Neuroethics (2025) 18:16
https://doi.org/10.1007/s12152-025-09585-6

ORIGINAL PAPER

The Right to Mental Integrity: Multidimensional, Multilayered and Extended

Guido Cassinadri

Received: 25 September 2024 / Accepted: 17 January 2025

Abstract In this article I present a characterization of the right to mental integrity (RMI), expanding and refining the definition proposed by Ienca and Andorno's (*Life Science Society Policy* 13 5, 2017) and clarifying how the scope of this right should be shaped in cases of cognitive extension (EXT). In doing so, I will first critically survey the different formulations of the RMI presented in the literature. I will then argue that the RMI protects from i) *nonconsensual* interferences that ii) *bypass reasoning* and iii) *produce mental harm*. Contrary to other definitions proposed in the literature, my formulation disentangles the RMI from the right to cognitive liberty (RCL) (Lavazza in *Frontiers Neuroscience* 12 82, 2018), the right to mental privacy (RMP) (Lavazza and Giorgi in *Neuroethics* 16 (1): 1-13, 2023), and the right to psychological continuity (RPC) (Zohny et al. in *Neuroethics* 16: 20, 2023), thus enabling a fine-grained assessment of their simultaneous or individual violation. Finally, I analyse how the extended mind thesis (EXT) reshapes the scope of the RMI, proposing a layered protection of extended mental integrity, which grants stronger protection to the organism-bound cognitive system and self in case of manipulative influences of the mind-extending device. To conclude, I present a variety of neurorights violations and mental harms inflicted to organism-bound and cognitively extended agents.

Keywords Neurorights · Right to mental integrity · Mental Integrity · Mental Interference · Extended Mind · Extended Cognition

G. Cassinadri (✉)
Institute for History and Ethics of Medicine, TUM School of Medicine and Health, Technical University of Munich, Munchen, Germany
e-mail: guido.cassinadri@tum.de; guido.cassinadri@santannapisa.it

G. Cassinadri
Scuola Superiore Sant'Anna, Piazza Martiri Della Libertà, 33, 56127 Pisa, Italy

Introduction

In the recent debate on the ethical and legal implications of neuroscience research and the regulation and use of neurotechnologies, the notion of 'neurorights' has been introduced [1–3]. This concept is an umbrella term used to encompass the set of rights that should guarantee adequate protection of the mind and brain of the human person. More specifically, "neurorights, can be defined as the ethical, legal, social, or natural principles of freedom or entitlement related to a person's cerebral and mental domain; that is, the fundamental normative rules for the protection and preservation of the human brain and mind" [1]. Although the lists and terminologies sometimes diverge, four neurorights have been presented and discussed: the right

Published online: 22 February 2025

to cognitive liberty (RCL), the right to mental privacy (RMP), the right to mental integrity (RMI), and the right to psychological continuity (RPC) [1–3].[1]

Within the debate on neurorights, many issues remain open. One example is the need or necessity for the introduction and implementation of new rights within the human rights framework [4]. However, in this discussion, I consider neurorights as moral rights and do not take on any metaethical commitment on their nature, remaining agnostic as to whether they are foundational normative features of the world grounded on the status [5] or authority [6] they confer on the rightholder, or whether they are grounded on their instrumental value in protecting the interests of the rightholder or society at large [7]. The open issue of interest here is the characterization of the right to mental integrity (RMI), which has been defined in different ways without conclusive agreement, as well as its relationship with other neurorights. Thus, I will first present the grounds of the RMI and its different definitions proposed in the debate. Then, I will refine and expand Ienca and Andorno's [2] original definition, clarifying its scope in relationship with other neurorights. In defining what counts as a violation of the RMI, I will consider (i) the type of intervention,(ii) the role of consent; and (iii) the effect of the intervention, distinguishing different types of mental harms.

Given that neurorights reflect fundamental human entitlements related to the brain and mind, their scope, content, and strength can be shaped by an ontology of the mind that includes external artificial substrates as constitutive components of the mental domain, such as the extended mind thesis (EXT) [8]. Among many examples, I will focus on the case of Rita Leggett, an epileptic patient who was implanted with an open-loop BCI advisory system, which alerted her of incoming seizures, enabling her to take proper medications thus living an almost normal life [9]. Given the high degree of integration of the device with her cognitive profile, her acquisition of agential capacities of planning, decision making and self-regulation as well as her conceptualization phenomenological incorporation of the device as a part of her self, Cassinadri and Ienca [10] argue that we can consider her as an agent with an extended mind and self. In this treatment, I will expand on their analysis, clarifying the conditions of self-extension, and distinguishing her from a similar patient whose self was estranged, rather than extended, by the same kind of device [11, 12]. In this way, I will clarify under which conditions a mind-extending tool falls within the expanded scope of the RMI.

Since Rita Leggett unfortunately underwent the unwanted explantation of her implanted BCI, causing her severe and irreversible phenomenological and functional mental harms, I will use her case to assess different forms of mental harms both on a functional and phenomenological level, across two layers of analysis: organism-bound and extended. I will argue that my formulation of the right to extended mental integrity enables us to better capture different kinds of mental harm, compared to other versions of the RMI proposed in the literature. To conclude, I will provide an overview of the various types of extended neurorights violations, defining the scope of the right to cognitive liberty, mental privacy, mental integrity, and psychological continuity.

The Grounds of the Right to Mental Integrity

Some authors who recognize the existence of the right to mental integrity (RMI) present it as the mental equivalent of the right to bodily integrity (RBI). Bodily integrity is conceived as the minimal and fundamental moral and legal protection from certain forms of interferences with one's body, which is possessed and controlled only by the individual himself, thus grounded on the notions of self-ownership and personal sovereignty [13, 14]. Those who support the recognition of the right to mental integrity (RMI) often argue that 1) this right can be distinguished from the right to bodily integrity despite 2) sharing its fundamental justification.

Regarding the first point, they argue that the RMI is distinguishable from the RBI because there are harms and interferences to the mind that are not reducible to harms to the physical brain [15, 16]. This is because there are many forms of detrimental mental effects that could not be captured or reduced to alterations

[1] Yuste et al. [113] proposed a different list of neurorights, including the *right to personal identity*, the *right to free-will*, the *right to mental privacy*, the *right to equal access to mental augmentation*, and the *right to protection from algorithmic bias*. See Bublitz [4] for a critical analysis of this list.

of brain structure, and therefore these harms must be characterized exclusively by their mental properties and domain. The need to distinguish mental integrity from bodily integrity stems from the fact that there are forms of major interferences in the mental domain that do not have a correspondent major change in the brain structure of the individual [3]. For example, brain-washing interventions aiming at changing one's inner psychological states are not morally despicable for their bodily effects, but specifically for their mental effects [15].

Thus, Bublitz [17, 18] and Bublitz and Merkel [15] argue that mental self-determination can be understood as the mental equivalent of bodily integrity and that it is not reducible to it. The right to mental self-determination can be characterized by both a negative and positive dimension: the first concerns the freedom of the individual from external interference, while the second involves the freedom of determining one's own inner realm. Ligthart et al. [3] similarly argue that since self-ownership and personal sovereignty ground the right to bodily integrity (RBI), then these normative principles may ground also the RMI.[2] These kinds of normative justifications for the RMI are usually grounded on the notions of mental self-determination [15], self-ownership and personal sovereignty [3], autonomy, and freedom [15, 19]. However, these justification of the RMI characterize it in terms of mental self-determination, thus overlapping it with the right to cognitive liberty (RCL), which "include(s) both negative freedom from coercion or interference and positive freedom to control one's own brain and mind" [3].

Ienca and Andorno [2] and Ienca [1], in opposition to the overlap of the RMI with the RCL, justify the former on the principle of avoidance of harm, which is more than simple mental self-determination. Some authors argue that the RMI is not only distinct from the RCL, but it even may pose a limit to it, by protecting non-negotiable and inalienable aspects of our mental domain, which requires absolute protection given their association with human dignity [20–22]. Thus, the principle of respect for human dignity [23] operates here as the fundamental grounding of this version of the RMI.

Different Formulations of the Right to Mental Integrity

I will now survey the major definitions of the right to mental integrity (RMI), which can be distinguished in different clusters, despite generally sharing a common justification. The main cluster is the one that defines it in terms of one's right to master and control over one's own brain/mental states and processes free from external alteration [19, 24–28]. The reason why many authors tend to define the RMI in these terms is that it mirrors its bodily counterpart, the right to bodily integrity (RBI), grounding both rights on self-determination. For example, Lavazza [26] defines "mental integrity" as "the individual's mastery of his mental states and his brain data" and the RMI as the principle that states that "no one can read, spread, or alter such states and data in order to condition the individual in any way" without the consent of the individual. This is the broadest characterization of RMI, encompassing any other definition of the concept.

This definition of RMI overlaps with the right to cognitive liberty (RCL), which includes both the negative freedom from coercion or interference and positive freedom to *control* one's own brain and mind [3]. Without further clarification of the kinds of modification covered by the RMI, its violations would be ubiquitous, given that any non-consensual and trivial modification of mental states would apply, thus trivializing the right itself [29]. However, just as there are nonsignificant interventions on our bodily sphere that arguably do not violate the RBI, such as accidentally touching someone's hair, there are also non-significant interventions on our mental sphere. To address this point, Douglas [16, 30] delineated a threshold of significance over which a modification of the mental domain would violate what he calls the "right against mental interference". This threshold is defined by the following criteria: the number of mental states or parts are interfered with, their centrality and/or importance, and the magnitude and permanence of the change [16]. However, as argued by Blumenthal-Barby and Ubel [29], this threshold remains unclear or at least difficult to implement in practice.

[2] Article 3 of the EU's Charter of fundamental rights state that "everyone has the right to respect for his or her physical and mental integrity.".

To clarify this threshold, a second characterization of the negative dimension of the RMI uses the notion of 'mental harm' [1, 3], as originally proposed by Ienca and Andorno [2].

For an action X, to qualify as a threat to mental integrity, it has to: (i) involve the direct access to and manipulation of neural signaling (ii) be unauthorized –i.e. must occur in absence of the informed consent of the signal generator, (iii) result in physical and/or psychological harm. ([2], p. 18).

In medical ethics, harm is classified according to its magnitude, severity, duration, and reversibility [31] and distinguished into three spheres: physical, psychological, and socio-economic. In its negative dimension, according to Ienca and Andorno [2], the RMI protects from brain-washing interventions, malicious brainhacking [32], memory modulation techniques [33], as well as against the neurosurgical risks of infection, bleeding, and rejection of the implanted neurostimulator and potential neuropsychiatric adverse effects including apathy, compulsive behavior and hallucinations stemming from [34]. Zohny et al. [35] argue that the kind of interventions that violates the RMI is only the one that produces a specific type of mental harm, namely alienation from one's mental states. They argue that since mental integrity refers to whole and coherent mental life, alienation is the specific form of harm from which the RMI protects ([35], p. 5).

Continuing in this direction of analyzing the meaning of 'mental integrity' and RMI, the most comprehensive cluster of definitions of the RMI unpacks this notion into its components, considering it an umbrella term encompassing personal identity, agency, autonomy, personhood [19, 28, 36–38]. Thus, the RMI protects the underlying mental states and processes of these elements from interference. For example, Lavazza and Giorgi ([19], p. 15) define the negative dimension of "mental integrity as the protection of and non-interference in certain mental and brain states and processes (correlates of overt mental functions) that are central to an individual's identity, autonomy and worth." Craig ([37], 112) argued that the notion of mental integrity implies a conceptual interconnection with the notions of mind, agency, personal identity, competency, and more broadly, self-authorship. First, it implies the psychological unity and the self-conscious awareness of continued existence over time, which consists with the notion of personal identity of some influential accounts [39]. Second, the autonomy competences consist in the self-authorship capacities for critically reflecting, making decisions and acting according to the reasons, beliefs, desires and commitments that characterize the personal identity of the individual [40]. On these lines, Fuselli [38] argues that the notion of mental integrity involves different interconnected aspects, related to the notions of personal identity, personal autonomy, agency, and cognitive liberty. First, integrity refers to something integral, namely something which is "not fragmented, not divided, not disjointed, not dispersed" ([38], 423). This unity is what defines the borders of the psychological continuity and personal identity of the individual overtime as well as what enables the formation of autonomous will-forming processes. In addition, the notion of agency, referred as the feeling of being the author of one's own choices, decisions and acts, is used in order to provide some contents for the notion of personal identity. To summarize these dimensions of mental integrity, I will use the label 'PIAAAS', encompassing personality, identity, agency, autonomy, authenticity and self [41], considering them as core components of mental integrity.

To conclude, the last characterization of RMI is related to mental health. Ienca and Androno ([2], p. 18) state that "mental integrity in this broader sense should guarantee the right of individuals with mental conditions to access mental health schemes and receive psychiatric treatment or support wherever needed". Following this direction, Wajnerman-Paz et al. [22] define the RMI as a positive right to (medical and non-medical) interventions that restore and sustain mental and neural function and promote its proper development as well as a negative right protecting people from interventions that threaten or undermine these functions or their development. Despite the merit of avoiding the problematic notion of 'control' over one's mental domain, they appeal to the equally problematic notion of 'normal mental functions', which is hard to define given the absence of a natural baseline and limit of development of human cognition due to the great degree of individual differences in cognitive capacities [42–44].

Despite this problem makes difficult to implement the positive dimension of the RMI, their proposal

highlights two important dimensions of mental harm that should be captured and the negative dimension of the RMI: the presence of suffering, distress and pain as well as the presence of a disfunction in exercising a mental capacity.

Types of Intervention and the Role of Consent

Ienca and Andorno [2] originally defined the i) *type of interventions* that violate the RMI in terms of "direct access and manipulation of neural signalling". This formulation arose from the fact that the seminal conceptualization of neurorights emerged from rising concerns about neurotechnological devices capable of accessing and altering brain states and processes. Direct interventions involve a device or a substance that directly interacts with brain states and processes (e.g., deep-brain stimulator or psychoactive substances) affecting them by any other routes but sensual perception. Indirect interventions can be defined as those that are perceived by the affected persons through their outward senses and pass through the mind of the person, being processed by a host of psychological mechanisms[3] ([15, 45], p. 58).

Bublitz [45], Bubltiz and Merkel [15] justify this distinction as morally and legally relevant because whether the interference is direct or indirect affects the degree and kind of control that the agent can exercise over external influences. The ability to control external influences consists in the capacity to detect, filter, engage with, and counteract interventions, which significantly differ in quality and quantity between direct and indirect interventions ([46], p. 63). Thus, direct interventions typically bypass psychological processes, directly affecting and changing the agent's cognitive processes.

However, Zohny et al. [35] argued that, since what is ultimately relevant is the degree of control that the target can exercise over the intervention itself, the directedness of the intervention on the neural signalling is not morally relevant per se. Instead, it is an imperfect proxy for this moral relevance. They present the case of technological interventions that bypass our senses without preventing our ability to perceive the information as if were sensory information: a brain-to-brain interface involves recording information from one brain, sending it to a computer, and then delivering it to a receiving brain via some stimulation technique ([35], p. 7). Therefore, what ultimately distinguishes legitimate alterations, from morally illegitimate intrusive interferences is the degree of rational and conscious cognitive control the receiver can exercise over them, allowing the receiver to deliberate, question, and challenge the information before it indirectly shapes her beliefs, desires, or traits [3, 16, 47].

Therefore, I use Douglas' [16] term 'by-passing influences' to refer to those mental influences that do not engage the autonomous and rational thought of the influenced subject. As we will see in the following sections, unlike Bublitz's [45] direct/indirect distinction, this criterion advantageously also applies to various forms of extended mind manipulations that target the agent's device. Despite the disagreement over which processes constitutes rational [48] and autonomous control [22, 35], these influences encompass both direct and indirect interventions [16, 47] and violate the subject's RCL, understood as the control over one's own brain and mental states and processes [49]. There can be trivial and/or legally protected influences on others' mental domain that mirror non-significant interferences with their right to bodily integrity, such as accidentally touching someone's hair. For example, wearing a perfume that triggers slight emotional responses in others is an action allowed by a legally protected interest of personal expression [45], exerting a non-significant by-passing influence on others' mental domains. However, if a chemist creates a perfume that makes everyone falling in love with him, this would no longer be an insignificant interference, thus violating others' RCL. As mentioned earlier, despite Douglas' [16] criteria for defining significant by-passing mental interferences,[4] there is no clear threshold distinguishing significant from non-significant ones. Unfortunately, there is no space here to develop a full account of this distinction.[5] Notwithstanding the open problem

[3] Indirect interventions encompass verbal communication, psychotherapy, visual and auditory stimuli.

[4] The number of mental states or parts are interfered with, their centrality and/or importance, and the magnitude and permanence of the change [16].

[5] One solution could be to consider all by-passing mental influences as significant, while recognizing the legitimacy

of defining which by-passing mental interference are significant and thus violate the RCL, given my focus on the RMI, I will further distinguish between *significant but non-harmful violations* of RCL from *significant harmful violations* of RCL, which also constitute violations of the RMI.

First, considering ii) *the role of consent*, Ienca and Andorno [2] argued that violations of the RMI occur against the will of the rightsholder. However, Zohny et al. [35], Blumenthal-Barby and Ubel [29], and Wajnerman-Paz et al. [22] question whether the absence of consent is a necessary element of any RMI violation. Blumenthal-Barby and Ubel [29] suggest that consent is neither a necessary nor a sufficient condition for defining the moral permissibility of mental interferences. Zohny et al. [35] argue that there can be a consensual and autonomous decision to receive a DBS stimulation that might lead to a form of alienation, thus undermining the mental integrity of the individual making the decision. Finally, Wajnerman-Paz et al. ([22], p. 2) propose that the RMI might even limit the RCL, as the former is tied to dignity and thus relates to inalienable aspects of our mind—features that are non-renounceable and cannot be violated, even with the authorization of the individual.

When considering mental integrity and its core features, it can be difficult to distinguish which aspects are inalienable and which are not. Arguably, what deserves absolute protection is, at the very least, one's cognitive liberty—that is, the ability to control one's own brain and mental states, and processes, since losing this fundamental ability would constitute a violation of one's dignity. In contrast, other dimensions and components of mental integrity might be autonomously sacrificed in a trade-off [50] decided by the patient who undergoes specific kind of neurotechnological therapies [35]. In such a trade-off, the patient might choose to simultaneously gain and lose different components of his own mental integrity, for example, by gaining therapeutic benefit in terms of mental well-being at the cost of a slight transformation of personality and/or behaviour. This would not qualify as a violation of the subject's RMI, provided that the decision stems from proper informed consent [28, 51]. In this context, the RMI "ought to prevent to a disproportionate relative harm compared to the potential therapeutic benefit." ([2], p. 19). However, if any portion of mental integrity is undermined via neurotechnological therapy without previous informed consent procedure that made the subject aware of this potential outcome, then it would qualify as a violation of the subject's RMI. To put it simply, in line with Ienca and Andorno [2], mental integrity can be undermined whether if consent is present or absent. In the first case, the subject has partially waived her own RMI, sacrificing a negotiable aspect of her mental integrity, while in the second case, it qualifies as a third-party violation of her RMI. Even if we accept the legitimacy of some third-party violations of RCL and RMI justified and balanced by societal reasons [52, 53], they cannot violate some core aspects of mental integrity, as they are absolutely protected by the *principle of respect for human dignity* [23].[6]

Distinguishing the Right to Mental Integrity from Other Neurorights

Now, I will analyse (iii) *the effects* from which the RMI ought to protect against the interferences in our mental domain. In doing so, I will define the scope of the RMI, distinguishing it from other neurorights. By clarifying these distinctions, I aim to offer a more informative and fine-grained framework capable of assessing the violations of simultaneous or individual neurorights, thereby better balancing countervailing interests as well as correlative rights and duties in complex scenarios [10].

As we have seen, Lavazza [26], Lavazza and Giorgi [19] present a broad characterization of the RMI in terms of "the individual's mastery of his mental states and his brain data". However, this definition is insufficiently informative and fine-grained, as it tends to overlap the RMI with right to cognitive liberty (RCL), and the right to mental privacy (RMP). To

Footnote 5 (continued)
only of those that express a personal protected interest, such as wearing perfume [45]. However, it remains open the problem of how to balance conflicting interests and rights. Another option may be to characterize all instances of manipulation as violations of the RCL, but a universally accepted account of manipulation is lacking [114].

[6] I remain agnostic here on whether and which aspects of mental integrity are alienable for societal reasons.

first distinguish the right to cognitive liberty (RCL) from the right to mental integrity (RMI), we must define and differentiate mental integrity and cognitive liberty. I consider mental integrity as the domain including all brain and mental states and processes (whether extended or not). Some of these states and processes form the core of mental integrity, namely the set of capacities and processes that a) enable the individual to exercise a certain degree of control and mastery over one's mental states and processes (i.e., to exercise one's own cognitive liberty) and that b) serve as the substrate of the individual personality, identity, authenticity, autonomy (PIAAAS) [41]. The domains of processes a) and b) overlap, as cognitive liberty is defined in similar terms to autonomy. Furthermore, I tentatively consider the cognitive processes underlying cognitive liberty and autonomy as the inalienable core aspects of mental integrity.

The right to cognitive liberty consists of the right to use one's own core set of capacities and processes for controlling and mastering one's own mental states and processes, free from external interference, while the RMI protects against mental harms affecting one's own mental domain and its core features. My definition of the RCL "include(s) both negative freedom from coercion or interference and positive freedom to control one's own brain and mind" [3]. Some authors argue that the RCL is more fundamental than the RMI, as it provides the foundational ground and justification for all other liberties by serving as their neurocognitive substrate, thereby resembling and conceptually updating the notion of 'freedom of thought' as the essential justification of other freedoms [3, 18]. Other authors argue that some core aspects of mental integrity are associated with human dignity [21] and therefore are inalienable, non-disposable, and deserving of absolute protection [22], thus posing a limit to the RCL. Here, I suggest that the inalienable aspects of mental integrity include the capacity to control one's own brain and mental states and processes, as losing this global capacity would infringe upon one's dignity. To put it simply, one has no right to autonomously functionally undermine the cognitive processes underlying their autonomy and cognitive liberty.

One's RCL can be violated in various ways and to different degrees. Any violation of other neurorights is primarily a violation of the RCL, as it implies an infringement on the negative freedom from interference with one's brain and mental domain. First, a significant bypassing influence that violates one's mental domain may be harmful or non-harmful. I will distinguish between phenomenological and functional harms: the former includes mental pain, suffering and distress, while the latter refers to the instillation of a dysfunction or incapacity in one's mental domain and competences [10, 22].[7] Douglas [16] imagines an evil manipulator that exerts a bypassing influence on Lucien, inducing him to experience a gloomy state (phenomenological harm) while he is dreaming. Then, a second manipulator intervenes with a non-harmful bypassing influence on Lucien, reducing his state of gloominess. These two effects have no long-standing effects on Lucien's mental domain. By overlapping the RMI with the RCL, reducing the former merely to control and mastery of mental states and processes, the distinction between phenomenologically harmful and non-harmful bypassing influences would not be captured, despite being morally significant. Moreover, consider the potential use of DBS to constrain paraphilias in convicted sex offenders [53]. While this measure might be considered as a legitimate violation of one's RCL balanced by societal interests [52], it might cease to be justified if the stimulation also produces high mental pain and suffering in the subject (phenomenological mental harm). Since the dimension of mental suffering, distress and pain is not covered by RCL, it can fall within the scope of protection afforded by the RMI, thereby better distinguishing a class of morally relevant violations of one's RCL. One might argue that inflicting such mental pain on a convicted individual would arguably qualify as a form of torture, which is prohibited as a violation of the *principle of respect for human dignity*. Therefore, the dimension of the RMI that protects from phenomenological mental harm is ultimately already encompassed by this principle. However, the principle of respect for human dignity does not render the RMI superfluous, as this principle is generally recognized as the foundation upon which human rights are based, rather than a specific right [23]. Consequently, the RMI captures the fact that our mental integrity is

[7] The distinction is blurred, since some forms of phenomenological mental harms can also practically cause an inability or dysfunctionality in the agent's mental sphere.

protected by virtue of its association with human dignity ([22], p. 2).

Second, when considering functional harms, we must distinguish between interventions that only violate the RCL from those that also violate the functional dimension of the RMI. An intervention violates the RCL whenever it exerts a significant bypassing influence on the mental domain of the subject. If this significant bypassing interference produces any effect that undermines the control, exercise, and/or development of some functions and capacities of one's own mental sphere, thus producing a functional mental harm, a violation of the RMI occurs. Such violations create a form of mental harm due to the disfunction introduced into one's mental sphere of competences [22].

As the RCL can be violated in different ways, one form is the violation of mental privacy. Informational privacy consists of everyone's entitlement to determine for themselves when, how, and to what extent personal information is communicated to others [54]. In light of the potential vulnerabilities brought about by neurodevices and the inferential potential of advanced data analytic techniques, Ienca and Andorno [2] and Yuste et al. [55] defined the right to mental privacy (RMP) as the individual's right against unconsented intrusion by third parties into their mental information, as well as against the unauthorized collection of such data. More specifically, it establishes that individuals have the right to control access to their own neural data, namely data about brain activity, function and structure, as well as non-neural data, from the analysis of which it is possible to extract information about the mental processes and states [56–58].[8] Given that the negative dimension of the RCL protects from interference with one's brain and mental domain [3], any mental privacy violation implies a violation of the RCL. Wajnerman-Paz [59] argues that privacy ultimately depends on cognitive processes we use for rationally filtering and selectively sharing information about ourselves. Thus, violating the subject's control over these cognitive processes amounts to a violation of her RCL and RMP. Wajnerman-Paz [59] also argues that these cognitive processes used for selectively sharing information about ourselves can be considered as a constitutive component of one's relational, self-constituting narrative identity [60], which is a core feature of mental integrity. This implies that any significant harmful interference that undermines the functionality of these cognitive processes constitutes a violation of RCL, RMP and RMI at its core features.[9] Therefore, there are forms of mental harm that undermine capacities for managing one's privacy mechanisms and safeguards and vice versa, namely some mental privacy violations can more or less directly lead to mental harm.

Lavazza and Giorgi [19] rightly highlight a strong connection between mental privacy and what I defined as core aspects of mental integrity, such as identity and autonomy. If an individual's mental contents are constantly exposed to public or third-party oversight, it is likely that this would cause self-censorship, which would prevent the exercise of relevant mental capacities and experience of mental states. This, in turn, undermines one's identity and autonomy [61], and may also lead to forms of mental harms such as mental distress and anxiety. In this neurotechnological "Big Brother" scenario, one's RCL, RMP and RMI would all be violated simultaneously. Lavazza and Giorgi [19] might further argue that even in less dystopic scenarios in which there is no constant public exposure of every mental state's content, any mental privacy violation directly and inherently implies a detrimental effect on someone's mental domain, thereby violating the RMI. Indeed, Lavazza and Giorgi ([19], p. 4) argue that "it makes sense to include mental privacy in mental integrity, in the broad sense of the term, since making an individual's mental processes public through technological means is tantamount to damaging that individual and undermining their identity, autonomy, and value". For instance, an individual might feel that her identity is compromised through a mind-reading technique that reveals the semantic content of her thoughts [62].

[8] Although I present here a broad definition of mental data, including both neural and non-neural data, I will only consider specific examples of brain-data breach and their relationship with mental integrity.

[9] "violations of a person's mental privacy, disrupting the cognitive control she has over what information about herself she shares or receives may actually affect the very process underlying the formation of her identity" ([59], p. 3).

However, contra Lavazza and Giorgi [19], I argue that it makes more sense to conceptually distinguish between the RMI and the RMP for several reasons. First, not every mental privacy violation necessarily infringes upon the RMI, and vice versa. For example, if a subject is unaware of a breach to her mental privacy, then she will not practice self-censorship on her thinking processes, thus undermining her autonomy and identity. Moreover, it is possible to inflict mental harm without violating someone's privacy, as in the case of Rita Leggett [10].

Second, even if a subject becomes aware of a privacy breach involving information about her neural data, the breach may not necessarily lead to self-censorship, depending on the type of neural data, such as in the case of data concerning brain structure [57]. Additionally, depending on the type of neural data, a subject aware of a mental privacy breach may personally dismiss its impact on her identity and autonomy.[10] To conclude this point, consider the hacking of a web-based cloud server with neural data gathered from a BCI [63]. Even if this mental privacy violation could, in principle, expose subjects to future interference, such as the harmful disabling of their BCI, these consequences do not automatically and necessarily follow.[11] Thus, their RMI is not automatically violated.

Third, even if we were to empirically discover that every mental privacy violation always leads, in one way or another, to a form of mental harm (which is unlikely), it would still make sense to conceptually distinguish the RMI and RMP given their different normative foundations. The RMI is grounded on the principle of harm avoidance [1, 2] and human dignity [20–22], while the RMP is grounded on the right to control one's own information [1, 2, 54]. To conclude, as individual violations of the RMI and RMP separately occur in different degrees, forms and magnitudes, encompassing the RMP within the RMI intended in the "broad sense of the term" would conceal normatively relevant differences between different kinds of interferences.

To support the conceptual distinction between the RMI and the RMP, let us consider different scenarios related to brainjacking of a DBS (deep-brain stimulator), namely the unauthorized control of another's electronic brain implant [32, 64]. In the case of blind attacks on a DBS, the hacker can damage brain tissue by over-stimulation without any knowledge of the patient condition, thus violating his RCL and RMI without infringing on the victim's privacy. In another scenario, an evil spy may temporarily deprive the patient of access to the external programmer, which is the device used by patients and clinicians to program stimulation parameters, without interfering with it. This action would violate the patient's RCL and RMP, due to the third-party's access to relevant mental information and the control of a mind-altering tool, but not his RMI, since no actual mental harm has yet occurred. However, if the deprivation is constant and prolonged in time, then the RMI would also be violated, as the patient would be deprived of an important contributor to his mental well-being. This example highlights that a violation of one's control over her own mental capacities (RCL) can gradually produce a functional mental harm, depriving the agent of some of her mental functions and capacities, thus violating her RMI. Finally, a malicious hacker might realize a targeted attack, exploiting illicitly accessed knowledge concerning a patient's mental data [56] related to his chronic pain condition and his DBS settings, causing him additional pain by activating the stimulation [64]. This case represents a violation of the patient's RCL, by depriving him of the control over his mental sphere, RMP, by accessing intimate information on his mental condition, and RMI, by inflicting him additional pain. Thus, a mental privacy violation does not automatically imply the violation of the right to mental integrity (RMI) of a subject, and the latter comes in degrees of seriousness.

Lavazza and Giorgi's [19] characterization of the RMI and its relationship with the RMP fails to grasp the differences between these scenarios and their related degrees of seriousness, simply considering

[10] Consider, for example, the case of a patient interviewed as part of the Hybrid Mind project (https://hybridminds.webflow.io/) who dismissed the mental privacy concerns potentially affecting their DBS, arguing that they were more worried about the privacy issues related to their smartphone [115]. Patient interviews reveal a high degree of individual variability regarding the subjective importance and normative weight attributed to ones' mental privacy, depending on the specific neural data at stake.

[11] I will present below different examples of brainjacking to illustrate this point. The general point is that it would be unfair to morally hold the hacker responsible and criminally prosecute him for the infliction of mental harm before this mental harm has occurred.

them as violations of mental integrity, intended in a broad sense. I acknowledge that some mental privacy violations can imply a violation of RMI, by causing distress and/or violating one's identity and autonomy via self-censorship [19], and some violations of RMI also imply a violation of RMP, when a harmful mental interference undermines the cognitive capacities necessary for managing one's mental privacy [59]. However, given the possible individual violations of the RMI and RMP, it makes sense to conceptually distinguish concerns about privacy and data ownership (mental privacy) from concerns about mental harm (mental integrity) [29, 35]. While unwanted access to one's mental states might expose the data subject to mental and PIAAAS-related harms, these do not automatically and necessarily follow. To conclude, given that the individual violations of RMI and RMP come in different degrees, forms and magnitudes, by simply combining them as violations of mental integrity "intended in a broad sense", as proposed by Lavazza and Giorgi [19], would hide normatively relevant differences rather than clarifying them.

Regarding the degree of seriousness of mental harms, I conclude by considering a variation of the last brainjacking scenario in which the mental harm is of a special kind, namely the infliction of a transformation of his character traits, such as behaviours, values and preferences [65]. PIAAAS-related harms are a special subset of phenomenological and/or functional mental harms on the personality, identity, agency, autonomy, authenticity and self of the subject [41]. According to Ienca and Andorno ([2], p. 21), the right to psychological continuity (RPC), a special case of the RMI, provides protection from a special class of mental harms that undermine the personality and personal identity, as well as the coherence of the individual's behaviour and the continuity her habitual thoughts, preferences, and choices. DBS-induced side effects can induce relevant changes in one's PIAAAS dimensions [65], potentially violating the subject's RPC if they occur without proper informed consent. Moreover, memory-modulating techniques [66] can negatively impact a person's identity by selectively removing, altering, adding, or replacing individual memories that are relevant to their self-recognition as persons [2]. Zohny et al. [35] narrows the scope of the mental harms covered by the RMI to alienation, thus overlapping and reducing the RMI with the RPC. However, it is important to not reduce all forms of mental harm captured by the RMI to those serious cases that also violate the RPC. Given the multidimensional character of mental integrity, its different components may have different normative weight. These different normative weights are relevant both in cases of a patient's autonomous and informed decision regarding a therapeutical path that involves trade-offs between different components of mental integrity, as well as in case of third-party neurorights violation [10].

A Multidimensional, Multilayered and Extended Right to Mental Integrity

To sum up my previous considerations, the RMI that I propose is characterized by a positive and a negative dimension. The former is formalized by Wajnerman-Paz et al. [22] as well as Ienca and Andorno [2] in terms of access to medical and non-medical mental health and psychiatric treatments and interventions that restore and sustain mental and neural function and promote its proper development. The negative dimension protects from:

A) Direct or indirect significant by-passing mental influences;
B) That happen in absence of informed consent of the rightholder;
C) producing mental harm, which can be
    i) phenomenological and/or functional;
    ii) either organism-bound or extended.

Phenomenological mental harms include mental pain, suffering and distress, while the functional mental harms cover induced disfunctions or incapacities in one's mental sphere of competences [22] (Fig. 1).

Mental harms come in degrees depending on the magnitude, severity, duration, and reversibility [31] as well as on the centrality of the impacted aspects of the mental domain ([19], p. 15). PIAAAS-related harms are a special subset of phenomenological and/or functional mental harms that impact the personality, identity, agency, autonomy, authenticity and self of the subject, which are core aspects of mental integrity [41]. The right to psychological continuity (RPC) protects from the disruption of one's personality, personal identity and self [1]. Thus, my version

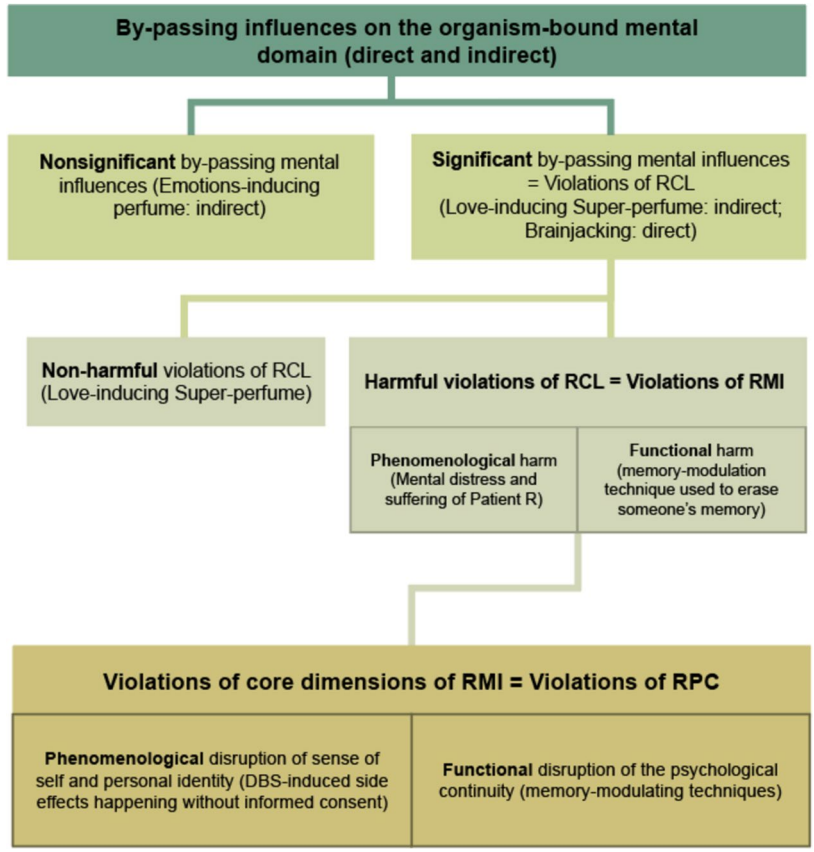

Fig. 1 Varieties of organism-bound mental interferences

of the RMI refines Ienca and Andorno's [2] original formulation, by including the two forms of harms individuated by Wajnerman-Paz et al. [22] and considering PIAAAS related harms as special violations of the RMI.

According to the extended mind and extended cognition theses (EXT), if we embrace a functionalist theory of the mind, then, under specific coupling conditions between a cognitive agent and a device that functionally contributes to the realization of some of the agent's mental states and cognitive processes, such a tool may be considered as a constitutive part of the extended realization base of such mental states and cognitive processes [8, 67, 68]. We can distinguish between three forms of cognitive coupling between an agent and a tool: action-perception loops, extended circuits, and hybrid circuits [69]. The former is characterized by perceptual and bodily engagement with an external representational device, such as a calculator, the second applies to devices directly implanted in an individual's brain and subpersonally integrated into their cognitive routines, such as a closed-loop DBS system [70],while the latter applies to devices that are both directly integrated into the subpersonal processes of the brain and involved in action-perception looping operations, such as open-loop BCI used for alerting epileptic patients of incoming seizures [10]. The conditions under which an artifact becomes a constitutive component of the agent's extended cognitive system, distinguishing it from a merely embedded and casually integrated one [71, 72] are highly debated in the literature [73]. An emergent solution consists of embracing an integrationist and multidimensional approach to assessing and explaining hybrid cognitive ensembles[12] [72, 74–78]. According to this integrationist approach cognitive

[12] Heersmink [74, 75] considers the dimensions of reliability, durability, trust, procedural and representational transparency, individualization, bandwidth, speed of information flow, distribution of computation, and cognitive and artifactual transformation. Heinrichs [76] introduced the dimensions of irreplaceability of the tool and dependency of the agent.

extension does not depend simply on discrete conditions, rather, different degrees of integration between the agent and the tool across various dimensions correspond to more or less integrated cognitive systems. Despite the blurred threshold, highly integrated cognitive systems correspond to extended cognitive systems ([79], p. 434).

Douglas [16], Tesink et al. [80], and Cassinadri and Ienca [10] argue that the thesis of extended cognition (EXT) implies an expansion of the scope of the right to mental integrity, as the device becomes a constitutive component of the agent's mind. Thus, for example, removing a mind-extending BCI would constitute an infringement of the patient's right to (extended) mental integrity [10, 80].[13] This has long been acknowledged in the debate on the moral implications of EXT in terms of the device's acquisition of moral status by virtue of its cognitive status "since the degree of dependency and integration [is] proportional to the artifact's moral status" [78, 79, 81, 82]. However, the tool's acquisition of moral status is not so straightforward as it is usually presented [83, 84]. Among the authors who recognize the extension of the moral status of a mind-extending tool, Hernández-Orallo and Vold ([81], p. 512) also consider the risk of the tool's 'interference and control' over the agent, particularly for those AI-extenders that can "model and monitor human behavior to find the targeted interventions that optimize some metric of cognitive enhancement, and can degenerate into sophisticated ways of surveillance and manipulation, well beyond the relatively decoupled smartphones and 'nudging' personal assistants of today". Arguably, this kind of manipulation may count as by-passing mental influences, potentially undermining both the RCL as well as the RMI [20, 49]. Thus, there can be highly integrated devices that, on one hand, may be considered as constitutive components of the agent's minds according to an integrationist framework of EXT [69, 75, 76], thus acquiring moral status, while on the other, exert a detrimental and/or manipulative influence on the agents' interests [20, 85] and cognitive profile [86], namely on their mental integrity. This concern applies particularly well to those implantable Brain-Computer Interfaces [11, 12] and DBS systems [70], which are highly integrated to the brain of patients, modulating their cognitive and affective states and processes [87], and potentially giving rise to unwanted side-effects such as personality changes as well as feelings of ambiguous agency, thus potentially leading to PIAAAS-related harms [34, 65]. So, how should we balance the acquisition of the device's moral status with the protection of the organism-bound agent from the potential manipulations of such a device?[14]

The simple solution consists of hierarchically assigning moral status and priority of protection first to core biological cognitive vehicles of the organism-bound agent, then to external and mind-extending devices, and finally to external causal enablers of our cognitive capacities [83].[15] Regarding the expanded scope of the RMI implied by EXT, this approach leads to a layered interpretation of the right to extended mental integrity (EXT-RMI), which allows us to recognize both the extended moral status of the mind-extending tool as well as the moral priority of the organism-bound agent over her potentially manipulative mind-extending device. The right to extended mental integrity should therefore be characterized by two hierarchical layers:

i) The first layer protects the organism-bound agent *from* harmful by-passing influences, either direct or indirect. This layer protects the organism-bound agents *from* potentially detrimental influ-

[13] Clark ([116], p. 215) acknowledged long ago that damaging or removing a mind-extending tool «has an especially worrying moral aspect: it surely is harm to the person, in about as literal a sense as can be imagined». Carter and Palermos [117] formalized it in legal terms as a form of 'extended personal assault'.

[14] Despite Douglas [16], Tesink et al. [80], and Cassinadri and Ienca [10] first argued that EXT implies an extension in scope of the RMI, they did not address the case in which a mind-extending tool also exerts a mentally detrimental influence on the organism-bound agent, a concern first addressed by Biber and Capasso [20].

[15] The moral status of a mind-extending tool is always relational and derivative from the moral status of the organism-bound agent to which the device is coupled [83, 92]. This means that since the organism-bound agent is the source of the moral worth and consideration of the device, once the device exercises a manipulative or controlling influence on the organism-bound agent it undermines its own source of moral value.

ences of a device integrated to their cognitive system.

ii) The second layer affords a protection *to* the devices against by-passing interferences that alter their functioning, thereby undermining the extended mental integrity of extended mental agents.

Neurorights Violations Against Extended Minds and Selves

At this point, two questions emerge: When does a mind-extending tool respect or undermine the organism-bound agent's mental integrity? When does a device expand the scope of the RMI? Some authors address the first question assuming a distinction between elements that constitutively contribute the cognitive system, and those that are a constitutive component of the self, identity and personhood ([88], p. 2), [89], thus differentiating instances of extended cognition (EXT) from the ones of extended self (EXT-S) [90, 91] and extended personhood (EXT-P), namely the morally relevant boundaries of the individual [92, 93]. In contrast, Tesink et al. ([80], p. 5) simply assume that the criteria for cognitive extension —such as "close physical proximity to the brain", "continuous intimate interaction with (other) mental processes" and "significant role in the mental functioning of a person"—are sufficient for both EXT as well as for an extended right to mental integrity (EXT-RMI), presenting the case of Rita Leggett in support of their argument. However, in the clinical trial in which 'patient R' (Rita Leggett) participated, there was also another patient that can serve as a counterexample to Tesink et al. [80]. Gilbert et al. [11] and Postan [12] present the case of this patient ('patient S'), who was as well suffering from chronic epilepsy, but had previously not characterised herself as epileptic: she "pretended that [her epilepsy] didn't really exist" [11]. In turn, the use of an open-loop BCI advisory system made her "feeling sick" all the time, as if she "didn't have control over what [she] was going to do" [11].

Postan [12] argues that the participant's response indicates that the disruption to her existing self-conception is indeed contrary to her interests because it threatens her ability to 'live with' who she is and to have an intelligible sense of self, thus grounding her engagement in the world. Gilbert et al. [11] described her case in terms of 'self-estrangement', which refers to an ethically significant, abrupt, and involuntary change in an individuals' qualitative experience of self, rendering them unable to access or identify with who they were beforehand. This is a form of mental harm that Zohny et al. [35] would characterize in terms of alienation. Since this kind of organism-bound phenomenological and functional identity-related harm was induced by a tool that respected the shallow criteria for cognitive extension (EXT) mentioned by Tesink et al. [80], these criteria proved insufficient for capturing also the extension of the self (EXT-S). Therefore, to properly expand the scope of the right to mental integrity, we can pursue two different directions. Either we restrict the cases of EXT to the ones that also extend the self and personhood of the agent, or we argue that the extended right to mental integrity protects only some special portions of the agent's mind, as proposed by Douglas [16]. Given that EXT is primarily a thesis developed within the philosophy of cognitive science for explanatory purposes of cognitive phenomena [68], I propose to follow the second option, arguing that the RMI expands its scope over those devices that extend the mind, self, and personhood of the rightholder.[16]

Given that both patient R and patient S met the criteria of i) close physical proximity to the brain, ii) continuous intimate interaction with (other) mental processes, and iii) significant role in the mental functioning of a person [80], it is useful to explore the relevant differences between them to better define the criteria for the extension of self and personhood. The first one arises at a phenomenological level: while patient R felt empowered and incorporated the device in her sense of self [9], patient S, by contrast, felt alienated and estranged [11]. This difference stems from their initial lived relationships with their own pathological conditions—R acknowledged and accepted hers, whereas S ignored and rejected hers. This, in turn, affected their levels of trust in the device: R embraced it, while S rejected it. Thus, while phenomenological incorporation of the device, may not be necessary for extending cognition [94], it appears to be a necessary condition for the extension of the specific portion of the mind called 'self' [88, 95]. Another relevant condition for the extension

[16] Biber and Capasso ([20], p. 509) argue that the criteria for cognitive extension, taken in isolation, risk to neglect the normative and justificatory role that such criteria should have, beyond their cognitive role.

of the self is the degree of 'reflective transparency', defined as the agent's "ability to control and reflectively focus upon how whether she is meeting her goals and how well her various devices support her aims" [88]. This allows the agent to evaluate the alignment between the device's influence and her set of values, preferences and commitments.

Since this problem of alignment has been largely discussed in the literature on DBS-induced personality changes [70, 96], I will rely here on Pugh et al.'s [65] framework of authenticity and autonomy of the self. Assessing alignment is not an easy task, as the true (*authentic*) self is defined by the set of cohering and evolving elements of the individual's nexus of values, beliefs, preferences [65, 97]. Within this framework, if a transformation of some components of the self happens in respect and continuity with the evolving core set of desires, values and commitments, then the authenticity and autonomy of the self are empowered, rather than endangered. This is what happened to patient R, but not to patient S. Given that the second had already a dysfunctional and alienated relationship with a component of her self—namely, her pathological condition—the BCI exacerbated further disruption of her sense of self, causing a phenomenological mental harm. For another example, consider again the case of a sex-offender forced by a judge to wear a brain stimulator that prevents his sexual attacks [52, 53]. If the sex-offender reflectively disapproves the 'offending' part of himself and rationally appreciate the treatment according to his values and beliefs, he may perceive himself as a new, authentic, more autonomous, and thus extended version of himself. The device may enhance his autonomy, aligning his behavior and first-order desires to his higher values and commitments [98]. Therefore, a brainjacking on his device would arguably qualify as a violation of his right to extended mental integrity (EXT-RMI), as would be the case for Patient R. Alternatively, the sex-offender may disapprove of this forced treatment as a violation of his autonomy and RCL, given its misalignment to his values and/or preferences, as in case of Patient S. Like patient S, this second sex offender may experience and consider the device as an external, alienating and constraining influence on his mind and self [84], which would likely fall outside the expanded scope of the EXT-RMI, as it would fail be part of his extended mental integrity.

The Variety of Mental Harms and Violations of Extended Neurorights

I will now conclude by providing an overview of the different forms of mental harms that impliy a violation of my version of the right to extended mental integrity (EXT-RMI), in relation to other neurorights. Consider again the case of R's unwanted explantation [9]. First, since the BCI played a crucial causal role in managing her mental domain and arguably extended her mind and self, she lost control over a significant portion of her mental domain, thus having her EXT-RCL violated [10]. Second, since the device enabled her to live an almost normal life by predicting her incoming seizures, she suffered from the return of epileptic attacks, which qualify as an *organism-bounded functional* mental harm. Third, she experienced an *organism-bounded phenomenological* mental harm, as the pain and distress she felt occurred within the boundaries of her organism after the explantation. Part of this mental harm also involved a disruption of her sense of self, as her self-narrative, in which she conceptualized herself as endowed with new capacities (planning, self-regulation and decision-making), was disrupted [10]. Additionally, given that some of her cognitive capacities were extended by the device, she suffered from *extended functional* mental harm. Considering the set of embodied and cognitive capacities as a component of the self [99], part of this harm consists in the amputation of a portion of the material vehicles constituting her self, constituted by her extended capacities [88, 95]. Therefore, the self-related mental harms must be assessed along two axes: 1) organism-bound and extended, 2) phenomenological and functional (Fig. 2).

These self-related harms imply a violation of her RPC, which protects "people's personal identity and the continuity of their mental life from unconsented external alteration by third parties" [2]. However, she did not experience any mental privacy violations, as there was no unwanted access to, use of, or sharing of her brain and mental data [10]. Other versions of the RMI that include mental privacy within mental integrity [19], or that defines the latter merely in terms of control of one's mental states and processes [19, 26], deprivation of support to normal neural functions [22], or infringement of personality, agency [37] and autonomy [38], fail to capture the distinct, multiple, complex and simultaneous dimensions of mental

harms she suffered from. Lavazza and Giorgi [19], overlap the RMI with the RCL, defining it in terms of control and mastery of one's mental states and processes, while Zohny et al. [35] include only alienation as a relevant mental harm, thus conflating the RMI with the RPC. However, this overlap obscures morally distinct types of harm. The same applies to the extended right to mental integrity proposed by Tesink et al. [80], as they simply define the RMI in terms of a protection from non-consensual mental interferences, failing to assess the multiplicity and variety of harms that patient R experienced in a fine-grained manner. In contrast, my multidimensional and multilayered characterization of the extended RMI enables us to capture different kinds of harms, thereby facilitating the definition of the correlative duties on third parties to avoid them [10]. Indeed, the role of neurorights in medical ethics also includes recognizing the specific vulnerabilities related to the brains and minds, as well as implementing specific safeguards to prevent potential harms. Since the degree and strength of duties and responsibilities toward vulnerable individuals are influenced by their level of vulnerability [100–103], distinguishing, on a fine-grained level of detail, the variety of harms and vulnerabilities to which subjects are exposed clarifies the nature, force and scope of these responsibilities [10].

I will conclude considering the final relevant forms of extended neurorights violations, whether in cases of *extended circuit* or *action-perception* EXT [67, 69]. One advantage of abandoning the direct/indirect distinction [45] in favor of Douglas' [16] bypassing criterion is that it applies equally to all versions of EXT as well as to organism-bound cases. Starting from the *extended circuit* EXT, I note that *phenomenological* mental harms are generally *organism-bounded*, as an *extended phenomenological* mental harm would require the extension of the material substrate of consciousness [104]. Even if we can imagine the brainjacking of a futuristic consciousness-extending brain chip that induce a phenomenological mental harm, this case could still be reframed in terms of an indirect organism-bound phenomenological mental harm, similar to what happened to patient R. However, if this kind of extended conscious mind violations were to create new and unexpected forms of phenomenological mental harms, they would genuinely qualify as *extended phenomenological* mental harms.

Let's consider now cases of action-perceptions EXT in which the agent receives indirect bypassing interferences. Cloud-Otto is a contemporary version of Otto from Clark and Chalmers's [8] original thought experiment—an Alzheimer's patient who relies on his smartphone to cope with everyday cognitive tasks, adhering to the conditions for the extension of his mind and self [88]. The smartphone is connected to his own data repository, and he uses the device to access, track and revise his set of extended beliefs, desires and plans. This agent is vulnerable to various forms of by-passing extended mental interferences that could potentially violate his extended neurorights. We can draw a continuum of significance and harmfulness, ranging from completely insignificant and harmless by-passing extended mental interferences (such as a negligible change in an app setting on Cloud-Otto's smartphone) to significant and harmful ones that undermine extended cognitive functions and/or alter a portion of the agent's self. Significant by-passing influences of the mental domain would undermine the EXT-RCL; extended (functional) mental harms would violate also the EXT-RMI; and extended PIAAAS-related harms could potentially violate the EXT-RPC.

Despite the difficulty of distinguishing between significant and non-significant alterations, we can preliminarily rely on Carter's [105] framework. He

Mental harms of patient R	Functional	Phenomenological
Organism-bound	Return of epileptic seizures	Disruption of sense of self
Extended	Deprivation of extended cognitive abilities, which partly constitute her extended self.	

Fig. 2 Types of mental harms inflicted to patient R

presents two forms of 'extended thought manipulations' that would qualify as a violation of the agent's right to extended cognitive liberty (EXT-RCL) through the manipulation of autonomous propositional attitudes—attitudes attributable to the agent in light of the set of desires, preferences and commitments that characterize her self: i) the *inception* of non-autonomous extended propositional attitudes; and ii) the *eradication* of extended autonomous propositional attitudes. Carter [105] identifies two jointly sufficient conditions for non-autonomy: the bypassing of mental capacities [16] and impossibility for the subject to give up or attenuate the strength of the relevant attitude. An autonomous propositional attitude is non-autonomously *eradicated* if it is caused to be either eliminated or blocked from manifesting in ways that relevantly bypass a thinker's cognitive and conative faculties ([105], p. 7). These violations of EXT-RCL can be harmful or non-harmful and may or may not infringe on the subject's right to extended mental privacy (EXT-RMP) [106] (Fig. 3).

First, considering extended PIAAAS-related mental harms that would violate the EXT-RPC, Heersmink [107–109], adopting a practical and narrative view of personal identity [110–112], presents the case of life-logging technologies used by Alzheimer's patient to maintain and extend their narrative identity. These devices can record, store and display some of the subjects' personal experiences, providing them access to an externalized portion of their narrative self-contents, which they can use to inform their self-interpretation and engagement with the world [110, 111]. Thus, the erasure of relevant *mind-and-self* extending information stored in life-logging devices that play a crucial role in one's narrative identity would qualify as an *extended functional*

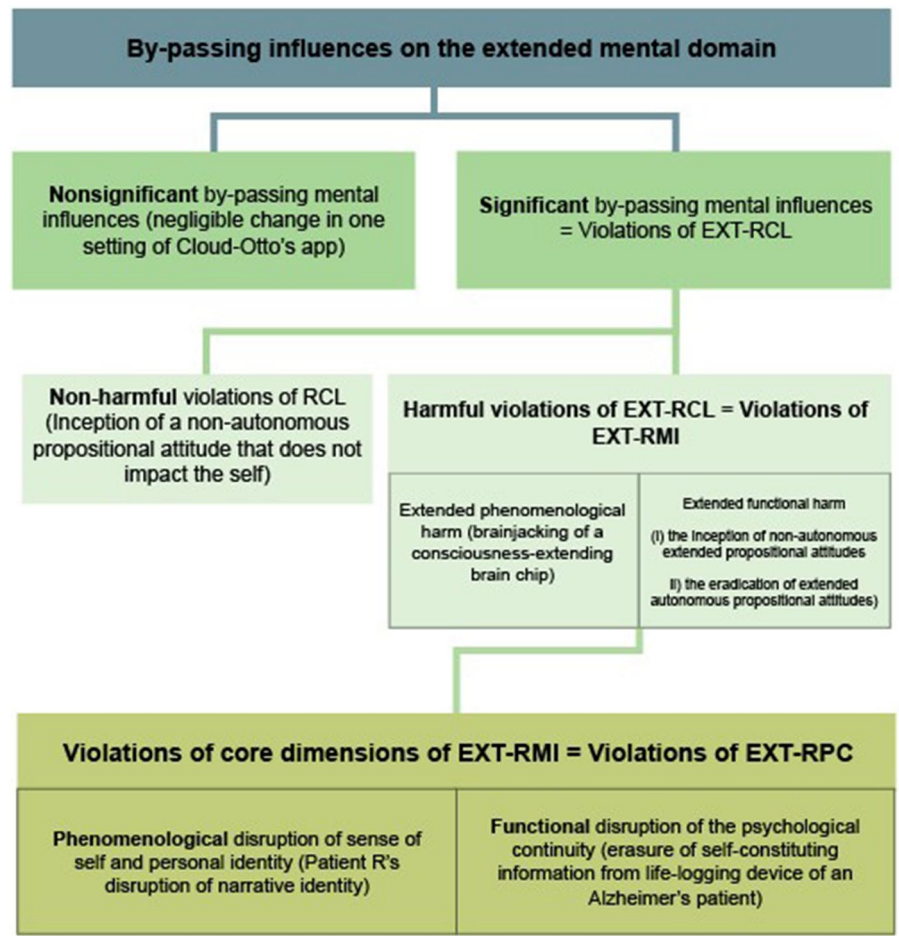

Fig. 3 Varieties of extended mental interferences

mental harm that violate the subject's EXT-RCL, EXT-RMI and EXT-RPC.

To conclude, as in the case of brainjacking discussed earlier, a privacy breach of information stored in mind-extending life-logging device is a violation of the EXT-RMP that does not automatically expose the subject to extended mental harm. Despite the potential for future manipulation and the violation of the subject' control over her cognitive domain (EXT-RCL) [61], these threats may not automatically result in mental harm, especially if she is unaware of the privacy breach. Therefore, it is conceptually better to distinguish the EXT-RMI from the EXT-RMP, given the distinction between the RMI and RMP [29, 35]. Nevertheless, there might be violations of EXT-RMI that undermine those cognitive capacities responsible for the management of one's external storage of mental information [59], thus undermining also the subject's EXT-RMP.

Conclusion

My analysis aimed to show that a multidimensional and multilayered characterization of the right to extended mental integrity enables us to pursue various desiderata. i) It clearly distinguishes the RMI from other neurorights, while simultaneously highlighting their intersections. ii) It identifies, at a more fine-grained level, the different types of mental harms and vulnerabilities to which human cognitive agents are exposed. iii) It clarifies the hierarchical levels of importance and prioritization of protection between the different aspects of mental integrity, distinguishing its core and inalienable aspects from its negotiable ones as well as the organism-bound portions of the mind from the extended ones. Nevertheless, further work is needed to clarify the unresolved grey areas: 1) the threshold of significance of mental interferences, whether organism-bound and extended; 2) the legitimacy of third-party violations of RCL and RMI; 3) the distinction between alienable and non-alienable core aspects of mental integrity; 4) the threshold at which a violation of control also results in a functional mental harm; 5) the utility of the fine-grained assessment of mental harms for distributing correlative responsibilities and duties on third parties. Despite these open issues, this work seeks to offer a promising and comprehensive framework for characterizing the right to mental integrity.

Acknowledgements I am deeply thankful to Marcello Ienca for his precious, constructive and constant feedback. I am also grateful to the members of the Institute of History and Ethics of Medicine at TUM for providing constructive feedback on this article.

Authors' Contributions G.C. performed all the work during the research and writing of the article.

Funding Open Access funding enabled and organized by Projekt DEAL. I benefited from a DAAD short term scholarship for this research.

Data Availability N/A.

Declarations

Ethics Approval and Consent to Participate N/A.

Consent for Publication N/A.

Competing Interests The author declares no competing interests.

References

1. Ienca, M. 2021. On Neurorights. *Frontiers in Human Neuroscience* 15 https://doi.org/10.3389/fnhum.2021.701258.
2. Ienca, M., and R. Andorno. 2017. Towards new human rights in the age of neuroscience and neurotechnology. *Life Science Society Policy* 13: 5.
3. Ligthart, S., Ienca, M., Meynen, G., Molnar-Gabor, F., Andorno, R., Bublitz, C., … Kellmeyer, P. 2023. Minding Rights: Mapping Ethical and Legal Foundations of 'Neurorights.' *Cambridge Quarterly of Healthcare Ethics* 32 (4): 461–481. https://doi.org/10.1017/S0963180123000245.
4. Bublitz, J. C. 2022. Novel Neurorights: From Nonsense to Substance. *Neuroethics* 15: 7. https://doi.org/10.1007/s12152-022-09481-3.

5. Kamm, F. M. 1992. Non-consequentialism, the person as an end-in-itself, and the significance of status. *Philosophy &Public Affairs* 21 (4): 354–389.
6. Owens, D. 2019. Property and Authority. *Journal of Political Philosophy*. 27 (3): 271–293.
7. Wenar, L. 2005. The value of rights. In *Law and social justice*, ed. J.K. Campbell, M. O'Rourke, and D. Shier, 3–179. Cambridge, MA: MIT Press.
8. Clark, A., and D. Chalmers. 1998. The extended mind. *Analysis* 58 (1): 7–19.
9. Gilbert, F., M. Ienca, and M. Cook. 2023. How I became myself after merging with a computer: Does human machine symbiosis raise human rights issues? *Brain Stimulation* 16 (3): 783–789. https://doi.org/10.1016/j.brs.2023.04.016.
10. Cassinadri, G., and M. Ienca. 2024. Non-voluntary BCI explantation: Assessing possible neurorights violations in light of contrasting mental ontologies. *Journal of Medical Ethics*. https://doi.org/10.1136/jme-2023-109830.
11. Gilbert, F., M. Cook, T. O'Brien, and J. Illes. 2019. Embodiment and estrangement: Results from a first-in-human "intelligent BCI" trial. *Science and Engineering Ethics* 25 (1): 83–96.
12. Postan, E. 2021. Narrative Devices: Neurotechnologies, Information, and Self-Constitution. *Neuroethics* 14: 231–251. https://doi.org/10.1007/s12152-020-09449-1.
13. Thomson, J. J. 1990. Trespass and first property. *The realm of rights*, 205–226. Cambridge: Harvard University Press.
14. Ripstein, A. 2006. Beyond the harm principle. *Philosophy & Public Affairs* 34 (3): 215–245.
15. Bublitz, J. C., and R. Merkel. 2014. Crimes Against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination. *Criminal Law, Philosophy* 8: 51–77.
16. Douglas, T. 2024. An Intuitive, Abductive Argument for a Right against Mental Interference. *Journal of Ethics* https://doi.org/10.1007/s10892-024-09476-7.
17. Bublitz, J. C. 2013. My Mind Is Mine!? Cognitive Liberty as a Legal Concept. In: Hildt, E., Franke, A. (eds) *Cognitive Enhancement. Trends in Augmentation of Human Performance*, vol 1. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-6253-4_19.
18. Bublitz, J. C. 2015. Cognitive liberty or the international human right to freedom of thought. In Levy, N. and Clausen, J. (Eds). *Handbook of Neuroethics*, Springer: 1309–1333.
19. Lavazza, A., and R. Giorgi. 2023. Philosophical foundation of the right to mental integrity in the age of neurotechnologies. *Neuroethics* 16 (1): 1–13.
20. Biber, S. E., Capasso, M. 2022. The Right to Mental Integrity in the Age of Artificial Intelligence: Cognitive Human Enhancement Technologies. In: Custers, B., Fosch, Villaronga, E. (eds) Law and Artificial Intelligence. Information Technology and Law Series, vol 35. T.M.C. Asser Press, The Hague. (Rainey et al., 2020).
21. Valera, L. 2022. Mental Integrity, Vulnerability, and Brain Manipulations: A Bioethical Perspective. In: López-Silva, P., Valera, L. (eds) *Protecting the Mind. Ethics of Science and Technology Assessment*, vol 49. Springer, Cham: 99–113. https://doi.org/10.1007/978-3-030-94032-4_9
22. Wajnerman-Paz, A., F. Aboitiz, F. Álamos, and V. P. Ramos. 2024. A healthcare approach to mental integrity. *Journal of Medical Ethics*. https://doi.org/10.1136/jme-2023-109682.
23. Andorno, R. 2009. Human dignity and human rights as a common ground for a global bioethics. *Journal of Medical Philosophy* 34 (3): 223–240. https://doi.org/10.1093/jmp/jhp023.
24. Birks, D., and A. Buyx. 2018. Punishing intentions and neurointerventions. *AJOB Neuroscience* 9 (3): 133–143.
25. Inglese, S., and A. Lavazza. 2021. What should We Do with people who cannot or Do Not want to be protected from neurotechnological threats? *Frontiers in Human Neuroscience* 15: 703092.
26. Lavazza, A. 2018. Freedom of thought and mental integrity: The moral requirements for any neural prosthesis. *Frontiers in Neuroscience* 12: 82.
27. Douglas T., Forsberg L. 2021. Three rationales for a legal right to mental integrity. In: Ligthart S et al (eds) *Neurolaw, Palgrave studies in law*. Neuroscience, and human behavior: 179–201.
28. Hildt, E. 2022. A Conceptual Approach to the Right to Mental Integrity. In: López-Silva, P., Valera, L. (eds) Protecting the Mind. Ethics of Science and Technology Assessment, vol 49. Springer, Cham: 87–99 https://doi.org/10.1007/978-3-030-94032-4_8.
29. Blumenthal-Barby, J., and P. Ubel. 2024. Neurorights in question: Rethinking the concept of mental integrity. *Journal of Medical Ethics* 50 (10): 670–675. https://doi.org/10.1136/jme-2023-109683.
30. Douglas. T. 2022. The scope of the right against mental interference. Presented at the workshop: The Ethics of Influence 2022.
31. Meslin, E. M. 1990. Protecting human subjects from harm through improved risk judgments. *IRB Ethics & Human Research* 12: 7–10. https://doi.org/10.2307/3563683.
32. Ienca, M., and P. Haselager. 2016. Hacking the brain: Brain–computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology* 18: 117–129. https://doi.org/10.1007/s10676-016-9398-9.
33. Nabavi, S., R. Fox, C. D. Proulx, J. Y. Lin, R. Y. Tsien, and R. Malinow. 2014. Engineering a memory with LTD and LTP. *Nature* 511 (7509): 348–352. https://doi.org/10.1038/nature13294.
34. Mackenzie, R. 2011. Who should hold the remote for the new me? Cognitive, affective, and behavioral side effects of DBS and authentic choices over future personalities. *AJOB Neuroscience* 2 (1): 18–20.
35. Zohny, H., D. M. Lyreskog, I. Singh, et al. 2023. The mystery of mental integrity: Clarifying its relevance to neurotechnologies. *Neuroethics* 16: 20. https://doi.org/10.1007/s12152-023-09525-2
36. Müller, O., and S. Rotter. 2017. Neurotechnology: Current developments and ethical issues. *Frontiers in Systems Neuroscience* 11: 93. https://doi.org/10.3389/fnsys.2017.00093.

37. Craig, J. N. 2016. Incarceration, direct brain intervention, and the right to mental integrity—A reply to Thomas Douglas. *Neuroethics* 9 (2): 107–118. https://doi.org/10.1007/s12152-016-9255-x.
38. Fuselli, S. 2020. Mental Integrity Protection in the Neuro-era. Legal challenges and philosophical background. *BioLaw Journal* 1: 413–429.
39. Parfit, D. 1984. *Reasons and persons*. Oxford: Clarendon.
40. Berofsky, B. 2007. *Liberation from self: A theory of personal autonomy*. Cambridge: Cambridge University Press.
41. Tubig, P., and Gilbert, F. 2023. "The Trauma of losing your own identity again": The Ethics of Explantation of Brain-Computer Interfaces. in Dubljevic, V. and Coin, A. (eds) Policy, Identity, and Neurotecnology: The Neuroethics of Brain-Computer Interfaces. Springer.
42. Chen, G., S. Padmala, Y. Chen, P. A. Taylor, R. W. Cox, and L. Pessoa. 2021. To pool or not to pool: Can we ignore cross-trial variability in FMRI? *NeuroImage* 225: 117496. https://doi.org/10.1016/j.neuroimage.2020.117496.
43. Levy, N. 2007. *Neuroethics: Challenges for the 21st century*. Cambridge: Cambridge University Press.
44. Zhang, C., C. Beste, L. Prochazkova, et al. 2022. Resting-state BOLD signal variability is associated with individual differences in metacontrol. *Science and Reports* 12: 18425. https://doi.org/10.1038/s41598-022-21703-5.
45. Bublitz, J. C. 2020. Why Means Matter: Legally Relevant Differences Between Direct and Indirect Interventions into Other Minds. In *Neurointerventions and the Law: Regulating Mental Human Capacity*, ed. N.A. Vincent, T. Nadelhoffer, and A. McCay, 49–89. Oxford: Oxford University Press.
46. Levy, N. 2020. Cognitive Enhancement: Defending the Parity Principle. In *Neurointerventions and the Law: Regulating Mental Human Capacity*, ed. N.A. Vincent, T. Nadelhoffer, and A. McCay, 33–49. Oxford: Oxford University Press.
47. Douglas, T. 2018. Neural and Environmental Modulation of Motivation: What's the Moral Difference? In David Birks & Thomas Douglas (eds.), Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice. Oxford: Oxford University Press.
48. Schmidt, A.T. 2019. Getting real on rationality—behavioral science, nudging, and public policy'. *Ethics* 129 (4): 511–543.
49. Ienca, M. 2023. On Artificial Intelligence and Manipulation. *Topoi* 42: 833–842. https://doi.org/10.1007/s11245-023-09940-3.
50. Cassinadri, G. 2024. ChatGPT and the technology-education tension: Applying contextual virtue epistemology to a cognitive artifact. *Philosophy and Technology.* 37: 14. https://doi.org/10.1007/s13347-024-00701-7.
51. Klein, E. 2016. Informed Consent in Implantable BCI Research: Identifying Risks and Exploring Meaning. *Science and Engineering Ethics* 22: 1299–1317. https://doi.org/10.1007/s11948-015-9712-7.
52. Ellegaard, M., Kragh, K. 2015. Moral Enhancement and Persistent Violent Offenders. Roskilde University. *Philosophy and Science Studies*. https://core.ac.uk/download/pdf/43031078.pdf.
53. McMillan, J. 2018. Containing Violence and Controlling Desire. In David Birks, and Thomas Douglas (eds), Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice, Engaging Philosophy (Oxford, 2018; online edn, Oxford Academic, 22 Nov. 2018), https://doi.org/10.1093/oso/9780198758617.003.001.
54. Westin, A. F. 1968. Privacy and freedom. *Wash. Lee Law Rev.* 25: 166.
55. Yuste, R., S. Goering, G. Bi, J. M. Carmena, A. Carter, J. J. Fins, et al. 2017. Four ethical priorities for neurotechnologies and AI. *Nature News* 551: 159–163. https://doi.org/10.1038/551159a.
56. Ienca, M., and G. Malgieri. 2022. Mental data protection and the GDPR. *Journal of Law and the Biosciences* 9 (1): lsac006. https://doi.org/10.1093/jlb/lsac006.
57. Ienca, M., J. J. Fins, R. J. Jox, F. Jotterand, S. Voeneky, R. Andorno, T. Ball, C. Castelluccia, R. Chavarriaga, H. Chneiweiss, A. Ferretti, O. Friedrich, S. Hurst, G. Merkel, F. Molnár-Gábor, J.-M. Rickli, J. Scheibner, E. Vayena, R. Yuste, and P. Kellmeyer. 2022. Towards a Governance Framework for Brain Data. *Neuroethics* 15 (2): 20. https://doi.org/10.1007/s12152-022-09498-8.
58. Magee, P., M. Ienca, and N. Farahany. 2024. Beyond neural data: Cognitive biometrics and mental privacy. *Neuron* 112 (18): 3017–3028. https://doi.org/10.1016/j.neuron.2024.09.004.
59. Wajnerman-Paz, A. 2021. Is Mental Privacy a Component of Personal Identity? *Frontiers in Human Neuroscience* 15. https://doi.org/10.3389/fnhum.2021.773441.
60. Baylis, F. 2012. The self in situ: a relational account of personal identity, in *Being Relational*: *Reflections on Relational Theory and Health Law*, eds J. Downie, and J. Llewellyn (Vancouver, BC: UBC Press): 109–131.
61. McCarthy-Jones, S. 2019. The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century. *Frontiers in Artificial Intelligence* 2: 19. https://doi.org/10.3389/frai.2019.00019.
62. Tang, J., A. LeBel, S. Jain, et al. 2023. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience* 26: 858–866. https://doi.org/10.1038/s41593-023-01304-9.
63. Kellmeyer, P. 2021. Big Brain Data: On the Responsible Use of Brain Data from Clinical and Consumer-Directed Neurotechnological Devices. *Neuroethics* 14 (1): 83–98. https://doi.org/10.1007/s12152-018-9371-x.
64. Pugh, J., L. Pycroft, A. Sandberg, T. Aziz, and J. Savulescu. 2018. Brainjacking in deep brain stimulation and autonomy. *Ethics of Information Technologies.* 20 (3): 219–232. https://doi.org/10.1007/s10676-018-9466-4.
65. Pugh, J., H. Maslen, and J. Savulescu. 2017. Deep Brain Stimulation, Authenticity and Value. *Cambridge Quarterly of Healthcare Ethics* 4: 640–657. https://doi.org/10.1017/S0963180117000147.
66. Farina, M., and A. Lavazza. 2022. Memory modulation *via* non-invasive brain stimulation: Status, perspectives, and ethical issues. *Frontiers in Human Neuroscience* 16: 826862. https://doi.org/10.3389/fnhum.2022.826862.
67. Chalmers, D. 2019. Extended Cognition and Extended Consciousness. In *Andy Clark and His Critics*, ed. M. Colombo, E. Irvine, and M. Stapleton, 9–21. Oxford: Oxford University Press.

68. Clark, A. 2008. Supersizing the mind. Embodiment, action, and cognitive extension. Oxford: Oxford University Press.
69. Cassinadri, G., and Fasoli, M. 2024. The extended mind thesis and the cognitive artifacts approach: a comparison. In Ienca, M. and Starke, G. (eds.) Developments in Neuroethics and Bioethics. Brains and Machines: Towards a unified Ethics of AI and Neuroscience. Elsevier. ISBN 2589–2959 https://doi.org/10.1016/bs.dnb.2024.02.004.
70. Goering, S., E. Klein, D. D. Dougherty, and A. S. Widge. 2017. Staying in the Loop: Relational Agency and Identity in Next-Generation DBS for Psychiatry. *AJOB Neuroscience* 8 (2): 59–70. https://doi.org/10.1080/21507740.2017.1320320.
71. Rupert, R. 2009. *Cognitive Systems and the Extended Mind*. Oxford: Oxford University Press.
72. Sterelny, K. 2010. Minds: Extended or Scaffolded? *Phenomenology and the Cognitive Sciences* 9: 465–481.
73. Colombo, M., E. Irvine, and M. Stapleton, eds. 2019. *Andy Clark and His Critics*. Oxford: Oxford University Press.
74. Heersmink, R. 2012. Mind and artifact: a multidimensional matrix for exploring Cognition artifact relations. In J. M. Bishop & Y. J. Erden (Eds.), Proceedings of the 5th AISB Symposium on Computing and Philosophy (54–61), Birmingham: AISB.
75. Heersmink, R. 2015. Dimensions of integration in embedded and extended cognitive systems. *Phenomenology and the Cognitive Sciences* 14: 577–598.
76. Heinrichs, J.-H. 2021. Neuroethics, cognitive technologies and the extended mind perspective. *Neuroethics* 14 (1): 59–72.
77. Cassinadri, G., and M. Fasoli. 2023. Rejecting the extended cognition moral narrative: A critique of two normative arguments for extended cognition. *Synthese* 202: 155. https://doi.org/10.1007/s11229-023-04397-8.
78. Farina, M., and A. Lavazza. 2022. Incorporation, transparency, and cognitive extension. Why the distinction between embedded or extended might be more important to ethics than to metaphysics. *Philosophy & Technology* 35: 10. https://doi.org/10.1007/s13347-022-00508-4.
79. Heersmink, R. 2017a. Distributed cognition and distributed morality: Agency, artifacts and systems. *Science and Engineering Ethics* 23 (2): 431–448.
80. Tesink, V., T. Douglas, L. Forsberg, S. Ligthart, and G. Meynen. 2024. Right to mental integrity and neurotechnologies: Implications of the extended mind thesis. *Journal of Medical Ethics*. https://doi.org/10.1136/jme-2023-109645.
81. Hernández-Orallo, J., and Vold. K. 2019. AI Extenders: The Ethical and Societal Implications of Humans Cognitively Extended by AI. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19). Association for Computing Machinery, New York: 507–513. https://doi.org/10.1145/3306618.3314238.
82. Peters. 2022. Reclaiming Control: Extended Mindreading and the Tracking of Digital Footprints. *Social Epistemology* 36 (3): 267–282. https://doi.org/10.1080/02691728.2021.2020366.
83. Heinrichs, J.-H. 2017. Against strong ethical parity: Situated cognition theses and transcranial brain stimulation. *Frontiers in Human Neuroscience* 11 (171): 1–13.
84. Cassinadri, G. 2022. Moral Reasons Not to posit Extended Cognitive Systems: A reply to Farina and Lavazza. *Philosophy and Technology*. 35: 64. https://doi.org/10.1007/s13347-022-00560-0.
85. Timms, R., and D. Spurrett. 2023. Hostile Scaffolding. *Philosophical Papers* 52 (1): 53–82. https://doi.org/10.1080/05568641.2023.2231652.
86. Heinrichs, J.H. 2024. Narrows, Detours, and Dead Ends—How Cognitive Scaffolds Can Constrain the Mind. In Heinrichs, J.-H., Beck, B., & Friedrich, O. (Eds.) (2024). Neuro-ProsthEthics: Ethical Implications of Applied Situated Cognition. J. B. Metzler Berlin, *Heidelberg*: 57-72.
87. Scangos, K. W., A. N. Khambhati, P. M. Daly, G. S. Makhoul, L. P. Sugrue, H. Zamanian, T. X. Liu, et al. 2021. Closed-loop neuromodulation in an individual with treatment-resistant depression. *Nature Medicine* 27: 1696–1700. https://doi.org/10.1038/s41591-021-01480-w.
88. Clowes, R. W. 2020. The internet extended person: exoself or doppelganger? Límite. Revista Interdisciplinaria de Filosofía y Psicología, 15 (22): 1–23. https://research.unl.pt/ws/portalfiles/portal/29762990/document_8_.pdf.
89. Clowes, R. W., P. R. Smart, and R. Heersmink. 2024. The ethics of the extended mind: Mental privacy, manipulation and agency. In *Neuro-ProsthEthics: Ethical Implications of Applied Situated Cognition*, eds. Jan-Hendrik Heinrichs, Birgit Beck, and Orsolya Friedrich. Heidelberg: J. B. Metzler Berlin.
90. Harris, K. 2021. Whose (Extended) Mind Is It, Anyway? *Erkenntis* 86: 1599–1613. https://doi.org/10.1007/s10670-019-00172-9.
91. Milojevic, M. 2020. Extended mind, functionalism and personal identity. *Synthese* 197: 2143–2170. https://doi.org/10.1007/s11229-018-1797-5.
92. Søraker, J. H. 2007. The moral status of information and information technologies: a relational theory of moral status. In S. Hongladarom, & C. Ess (Eds.), *Information technology ethics: cultural perpectives* Idea Group Publishing. https://doi.org/10.4018/978-1-59904-310-4.ch001.
93. Buller, T. 2013. Neurotechnology, Invasiveness and the Extended Mind. *Neuroethics* 6: 593–605. https://doi.org/10.1007/s12152-011-9133-5.
94. Facchin, M. 2022. Phenomenal transparency, cognitive extension, and redictive processing. *Phenomenology and Cognitive Science*. https://doi.org/10.1007/s11097-022-09831-9.
95. Heersmink, R. 2020. Varieties of the extended self. *Consciousness and Cognition* 85: 103001. https://doi.org/10.1016/j.concog.2020.103001.
96. Gilbert, F., J. N. M. Viaña, and C. Ineichen. 2021. Deflating the "DBS causes personality changes" bubble. *Neuroethics* 14 (Suppl 1): 1–17. https://doi.org/10.1007/s12152-018-9373-8.

97. Ekstrom, L. W. 1993. A coherence theory of autonomy. *Philosophy and Phenomenological Research* 53 (3): 599–616.
98. Frankfurt, H. G. 1971. Freedom of the will and the concept of a person. *The Journal of Philosophy* 68 (1): 5–20.
99. Newen, A. 2018. The embodied self, the pattern theory of self, and the predictive mind. *Frontiers in Psychology* 9: 2270. https://doi.org/10.3389/fpsyg.2018.02270.
100. Fins, J. J. 2009. Deep brain stimulation, deontology and duty: The moral obligation of non-abandonment at the neural interface. *Journal of Neural Engineering* 6: 050201.
101. Goering, S., A. I. Brown, and E. Klein. 2024. Brain Pioneers and Moral Entanglement: An Argument for Post-trial Responsibilities in Neural-Device Trials. *Hastings Center Report* 54 (1): 24–33. https://doi.org/10.1002/hast.1566.
102. Lázaro-Muñoz, G., D. Yoshor, M. S. Beauchamp, et al. 2018. Continued access to investigational brain implants. *Nature Reviews Neuroscience* 19: 317–318.
103. Sierra-Mercado, D., P. Zuk, M. S. Beauchamp, S. A. Sheth, D. Yoshor, W. K. Goodman, et al. 2019. Device removal following brain implant research. *Neuron* 103 (5): 759–761. https://doi.org/10.1016/j.neuron.2019.08.024.
104. Clark, A. 2009. Spreading the Joy? Why the Machinery of Consciousness is (Probably) Still in the Head. *Mind* 118 (472): 963–993.
105. Carter, J.A. 2021. Varieties of (extended) thought manipulation. In *Palgrave Studies in Law, Neuroscience, and Human Behavior*, ed. A. Neuroscience, M.J.B. Individual Rights, and J.C. Bublitz, 291–309. Cham: Springer International Publishing.
106. Palermos, S. O. 2023. Data, Metadata, Mental Data? Privacy and the Extended Mind. *AJOB Neuroscience* 14 (2): 84–96.
107. Heersmink, R. 2017b. Distributed selves: personal identity and extended memory systems. *Synthese* 194 (8): 3135–3151. http://www.jstor.org/stable/26748901.
108. Heersmink, R. 2022a. Extended mind and artifactual autobiographical memory. *Mind & Language.* 37: 659–673. https://doi.org/10.1111/mila.12353.
109. Heersmink, R. 2022b. Preserving Narrative Identity for Dementia Patients: Embodiment, Active Environments, and Distributed Memory. *Neuroethics* 15: 8. https://doi.org/10.1007/s12152-022-09479-x.
110. Schechtman, M. 1994. The truth about memory. *Philosophical Psychology* 7 (1): 3–18.
111. Schechtman, M. 1996. *The constitution of selves*. Ithaca, NY: Cornell University Press.
112. Schechtman, M. 2008. Diversity in unity: Practical unity and personal boundaries. *Synthese* 162 (3): 405–423.
113. Yuste, R., Genser, J., and Herrmann, S. 2021. It's Time for Neuro-Rights. *Horizons Journal International Related Sustainable Development* 154–166.
114. Noggle, R. 2022. The Ethics of Manipulation. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.). https://plato.stanford.edu/archives/sum2022/entries/ethics-manipulation/.
115. Starke, G. 2024. Hybrid Minds: Experiential and ethical implications of intelligent neural interfaces. conference talk at EACME conference 2024.
116. Clark, A. 1997. *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.
117. Carter, J. A., and O. S. Palermos. 2016. Is Having Your Computer Compromised a Personal Assault? The Ethics of Extended Cognition. *Journal of the American Philosophical Association* 2 (4): 542–560.