adaptivity, commodification and personalization of users' online experience can bring about notable epistemic consequences. Considering SNSs in this framework can capture their cognitive effects for users, because, more than mere artifacts, they can be considered *environments*, characterized by social practices and norms users ought to follow for their proper use.

2.3 The Embodiment and Normativity of Social Cognition

The debate regarding how human beings understand each other's minds has a long history in philosophy. This research project will especially focus over theories emphasizing the importance of embodied expressivity and social expectations. Gallagher (2001, 2007) and Zahavi (2007) argue that embodied cues such as facial expressions and gestures are central for social cognition. This theory is particularly interesting when applied to digital environments, not only because there are no proper embodied cues, but because understanding other minds through embodied expressivity is linked to a form of empathy (Zahavi 2007), which might lack in online environments and be an important factor for explaining online aggression, polarization and dehumanization.

Another extremely important philosophical theory of social cognition is known as "mindshaping". This theory argues that understanding other minds involves conformity to and enforcement of social norms (Mameli 2001; Zawidzki 2008, 2013; McGeer 2007, 2015, 2019, 2020; Haslanger 2018). This implies that much of social cognition is dependent on the sociocultural context of the social interaction. In the case of SNSs, this is particularly relevant, not only because different social contexts collapse and dissolve in the platform, but because the material characteristics of these environments afford new social practices and behavioral standards that can be useful for understanding the dynamics of social cognition online.

The combination of these philosophical frameworks provides a number of different dimensions to analyze the dynamics of online social cognition. Technological mediation can help understanding the experience of digital environments; active externalism can help understanding the integration and transformation of cognitive processing online; and embodied and regulative theories of folk psychology can help understanding the transformation of face-to-face social cognition in online settings.

2.4 Research Questions

RQ: How can we provide a perspicuous representation of (hostile) social cognition in social networking sites (SNSs)?

subRQ1: How do theories of technological mediation and active externalism complement each other?

subRQ2: What are the characteristic features of SNSs that technological mediation and active externalism highlight?

subRQ3: What are the design characteristics of SNSs impacting social cognition online?

subRQ4: What are (some) design-level interventions that might improve social cognition and reduce hostility and dehumanization online?

3. The research project

3.1 Year 1: Establishing a Framework: Mediation, Active Externalism and Social Cognition

The first year aims to develop an integrated philosophical framework for the dissertation. The candidate will elaborate and connect two different frameworks, that are the theory of technological mediation and active externalism in philosophy of mind. On the one hand, mediation theory investigates how technology shapes everyday human life actively and not neutrally (Verbeek 2011, 2019, 2020). On the other hand, active externalism argues that environmental elements such as artifacts co-constitute the cognitive processes traditionally understood as the sole result of the brain's activity. Approaches related to EM such as cognitive niche-construction theory (Sterelny 2004, 2010; Clark 2008; Pinker 2010; Bertolotti and Magnani 2017), cognitive integration (Menary 2008, 2010), and, relevantly for the research at hand, theories of cognitive extension through institutions and socially shared norms (Gallagher 2013; Slors 2019, 2020) will be important for establishing the project's framework. Versions of active externalism explicitly applied to the digital world will be taken into account (Clowes 2015, 2019; Smart 2017; Smart et al. 2017a, 2017b; Arfini, Bertolotti & Magnani 2019; Arfini 2020).

Among the various criteria in the literature for cognitive extension and integration (Rowlands 2009; Sterelny 2010; Sutton 2010; Heersmink 2017), there is one of particular importance for the research project, pointed out only by Arfini, Bertolotti & Magnani (2019), that is the immersive character of online platform such as social networking sites. SNSs are complex environments, which are identifiable as virtual online spaces but are simultaneously embedded in and informed by the traces of other people on the network. The complexity of SNSs is determined by their being virtual, privately accessible, and communal spaces, to the point that the user's cognitive processes can be dictated both by the material characteristics of the platform and by socially shared practices that the platform affords.

Active externalism is particularly useful in this context because due to certain characteristics of SNSs such as personalization and immersivity, the configuration of the user's newsfeed can be considered an extremely "comfortable" cognitive niche, being the place of cognitive load from users to the point that they lose track of where their minds and where the niche begins. This is not simply because of the technological characteristics of the platform but also because of the kinds of social practices and standards that the collectivity of users is nudged into by those technological aspects (Selinger and Whyte 2010; Lavi 2018) and of the choices implemented by the platform's designers (Williams 2018).

Combining the two approaches and highlighting their complementary aspects (such as the understanding of users as "cyborgs": Verbeek 2007; Clark 2003) will ground the candidate's research in the following years. A special focus in this phase will be given to how these approaches understand social cognition. Some postphenomenological research focuses on how technology plays an active role in the moral treatment of other people (see e.g. Nelson 2019); on the other hand, approaches to folk psychology related to active externalism underline how understanding other minds relies on shared social norms and context-dependent behavioral expectations (Zawidzki 2013; McGeer 2007, 2015, 2020). The intersection of these different branches of research will provide an adequate background to provide an answer to the research question, regarding how social cognition occurs in social networking sites. Ultimately, a theory of social cognition both based on embodied cues and akin to the phenomenological tradition (Gallagher 2001, 2007; Zahavi 2007) and relying on social regulations and norms will provide adequate *explanans* for the research question.

In this timeframe, two articles will be written for publication. One of these articles would be aimed at describing users in digital environments as cyborgs, building on Verbeek's

(2007, 2013) and Arfini, Bertolotti and Magnani's (2018) divergent understandings of the concept. Particular emphasis will be put on the idea that in digital environments users are phenomenologically indistinguishable from the content they produce and the information they provide, and reflect on how this intrinsic feature of digital platforms might affect the perception of other people online. This article may be submitted to *Minds and Machines*, *Philosophy and Technology* or *Science, Technology and Society.*

The aim of the second will be to delineate the relation between users, digital platforms, and other users met within the platform. I will combine the theory of symbiotic cognition (Slors 2019, 2020) and the notion of digital environments (Smart, Heersmink and Clowes 2017, 2018) and the human-technology relations as *alterity* (Ihde 1990; see also Verbeek 2013) and *cyborg* relations (Verbeek 2007, 2013; Arfini, Bertolotti and Magnani 2017). I will provide an account of human-digital environments relation as one of a *constitutive* kind, whereby users are considered as constitutive elements of digital platforms, while the latter is in itself an environment affording and directing diverse forms of interaction with information and others. This article would be submitted to either *Minds and Machines*, *Philosophy and Technology* or *Techné: Research in Philosophy and Technology*.

3.2 Year 2: Mediation and Extension in Practice: Selection and Analysis of Empirical Studies

The second year aims to gather a number of empirical studies regarding social cognition on social networking sites, from the fields of media and communication sciences and applied psychology. Research in this field is both increasingly vast and developing from a number of different theories, and therefore it is a rather chaotic field. For this reason, it would be appropriate, when possible, to establish some criteria, based upon the framework established the previous year, in order to select what empirical research is relevant for the candidate's research purposes.

The empirical research taken into account will cover three different areas connected to the research framework. Part of the research that will be taken into account will have as an object the characteristics of digital platforms such as SNSs highlighted by mediation theory and active externalism. Taking empirical research into account will be important to prove the usefulness of the philosophical framework previously developed to explain currently diversified research and provide a generalized picture of the dynamics analyzed by the research.

More importantly, the bulk of the research considered in this year will regard social cognition in digital environments and social networking sites. The organization of the empirical research at hand within the theoretical framework will provide a unitary structure to the empirical findings and guide the research project in its last years. One publishable article will be dedicated to the individuation of the most salient predictors of hostility and disinhibition online, such as the lack of embodied cues (Suler 2004; McCall 2013), context collapse (Marwick and boyd 2011; Vitak 2012; Davis & Jurgenson 2014; Costa 2018) and cognitive comfort (the ease of use and consumption of information online: see Ward 2013; Ward et al. 2017). This article may be submitted to *Minds and Machines*, *Techné*, *Technology in Society* or *Science, Technology and Society.*

3.3 Year 3: Social Cognition and Practices on Social Networking Sites

Thanks to the work previously developed, the candidate will elaborate in greater detail how digital environments mediate social cognition. The main principle guiding the link between the framework and the evidence is the consideration of SNSs use as a form of niche construction, where the digital environment is manipulated by users in their personalized account (as a form of cognitive integration); the characteristics of the digital niche affords certain standardized expressive behaviors that can consolidate as social practices guiding the user's activity (a form of cognitive symbiosis); and, ultimately, both the integration between user and platform, the embodied and embedded characteristics of the digital environment as such and the socially shared behaviors that the platform affords partake in social cognition (as a form of complex technological mediation). In sum, one can say that the platform's material characteristics afford new "language games" guiding social interaction, and that social cognition online must be understood in light of the complex interaction between the materiality of the platform and the social practices that the platform co-constitutes.

The ordering of the analyzed empirical studies will support the now fully formulated and detailed thesis regarding how social cognition takes place in social networking sites. The argument will be especially instrumental to understand hostile and violent social cognition on SNSs. This phenomenon will be explained by taking into account the regulative character of social cognition (Zawidzki 2013; McGeer 2015, 2019, 2020) and the features of SNSs investigated in year 2 through the philosophical framework of year 1.

One publication, built upon the research of the previous years, will be produced. Its object will be the comparison of how different online platforms enable different modalities of social interaction and shared practices. Such a comparison will support a closer understanding of how specific design features of online platforms influence interaction, and implications for possible design changes for reducing toxicity in interaction will be drawn.

3.4 Year 4: Summary and thesis

The aim of this year will be to summarize the research of the previous years, draw up the conclusions and suggest some future research directions. Importantly, after carefully considering the various technological and social factors contributing to social cognition on social networking sites, some ideas for implementational changes in the platform's design and policy will be provided at the conclusion of the research. These changes will keep as a general aim to make engagement on SNSs more civil and virtuous.

4. Philosophical, scientific and social impact of the project

- Connecting approaches from different traditions, namely active externalism and technological mediation, with the aim of mutually enriching them;
- Understanding some of the epistemically vicious consequences of the integration digital technologies in cognitive and social practices;
- Contributing to the dialogue between philosophical and empirical research in digital media studies;
- Proposing design changes and implementations for SNSs in order to reduce toxicity and hostility in online social interaction.

5. Summary for non-specialists

Online incivility and aggression have become a generalized concern in contemporary society. It seems that social media has made people increasingly polarized and divided, and that interaction online has become increasingly uncivil and aggressive. Not only online aggression in everyday interaction is a potential harm to people's well-being. Toxic interaction may have the implication of isolating social media users as citizens, and to hinder the healthy functioning of a democratic society, where there must be a common ground for debate and deliberation can be possible.

While such concerns are justified, there is a tendency, in both academia and especially at a broader societal level, to either assume or underexplain the negative impact of social media on social interaction. Understanding the exact nature of this relation is fundamental in order to provide effective solutions to socially and epistemically toxic practices such as cyber-aggression, polarization and the spread of mis- and malinformation.

This research project sets out to explore the connection between the social and cognitive impact of social media use and the apparent rise in online violence, aggression and incivility. My aim is to bypass the moral panics surrounding social media use and set out to explore how social media, understood both as a technology and as a locus of sociality, impacts interaction in toxic ways. Examining the way social media accommodates and transforms information processing and drives interaction in value-laden ways is critical in order to make online interaction and information consumption healthier, and to reduce violence online.

On the one hand, I will make use of a variety of approaches in philosophy of technology in order to have a rich framework for a perspicuous understanding of how social media transforms the way we approach information and others. I will consider approaches that treat technologies as (potential) means of extension and transformation of our cognitive capacities (a branch of theories known as *active externalism* in philosophy of mind); as well as theories that examine how certain technologies embody some social and moral values and change our perception of the world and ourselves (an approach commonly known as *postphenomenology*). On the other hand, I will compare different online platforms and social networking sites in order to grasp how the different design choices of the different platforms influence social interaction and enable (or impede) different kinds of practices and relations between their users.

A close analysis of the social and cognitive impact of digital technology through the lenses of active externalism and postphenomenology can not only account for a deeper understanding of online toxicity. It can also illustrate which features of these digital platforms affect cognition and interaction in negative ways. In this way, the research project can lay a path for possible design implementation that can enhance everyday interaction online.

Preliminary bibliography

Adams, F., and Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology,* 14: 43–64. https://doi.org/10.1080/09515080120033571

Andrews, K. (2009). Understanding Norms Without a Theory of Mind. *Inquiry* 52, 433-448. https://doi.org/10.1080/00201740903302584

Andrews, K. (2012). *Do apes read minds? Toward a new folk psychology*. Cambridge, MA: MIT Press.

Antoci A, Delfino A, Paglieri F, Panebianco F, Sabatini F (2016). Civility vs. Incivility in Online Social Interactions: An Evolutionary Approach. *PLOS ONE* 11(11): e0164286. https://doi.org/10.1371/journal.pone.0164286

Arfini, S., Bertolotti, T., Magnani, L. (2018). Cyber-Bullies as Cyborg-Bullies. In R. Luppicini (Ed.) *The Changing Scope of Technoethics in Contemporary Society* (60-73), IGI Global. DOI: 10.4018/978-1-5225-5094-5.ch004

Arfini, S., Bertolotti, T., Magnani, L. (2019) Online communities as virtual cognitive niches. *Synthese* 196 (1): 377-397. https://doi.org/10.1007/s11229-017-1482-0

Aro, J. (2016). The cyberspace war: propaganda and trolling as warfare tools. *EUROPEAN VIEW* 15: 121-132. DOI:10.1007/s12290-016-0395-5

Aydin, C., González Woge, M. & Verbeek, P. (2019) Technological Environmentality: Conceptualizing Technology as a Mediating Milieu. *Philosophy & Technology,* 32: 321–338. https://doi.org/10.1007/s13347-018-0309-3

Barbu, C.M., and Schneeberger, T. (2014) How can recommender systems help people learn while choosing? *Proceedings of the 1st International Workshop on Decision Making and Recommender Systems Volume: CEUR Workshop Proceedings* 1278: 49-51. Retrieved from: http://ceur-ws.org/Vol-1278/paper10.pdf

Bail, C.A., Guay, B, Maloney, E., Combs, A., Hillygus, S.S., Merhout, F., Freelon, D. and Volfosky, A. (2020). Assessing the Russian Internet Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. *PNAS* 117(1): 243-350. DOI:10.1073/pnas.1906420116

Beam, M. A., Child, J. T., Hutchens, M. J., & Hmielowski, J. D. (2018). Context collapse and privacy management: Diversity in Facebook friends increases online news reading and sharing. *New Media & Society* 20(7): 2296–2314. https://doi.org/10.1177/1461444817714790

Bergen, J.P., Verbeek, P. (2020) To-Do Is to Be: Foucault, Levinas, and Technologically Mediated Subjectivation. *Philos. Technol*. https://doi.org/10.1007/s13347-019-00390-7

Bergram, K., Gjerlufsen, T., Maingot, P., Bezençon, V., Holzer, A. (2020) DIGITAL NUDGES FOR PRIVACY AWARENESS: FROM CONSENT TO INFORMED CONSENT? *Twenty-Eigth European Conference on Information Systems (ECIS2020), Marrakesh, Morocco. Retrieved from: https://www.researchgate.net/publication/346915885_DIGITAL_NUDGES_FOR_PRIVACY_AWARENESS_FROM_CONSENT_TO_INFORMED_CONSENT*

Bermúdez, J. (2003). The Domain of Folk Psychology. *Royal Institute of Philosophy Supplement* 53: 25-48. doi:10.1017/S1358246100008250

Bertolotti, T. (2011). Facebook Has It: The Irresistible Violence of Social Cognition in the Age of Social Networking. *International Journal of Technoethics (IJT), 2*(4), 71-83. doi:10.4018/jte.2011100105

Bertolotti T. (2015) Niche Construction Through Gossip and Mobbing: The Mediation of Violence in Technocognitive Niches. In: *Patterns of Rationality. Studies in Applied Philosophy, Epistemology and Rational Ethics*, (145-169) Springer, Cham. DOI: 10.1007/978-3-319-17786-1_8

Bertolotti, T., Magnani, L. (2013) Terminator Niches. *Proceedings of the Virtual Reality International Conference: Lava Virtual*: 1-10. https://doi.org/10.1145/2466816.2466850

Bertolotti, T., Magnani, L. (2017) Theoretical considerations on cognitive niche construction. *Synthese* 194: 4757–4779. https://doi.org/10.1007/s11229-016-1165-2

Bertolotti, T.; Arfini, S.; Magnani, L. (2017) Of Cyborgs and Brutes: Technology-Inherited Violence and Ignorance. *Philosophies* 2(1): 1-14. https://doi.org/10.3390/philosophies2010001

Brown, J.J. Jr., and Hennis, G. (2019). Hateware and the Outsourcing of Responsibility. In Reyman, J., & Sparby, E. (Eds.) *Digital Ethics: Rhetoric and Responsibility in Online Aggression* (1st ed.) (18-31). Routledge.

Bucher, T. (2018) *If...Then. Algorithmic Power and Politics*. New York: Oxford University Press.

Cho, D., and Kwon, K.H. (2015). The impacts of identity verification and disclosure of social cues on flaming in online user comments. *Computers in Human Behavior* 52(A): 363-372. https://doi.org/10.1016/j.chb.2015.04.046.

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis* 58: 7–19. https://doi.org/10.1111/ 1467-8284.00096

Clark, A. (2003). *Natural-born cyborgs: minds, technologies, and the future of human intelligence*. Oxford; New York: Oxford University Press. ISBN 0-19-514866-5

Clark, A. (2005). Word, niche and super-niche: How language makes minds matter more. *Theoria* 54: 255–268. https://doi.org/10.1387/theoria.561.

Clark, A. (2008). *Supersizing the mind: embodiment, action, and cognitive extension*. Oxford; New York: Oxford University Press. ISBN 978-0-19-533321-3

Clowes, R. (2015) Thinking in the Cloud: The Cognitive Incorporation of Cloud-Based Technology. *Philos. Technol.* 28: 261–296. https://doi.org/10.1007/s13347-014-0153-z

Clowes, R.W. (2019) Immaterial engagement: human agency and the cognitive ecology of the internet. *Phenom Cogn Sci* 18: 259–279. https://doi.org/10.1007/s11097-018-9560-4

Costa, E. (2018). Affordances-in-practice: An ethnographic critique of social media logic and context collapse. *New Media & Society* 20(10): 3641–3656. https://doi.org/10.1177/1461444818756290

D'Errico, F., & Paciello, M. (2018). Online moral disengagement and hostile emotions in discussions on hosting immigrants. *Internet Research* 28: 1313–1335. https://doi.org/10.1108/IntR-03-2017-0119.

Davis, J. L., & Jurgenson, N. (2014) Context collapse: theorizing context collusions and collisions, *Information, Communication & Society* 17(4): 476-485. https://doi.org/10.1080/1369118X.2014.888458

Dubois, E., Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society* 21(5): 729-745. https://doi.org/10.1080/1369118X.2018.1428656

Fenici, M., and Zawidzki, T. (2020). The origins of mindreading: how interpretive socio-cognitive practices get off the ground. *Synthese*: 1-23. DOI: 10.1007/s11229-020-02577-4

Fernandez Castro, V. (2020). Regulation, Normativity and Folk Psychology. *Topoi* 39: 57–67. https://doi.org/10.1007/ s11245-017-9511-7

Flaxman, S., Goel, S., and Rao J.M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quart.* 80(S1): 298–320. https://doi.org/10.1093/poq/nfw006

Fletcher-Watson, S., De Jaegher, H., van Dijk, J., Frauenberg, C., Magnée, M. Ye, J. (2018) Diversity Computing. *Interactions* 25(5): 28-33. DOI:https://doi.org/10.1145/3243461

Gallagher, S. (2001). The practice of mind: Theory, simulation or primary interaction? *Journal of Consciousness Studies* 8 (5-7): 83-108. DOI: 10.1.1.710.5008

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition* 17(2): 535-543. https://doi.org/10.1016/j.concog.2008.03.003

Gallagher, S. (2013) The socially extended mind. *Cognitive Systems Research* (25–26): 4-12. http://dx.doi.org/10.1016/j.cogsys.2013.03.008

Gallagher, S. (2018) The Extended Mind: State of the Question. *The Southern Journal of Philosophy* 56(4): 421-447. https://doi.org/10.1111/sjp.12308

Gallagher, S., and Hutto, D. D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), *Converging evidence in language and communication research (CELCR): Vol. 12. The shared mind: Perspectives on intersubjectivity* (17–38). John Benjamins Publishing Company.

Gallagher, S., Varga, S. (2014). Social Constraints on the Direct Perception of Emotions and Intentions. *Topoi* 33: 185–199. https://doi.org/10.1007/s11245-013-9203-x

Halpin, H., and Monnin, A. (Eds.) (2013) *Philosophical Engineering. Toward a Philosophy of the Web*. Wiley-Blackwell. ISBN: 978-1-118-70018-1

Harel, T. O., Jameson, J. K., & Maoz, I. (2020). The Normalization of Hatred: Identity, Affective Polarization, and Dehumanization on Facebook in the Context of Intractable Political Conflict. *Social Media + Society*. https://doi.org/10.1177/2056305120913983

Haslanger, S. (2019) Cognition as a Social Skill. *Australasian Philosophical Review*, 3(1): 5-25. https://doi.org/10.1080/24740500.2019.1705229

Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. *Phenomenology and the Cognitive Sciences* 14(3): 577–598. https://doi.org/10.1007/s11097-014-9355-1

Heersmink, R. (2016) The Internet, Cognitive Enhancement, and the Values of Cognition. *Minds & Machines* 26: 389–407. https://doi.org/10.1007/s11023-016-9404-3

Heersmink, R., Sutton, J. (2020) Cognition and the Web: Extended, Transactive, or Scaffolded?. *Erkenn* 85: 139–164. https://doi.org/10.1007/s10670-018-0022-8

Highfield, T. (2018) *Social Media and Everyday Politics*. Cambridge; Malden: Polity Press. ISBN: 978-0-745-69134-3

R. Hoffmann, C., & Bublitz, W. (Eds.) (2017). *Pragmatics of Social Media*. Berlin, Boston: De Gruyter Mouton

Hutchins, E. (2010a). Cognitive ecology. *Topics in Cognitive Science,* 2: 705–715. https://doi.org/10.1111/j.1756-8765.2010.01089.x

Hutchins, E. (2014). The cultural ecosystem of human cognition. *Philosophical Psychology* 27(1): 34-49. https://doi.org/10.1080/09515089.2013.830548

Hutto, D. D. (2008). *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons*. Bradford.

Ihde, D. (1990). *Technology and the Lifeworld: From Garden to Eart*h. Bloomington; Indianapolis: Indiana University Press. ISBN:0-253-32900-0

Ihde, D., Malafouris, L. *Homo faber* Revisited: Postphenomenology and Material Engagement Theory. *Philos. Technol.* 32, 195–214 (2019). https://doi.org/10.1007/s13347-018-0321-7

Ivcevic, Z., Ambady, N. (2012) Face to (Face)Book: The Two Faces of Social Behavior? J*ournal of Personality* 81(3): 290-301. DOI: 10.1111/j.1467-6494.2012.00804.x

Kiri Gunn, H. (2020). How Should We Build Epistemic Community? *Journal of Speculative Philosophy* 34 (4): 561-581. https://doi.org/10.5325/jspecphil.34.4.0561

Kirkpatrick D. (2010). *The Facebook Effect: The Inside Story of the Company that is Connecting the World*. New York: Simon and Schuster.

Kowalski, R. M., Giumetti, G., Schroeder, A., & Lattanner, M. (2014). Bullying in the digital age: a critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin* 140: 1073- 1137. DOI: 10.1037/a0035618

Kramer, A.D.I., Guillory, J.E., and Hancock, J.T (2014) Experimental evidence of massive-scale through social networks. *PNAS* 111 (24): 8788-8790. https://doi.org/10.1073/pnas.1320040111

Laland, K., O'Brien, M.J. (2011). Cultural Niche Construction: An Introduction. *Biological Theory 6*:191-202. https://doi.org/10.1007/s13752-012-0026-6

Laland, K., Matthews, B., Feldman, M. W. (2016). An introduction to niche construction theory. *Evolutionary Ecology 30*:191–202. https://doi.org/10.1007/s10682-016-9821-z

Langos C. (2012). Cyberbullying: the challenge to define. *Cyberpsychology, behavior and social networking* 15(6): 285–289. https://doi.org/10.1089/cyber.2011.0588

Lavi, M. (2018) Evil nudges. *Vanderbilt Journal of Entertainment and Technology Law* 21(1): 1+. Retrieved from: https://www.researchgate.net/publication/338223653_Evil_Nudges

Leonardi, P.M. (2018) Social Media and the Development of Shared Cognition: The Roles of Network Expansion, Content Integration, and Triggered Recalling. *Organization Science* 29(4):547-568. https://doi.org/10.1287/orsc.2017.1200

Lo Presti, P. (2016). An ecological approach to normativity. *Adaptive Behavior 24(1)*: 3–17. https://doi.org/10.1177/1059712315622976

Lapidot-Lefler, N., and Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior* 28(2): 434–443. https://doi.org/10.1016/j.chb.2011.10.014

Lynch, M. P. (2014) Neuromedia, Extended Knowledge and Understanding. *Philosophical Issues, 24*: 299-313 https://doi.org/10.1111/phis.12035

Lynch, M. P. (2016) *The Internet of Us. Knowing More and Understanding Less in the Age of Big Data*. New York: Liveright. ISBN: 978-1-63149-277-8

Maher, P., Igou, E., and van Tilburg, W. (2018). Brexit, Trump, and the polarizing effect of disillusionment. *Social Psychological & Personality Science* 9(2): 205-213. https://doi.org/10.1177/1948550617750737

Marwick, A. E., & boyd, danah. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1): 114–133. https://doi.org/10.1177/1461444810365313

McCall, C. A. (2013). Social Cognition in the Cyborg Age: Embodiment and the Internet. *Psychological Inquiry* 24(4): 314-320. https://doi.org/10.1080/1047840X.2013.839921

McGeer, V. (2015). Mind-making practices: the social infrastructure of self-knowing agency and responsibility. *Philosophical Explorations,* 18(2): 259-281. https://doi.org/10.1080/13869795.2015.1032331

McGeer, V. (2019) Scaffolding agency: A proleptic account of the reactive attitudes. *Eur J Philos*. 27: 301– 323. https://doi.org/10.1111/ejop.12408

McGeer, V. (2020) Enculturating Folk Psychologists. *Synthese*. DOI: 10.1007/s11229-020-02760-7

Menary, R. (2007) *Cognitive Integration. Mind and Cognition Unbounded.* Basingstoke, Hampshire: Palgrave Macmillan Houndmills. ISBN 978–0–230–54935–7

Menary, R. (2010). *The Extended Mind*, Cambridge, MA: MIT Press. ISBN 978-0-262-01403-8

Menary, R. (2013). Cognitive integration, Enculturated Cognition and the Socially Extended Mind. *Cognitive Systems Research 25–26*, 26–34. https://doi.org/10.1016/j.cogsys.2013.05.002

Michelfelder, D.P. (2010) Philosophy, privacy, and pervasive computing. *AI & Soc* 25: 61–70 https://doi.org/10.1007/s00146-009-0233-2

Miller, B. (2021). Is Technology Value-Neutral? *Science, Technology, and Human Values* 46 (1): 53-80. https://doi.org/10.1177/0162243919900965

Miller, B., and Record, I. (unpublished). People, Posts, and Platforms. Reducing the spread of online toxicity by contextualizing content and setting norms.

Miller, R. L., Brickman, P., & Bolen, D. (1975). Attribution versus persuasion as a means for modifying behavior. *Journal of Personality and Social Psychology* 31(3): 430–441. DOI: 10.1037/h0076539

Moor, P.J., Heuvelman, A., and Verleur, R. (2010). Flaming on YouTube. *Comput. Hum. Behav.* 26(6): 1536–1546. https://doi.org/10.1016/j.chb.2010.05.023

Nelson, L.S. (2018) *Social Media and Morality. Losing Our Self Control*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316691359

Nichols, S., and Stich, S. P. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford University Press.

Nisbet, E. C., Cooper, K. E., and Garrett, R. K. (2015). The partisan brain: How dissonant science messages lead conservatives and liberals to (dis)trust science. *The Annals of the American Academy of Political and Social Science,* 658(1): 36-66. https://doi.org/10.1177/0002716214555474

Notten, N., Peter, J., Kraaykamp, G., and Valkenburg, P.M. (2009). Research Note: Digital Divide Across Borders—A Cross-National Study of Adolescents' Use of Digital Technologies. *European Sociological Review* 25(5): 551–560. https://doi-org.ru.idm.oclc.org/10.1093/esr/jcn071

Pacilli, M. G., Roccato, M., Pagliaro, S., & Russo, S. (2016). From political opponents to enemies? The role of perceived moral distance in the animalistic dehumanization of the political outgroup. *Group Processes & Intergroup Relations* 19(3), 360–373. https://doi.org/10.1177/1368430215590490

Peters, U. (2019). The complementarity of mindshaping and mindreading. *Phenomenology and the Cognitive Sciences* 18(3): 533–549. https://doi.org/10.1007/s11097-018-9584-9

Peters, U. (2021). How (Many) Descriptive Claims About Political Polarization Exacerbate Polarization. *Journal of Social and Political Psychology* 9(1): 24-36. https://doi.org/10.5964/jspp.5543

Phillips, W. (2015) *This is Why We Can't' Have Nice Things. Mapping the relationship between online trolling and mainstream culture*. Cambridge MA: The MIT Press. ISBN: 9780262529877

Phillips, W., and Milner, R. (2017). *The Ambivalent Internet. Mischief, Oddity, and Antagonism Online*. Polity Press

Phillips, W., and Milner, R. (2021). *You Are Here. A Field Guide for Navigating Polarized Speech, Conspiracy Theories, and Our Polluted Media Landscape*. The MIT Press.

Pinker, S. (2010) The cognitive niche: Coevolution of intelligence, sociality, and language. PNAS 107(2): 8993-8999. https://doi.org/10.1073/pnas.0914630107

Rains, S.A., Kenski, K., Coe, K., and Harwood, J. (2017). Incivility and Political Identity on the Internet: Intergroup Factors as Predictors of Incivility in Discussions of News Online, *Journal of Computer-Mediated Communication* 22(4): 163-178. https://doi.org/10.1111/jcc4.12191

Record, I., & Miller, B. (2018) Taking Iphone Seriously: Epistemic Technologies and the Extended Mind. In, Carter, J. Adam, Clark, Andy, Kallestrup, Jesper, Palermos, S. Orestis and Pritchard, Duncan (eds.) *Extended Epistemology.* Oxford. Oxford University Press. DOI: 10.1093/oso/9780198769811.001.0001

Rini, R. (2017). Fake News and Partisan Epistemology. Kennedy Institute of Ethics Journal 27 (S2): 43-64. DOI:10.1353/ken.2017.0025

Rost, K., Stahel, L., & Frey, B. S. (2016). Digital Social Norm Enforcement: Online Firestorms in Social Media. *PloS one* 11(6): e0155923. https://doi.org/10.1371/journal.pone.0155923

Rowlands, M. (2009), THE EXTENDED MIND. *Zygon®* 44: 628-641. https://doi.org/10.1111/j.1467-9744.2009.01021.x

Rupert, R. (2004). Challenges to the Hypothesis of Extended Cognition. *The Journal of Philosophy* 101(8): 389-428. http://www.jstor.org/stable/3655517

Sanfilippo, M., Yang, S. and Fichman, P. (2017), Trolling here, there, and everywhere: Perceptions of trolling behaviors in context. Journal of the Association for Information Science and Technology, 68: 2313-2327. DOI:10.1002/asi.23902

Selinger, E., and Whyte, K.P. (2010) Competence and Trust in Choice Architecture. *Know Techn Pol* 23: 461–482. https://doi.org/10.1007/s12130-010-9127-3

Smart, P (2017). Extended Cognition and the Internet. *Philos. Technol.* 30: 357–390. https://doi.org/10.1007/s13347-016-0250-2

Smart, P. (2018) Emerging digital technologies: Implications for extended conceptions of cognition and knowledge. In, Carter, J. Adam, Clark, Andy, Kallestrup, Jesper, Palermos, S. Orestis and Pritchard, Duncan (eds.) *Extended Epistemology.* Oxford. Oxford University Press, pp. 266-304. DOI: 10.1093/oso/9780198769811.001.0001

Smart, P., Engelbrecht, P., Braines, D., Hendler, J., Shadbolt, N. (2008) The Extended Mind and Network-Enabled Cognition. Retrieved from: https://www.researchgate.net/publication/39997176_The_Extended_Mind_and_ Network-Enabled_Cognition/citations

Smart, P., Clowes, R. W. & Heersmink, R. (2017a). Minds Online: The Interface between Web Science, Cognitive Science, and the Philosophy of Mind. *Foundations and Trends in Web Science* 6 (1-2):1-234. http://dx.doi.org/10.1561/1800000026

Smart P., Heersmink R., Clowes R.W. (2017b) The Cognitive Ecology of the Internet. In: Cowley S., Vallée-Tourangeau F. (eds) *Cognition Beyond the Brain*. Springer, Cham. 251-282. https://doi.org/10.1007/978-3-319-49115-8

Skorburg, J. A. (2019). Where are virtues? *Philosophical Studies* 176 (9):2331-2349. DOI: 10.1007/s11098-018-1128-1

Slors, M.V.P. (2019). Symbiotic cognition as an alternative for socially extended cognition. *Philosophical Psychology* 32 (8). https://doi.org/10.1080/09515089.2019.1679591

Slors, M.V.P. (2019). A Cognitive Explanation for the Perceived Normativity of Cultural Conventions. *Mind and Language* 35 (1): .https://doi.org/10.1111/mila.12265

Slors, M.V.P. (2020) From Notebooks to Institutions: The Case for Symbiotic Cognition. *Front. Psychol.* 11: 674. ***https://doi.org/10.3389/fpsyg.2020.00674***

Sperber, D. and Wilson, D. (2002). Pragmatics, modularity, and mind-reading. *Mind and Language*, 17(1–2), 3–23. https://doi.org/10.1111/1468-0017.00186

Sterelny, K (2004), 'Externalism, Epistemic Artefacts and The Extended Mind', in R.Schantz (ed.), *The Externalist Challenge* (239-254). Walter de Gruyter, Berlin. 239-254. DOI:10.1515/9783110915273

Sterelny, K. (2010). Minds: Extended or Scaffolded? *Phenomenology and the Cognitive Sciences,* 9:465–481. https://doi.org/10.1007/s11097-010-9174-y

Sterelny, K (2014) Constructing the Cooperative Niche. In C. T. Wolfe et al. (eds.), *Entangled Life*. New York: Springer, pp. 261-279.

Sterelny, K 2017, 'Artifacts, Symbols, Thoughts', Biological Theory 12: 236–247. https://doi.org/10.1007/s13752-017-0277-3

Suler, J. (2004) The Online Disinhibition Effect. *Cyberpsychology & Behavior*, 7(3): 321-326. https://doi.org/10.1089/1094931041291295

Sunstein, C.R. (2019) *Conformity. The Power of Social Influence*. New York: New York University Press. ISBN:9781479867837

Sutton, J. (2013). Skill and Collaboration in the Evolution of Human Cognition. *Biological Theory* 8: 28-36. https://doi.org/10.1007/s13752-013-0097-z

Szabla, M. & Blommaert, J. (2020). Does context really collapse in social media interaction? *Applied Linguistics Review* 11(2): 251-279. https://doi-org.ru.idm.oclc.org /10.1515/applirev-2017-0119

Taddeo, M. (2017) Trusting Digital Technologies Correctly. *Minds & Machines* 27: 565–568. https://doi.org/10.1007/s11023-017-9450-5

Taddeo, M. (2018) Deterrence and Norms to Foster Stability in Cyberspace. *Philos. Technol.* 31: 323–329. https://doi.org/10.1007/s13347-018-0328-0

Taddeo, M. (2019) The Civic Role of Online Service Providers. *Minds & Machines* 29: 1–7. https://doi.org/10.1007/s11023-019-09495-6

Talisse, R. (2019). *Overdoing democracy: Why we must put politics in its place*. Oxford, United Kingdom: Oxford University Press.

Thomas, H.J., Connor, J.P. & Scott, J.G. (2015) Integrating Traditional Bullying and Cyberbullying: Challenges of Definition and Measurement in Adolescents – a Review. *Educ Psychol Rev* 27: 135–152. https://doi.org/10.1007/ s10648-014-9261-7

Tooby, J. and Cosmides, L. (1995). The language of the eyes as an evolved language of mind. Foreword to: S. Baron-Cohen (ed.), *Mindblindness: an essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Vaccari, C., Valeriani, A., Barberá, P., Jost, J. T., Nagler, J., & Tucker, J. A. (2016). Of Echo Chambers and Contrarian Clubs: Exposure to Political Disagreement Among German and Italian Users of Twitter. *Social Media + Society.* https://doi.org/10.1177/2056305116664221

van Deursen, A. J., & van Dijk, J. A. (2019). The first-level digital divide shifts from inequalities in physical access to inequalities in material access. *New Media & Society*, *21*(2), 354–375. https://doi.org/10.1177/ 1461444818797082

Verbeek, P.P. (2006) Materializing morality. Design ethics and technological mediation. *Science, Technology and Human Values* 31(3): 361-380. https://doi.org/10.1177/0162243905285847

Verbeek, P. (2008) Cyborg intentionality: Rethinking the phenomenology of human–technology relations. *Phenom Cogn Sci* 7, 387–395. https://doi.org/10.1007/s11097-008-9099-x

Verbeek, P.P. (2005), *What Things Do – Philosophical Reflections on Technology, Agency, and Desig*n. The Pennsylvania State University Press, University Park, Pennsylvania, 2005 ISBN-13:978-0271025407

Verbeek, P.P. (2011), *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago and London: University of Chicago Press. isbn-10: 0-226-85293-8 (paper)

Verbeek, P. P. C. C. (2015). Cover story: Beyond Interaction: a short introduction to mediation theory. *Interactions (ACM)*, *22*(3), 26-31. https://doi.org/10.1145/2751314

Vitak, J. (2012) The Impact of Context Collapse and Privacy on Social Network Site Disclosures, *Journal of Broadcasting & Electronic Media*, 56(4): 451-470. https://doi.org/10.1080/08838151.2012.732140

Ward, A.F.. (2013) *One with the Cloud: Why People Mistake the Internet's Knowledge for Their Own*. Doctoral dissertation, Harvard University. http://nrs.harvard.edu/urn-3:HUL.InstRepos:11004901

Ward, A.F, Duke, K., Gneezy, A., and Bos, M.W. (2017) Brain drain: The Mere Presence of One's Own Smartphone Reduces Available Cognitive Capacity. *Journal of the Association for Consumer Research* 2(2): 140-154. https://doi.org/10.1086/691462

Westra, E. (2020) Folk personality psychology: mindreading and mindshaping in trait attribution. *Synthese*. https://doi.org/10.1007/s11229-020-02566-7

Wilkerson, W.S. (2001). Simulation, theory, and the frame problem: The interpretive moment. *Philosophical Psychology* 14(2): 141-153 DOI:10.1080/09515080120051535

Williams, J. (2018) *Stand Out of Our Light. Freedom and Resistance in the Attention Economy*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108453004

Worden, Kirsten J. (2019). Disengagement in the Digital Age: A Virtue Ethical Approach to Epistemic Sorting on Social Media. *Moral Philosophy and Politics* 6(2): 235-259. DOI:10.1515/mopp-2018-0066

Zahavi, D. (2007) Expression and Empathy. In Hutto, D.D. & Ratcliffe, M. (eds.) *Folk Psychology Re-Assessed* (25-47). Springer.

Zahavi, D. (2011) Empathy and Direct Social Perception: A Phenomenological Proposal. *Review Philosophical Psychology* 2: 541-558. https://doi.org/10.1007/s13164-011-0070-3

Zawidzki, T. (2008). The function of folk psychology: Mind reading or mind shaping? *Philosophical Explorations* 11(3): 193–210. https://doi.org/10.1080/13869790802239235

Zawidzki, T. (2013). *Mindshaping. A New Framework for Understanding Human Social Cognition*. Cambridge, MA: The MIT Press. ISBN 978-0-262-01901-9

Zawidzki, T. (2018) Mindshaping. In Newen, A., de Bruin, L., and Gallagher, S. (Eds.) *The Oxford Handbook of 4E Cognition*. Oxford University Press. DOI: 10.1093/oxfordhb/9780198735410.013.39

Curriculum Vitae

Giacomo Figà Talamanca

Contact

Radboud Universiteit Nijmegen,
Erasmusplein 1
6526 HT Nijmegen
Phone: (+39) 3408917546
Email: giacomo.f.talamanca@gmail.com

Personal Information

Place and Date of Birth: Roma, Italy, September 4, 1995
Nationality: Italian.
Languages: Italian (mother tongue); English (C1 – IELTS Academic 2018: 7.5)

Current Position

Jan 2019 – Research Master's in Philosophy of Mind | Radboud Universiteit
Department of Philosophy: currently researching the application of theories of active externalism and theories of social cognition to Digital Environments. Supervisor: Marc V.P. Slors.

Areas of Interest

- Extended Cognition and the Internet
- Niche-Construction Theory
- Societal Impact of Digital Technologies
- Vice Epistemology
- Psychology and Epistemology of Conspiracy Theories

Education

Nov 2020 – May 2021 **Visiting Student under Honors Academy Project**

University of Pavia, Pavia, Italy. Supervisors: Marc V. P. Slors and Selene Arfini. Title of the project: “Rethinking Filter Bubbles”.

Jan – July 2017 **Erasmus+ Exchange Period**
Radboud University, Nijmegen, The Netherlands. Coordinator: Prof. Emidio Spinelli, Sapienza Università di Roma.

Sept 2014 – March 2018 **Bachelor’s Degree in Philosophy**
Sapienza Università di Roma, Roma, Italy. Supervisor: Tito Magri. Thesis Title: “Self-Conscious Animals. The Role of the First-Person Perspective in the Identity of Human Beings.” – *Full Marks and Honor.*

Publications

Work in progress “Rethinking filter bubbles.” (first author with Arfini, S.).

Work in progress “Mindshaping Jerks: the Folk Psychology of Online Hostility.”

Forthcoming “Cognitive integration, epistemic parasitism. Social networking sites and the exploitation of ignorance.” (first author with Hunting, E.M.) In Arfini, S. and Magnani, L. (Eds.) *Embodied, Extended, Ignorant Minds. New Studies on the Nature of Not-Knowing,* Synthese Library, Cham: Springer.

Under Review “Wittgenstein’s Lebensform as Niche Construction: an evolutionistic reading.” Submitted to the *European Journal of Philosophy.*

2021 “Joint Action Without Mutual Beliefs: Cooperative Hunting in Chimpanzees as an Evolutionary Precursor to Common Ground.” *Kriterion Journal of Philosophy.* DOI: https://doi.org/10.1515/krt-2021-0004

Awards

Jul. 2020 Scholarship granted for Honors Academy Project: “Rethinking Filter Bubbles”. Supervisors: Prof. Marc V.P Slors (Radboud Universiteit, Nijmegen), Dr. Selene Arfini (University of Pavia)

Jan. 2017 Scholarship granted for Erasmus+ exchange semester from Sapienza Università di Roma to Radboud University, Nijmegen

Teaching Activities

Apr 2020 **Teaching Assistant**

– Jul 2021 Radboud University, Nijmegen, The Netherlands. Bachelor Psychology Course. Supervisor: Marc V. P. Slors

Mar 2021 **Seminar "Rethinking Filter Bubbles"**
University of Pavia, Pavia, Italy. Course of Cognitive Philosophy.
Supervisor: Selene Arfini

Presentations and Summer Schools

18-19 June 2021 **Ethics of Conversation and Disagreement.** Online workshop, Tilburg University, Turkish-German University Istanbul. (Schedule forthcoming. Link: https://dgphil.de/veranstaltungen/cfp-tagungen/lesen/ ?tx_ttnews%5Byear%5D=2021&tx_ttnews%5Bmonth%5D=04&tx_ttnews%5Bday%5D=22&tx_ttnews%5Btt_news%5D=5315&cHash=2054a9ef2941e40bfa9bd6ed9658c00e)

22 Apr. 2021 **Graduate Conference in Theoretical Philosophy 2021**. Delft University of Technology, Eindhoven University of Technology (online event). Link: https://www.ozsw.nl/activity/graduate-conference-in-theoretical-philosophy-2021/

04-07 Nov. 2020 **Philosophy of Human-Technology Relations 2020 conference**. University of Twente, Enschede (online event). Link: https://www.utwente.nl/en/phtr/. Presentation n. 67c.

05-09 Aug. 2019 **Summer School "Philosophical Lessons for and from the Post-Truth Era".** Radboud University, Nijmegen, The Netherlands. Supervisor: Simon Truwant (KU Leuven).

Conferences and Extracurricular Meetings

21 – 22 Apr. 2021 **"Beyond Fake News: Mitigating the Spread of Epistemically Toxic Content"**. Online Workshop, Zefat Academic College and University of
Haifa.

04 Feb. – 07 Apr. 2021 **Manipulation Online: Philosophical Perspectives on Human-Machine Interaction.** Online Workshop Series.

12 Oct. 2020 **Digital Identities, Digital Ways of Living: Philosophical Analyses.** San Raffaele School of Philosophy, Milan (online event).

13-14 Dec. 2019 **2nd Annual Political Epistemology Conference.** Amsterdam, The Political Epistemology Network.

07-08 Dec. 2019 **MindGrad 2019 - Ourselves and Others.** University of Warwick, Coventry.

Dissemination

May 2020 – **Armchair Opinions**
Short pieces, podcast sessions, philosophical debates. Link: https://armchairopinions.org/