



Neurorights, Mental Privacy, and Mind Reading

Cohen Marcus Lionel Brown 

Received: 22 August 2023 / Accepted: 1 July 2024 / Published online: 9 July 2024
© The Author(s) 2024

Abstract A pressing worry in the ongoing neuro-rights debate is the language used to advocate for newly proposed rights. This paper addresses this concern by first examining the partial and ambiguous associations between mind reading and neurotechnology, often cited by advocates in support of the right to mental privacy. Secondly, it addresses the conceptual foundations of mind reading, distinguishing between natural, digital, and neurotechnological forms. These distinctions serve to highlight the normative parallels in privacy vulnerabilities between neurotechnology and other mind-reading methods, with an emphasis on multimodal digital systems. I suggest that authentic safeguards for the mental realm demand an expansion of the protective ambit beyond brain-targeted devices to recognize the spectrum of mind-reading applications. Ultimately, this urges re-evaluation of the scope and justification of a right to mental privacy owing to the need for coherent frameworks in an increasingly interconnected digital landscape.

Keywords Mental privacy · Neurorights · Human rights · Neurotechnology · Digital data

Introduction

New and emerging brain interventions are raising anxieties amongst scholars, policy-makers, and the wider public. One response has been to lobby for the creation of ‘neurorights’ – new human rights that promise to safeguard against the supposedly novel risks of neurodata and neurotechnology. Proponents claim that current frameworks established by the 1948 UN Declaration of Human Rights do not accommodate the profound capabilities of modern neuroscience and associated threats to fundamental human goods such as privacy, freedom, and personhood. They argue that existing legal instruments are lagging with respect to neurotechnology, and that the solution is the creation of novel rights pertaining specifically to devices and processes that target the human brain. However, the proposal has aroused controversy amongst experts at the intersection of law, ethics, and human rights. Although there is firm consensus that emerging risks necessitate safeguards, endorsement of neurorights has not been unanimous, with many neuroethicists voicing concern.

One worry pertains to the language used around newly proposed rights, which includes the ‘right to mental privacy’ [1]. For example, ‘brain reading’ is commonly accepted as shorthand to describe neurotechnological processes that image, record, or monitor the brain. But recent discussions of mental privacy have increasingly invoked the related and contentious term ‘mind reading’ to describe a range of processes

C. M. L. Brown (✉)
University of Wollongong, HSI Philosophy, Wollongong,
NSW, Australia
e-mail: cohen@uow.edu.au

that could have the capacity to reveal more intimate details than traditionally obtainable by clinical neurotechnology [2–8]. This includes mental and personal content such as thoughts, emotions, memories, social biases, group sympathies, moral dispositions, political and sexual orientation, and more. Of course, the term ‘mind reading’ is controversial: at one extreme, it conjures dystopian science-fiction, while at the other, it captures the basic inferences inherent in everyday social interaction. The former speaks to radical prospects, while the latter marks a universal human practice. While there may be space between these poles for a productive formulation, clarity is needed. One aim of this paper is therefore to address the ambiguity around talk of mind reading in the neuro-rights debate. I draw on particular accounts to show that a general understanding of mind reading is locatable in the broader neuroethics literature, and offer additional focus by distinguishing between ‘natural’, ‘digital’, and ‘neurotechnological’ kinds. These distinctions are not simply theoretical but have significance beyond pure disciplinary interests.

Such considerations bear on the case for neurorights. Specifically, claims about the risks of mind-reading neurotechnologies are often invoked to argue the need for a new, mind-specific privacy right: the right to mental privacy [2, 9–12]. This prompts questions about the concepts of mind reading and mental privacy and their strong coupling with neurotechnology in calls for new rights. For instance, how should mind reading be understood in the context of neuro-rights? Does it uniquely concern neurotechnology, as much neurorights-speak seems to imply, or is mental privacy elsewhere at risk? To clarify this space, this paper identifies relevant concerns about the partial and ambiguous associations between mind reading, mental privacy, and neurotechnology. I argue that loose conceptual notions are driving misdirection around the scope and content of privacy protections in the neurorights debate. The danger is that a mental privacy right that prioritizes neurodata or neurotechnology above other modes of mental incursion will obscure comparative privacy threats, such as those raised by multimodal digital technologies.

To be clear, the larger question is not whether privacy ought to be protected, but whether new human rights are the appropriate tools to secure the sort of protection we should want with respect to emerging technologies. As some theorists argue: “we need

new laws, not new rights” [13]. But given there are grounds for a mental privacy right, a risk is that, if formulated in neuroexceptionalist terms, it may not effectively provide the safeguards it claims to offer. If what matters for mental privacy is that minds be protected from violation, then the method by which mind-accessing occurs should be irrelevant. The imperative is that a person’s mental life be free from unwanted or unjustified incursion – not the specific form that any such incursion takes.¹ But if this is the case, then it suggests that more general privacy approaches are favourable. On the one hand, a generic neuroright might be seen as unnecessary (given overlap with existing rights), while a specific one risks being too attached to particular technologies. Second, it is not only that a neuro-specific mental privacy right would be incomplete, but that it could facilitate conditions that result in active harm. If cognitive and material resources are disproportionately allocated, neglected risks are likely to compound. Therefore, addressing these conceptual issues is important for discursive progress, but also for the practical aims of translational neuroethics in lending guidance to policy-makers, governance bodies, and broader stakeholders on the matter of neurorights.

The Neurorights Proposal

It is uncontroversial to state that the most vocal endorsement of neurorights has come from influential advocacy group the Neurorights Foundation. The group have proposed five basic neurorights: 1) The Right to Mental Privacy; 2) The Right to Personal Identity; 3) The Right to Free Will; 4) The Right to Fair Access to Mental Augmentation; 5) The Right to Protection from Bias [14]. However, this proposal is not unique, with independent arguments raised by other scholars. Ienca & Adorno [12] offer a separate, alternative set of neurorights: The Right to Cognitive Liberty, The Right to Mental Privacy, The Right to Mental Integrity, and The Right to Psychological

¹ That is assuming one has not forfeited their right to privacy by breaching certain laws, or that there are no priority interests such as justice or public safety. Additionally, it is worth recognizing that privacy is not an absolute right but one that is instrumental in facilitating moral and social goods such as freedom, security, democracy, etc.

Continuity. Since these separate bids, the general project has been backed in joint collaborations between Neurorights Foundation members and those in the wider neuroethics community. This includes a paper by an international coalition of scholars that recommends the establishment of explicit neurorights such as “mental liberty, mental privacy and mental integrity” to keep “internal mental space free of unwanted recording and manipulation” ([15], 377). While uncertainty lingers over the degree of consensus and compatibility between these separate visions, advocates generally agree that brain-based human rights reform is necessary in response to emerging neurotechnologies.

Global governance bodies and state authorities have taken keen interest. On November 9, 2021, The Council of Europe and The Organisation for Economic Co-operation and Development (OECD) held a roundtable discussion titled “Neurotechnologies and Human Rights Framework: Do We Need New Rights?” and that same year, the Chilean government advanced constitutional reforms to protect brain ‘activity’ and ‘information’ [16, 17]. In October 2022 the UN Human Rights Council adopted draft resolution A/HRC/51/L.3 on neurotechnology and human rights [18]. Then, in 2023, UNESCO published a report titled *The risks and challenges of neurotechnologies for human rights*, and on July 13, hosted an International Conference on the Ethics of Neurotechnology on the theme “Towards an Ethical Framework in the Protection and Promotion of Human Rights and Fundamental Freedoms” [19]. Recent developments at the time of writing include the explicit adoption of the Neurorights Foundation’s framework by Colorado Medical Society, the advancement of neurotechnology bills in both Colorado and California, and related amendments to the Brazilian Constitution [20–23].

Despite these advances, neurorights proposals have drawn assorted criticism from scholars. A key claim is that advocates have not adequately substantiated the need for new rights, and that the proposed articles can either be reduced to or integrated with existing protective measures [13, 24–29]. In the case of mental privacy, standard but non-exhaustive instruments include: Article 8(1) of the European Convention on Human Rights (ECHR), which includes the “right to respect for private life”; Article 9 of the ECHR, that protects a “right to freedom of thought”; Article 8 of the European Charter for Fundamental Rights

(ECFR), “the right to the protection of personal data”; Article 10 of the ECFR which includes the “right to mental integrity”; Article 18 of the Universal Declaration of Human Rights (UDHR), which states that everyone has the “rights to freedom of thought and conscience”. Similar precedents exist under ECtHR case-law against the non-consensual collection, storage, use, and sharing of biodata.² In addition to universal precepts, many domestic laws arguably make suitable provisions against mental incursion, e.g., US Fourth and Fifth Amendments; the American Convention on Human Rights; Canada’s Personal Information Protection and Electronic Documents Act (PIPEDA); Germany’s Federal Data Protection Act (BDSG); the Australian Privacy Act.³

Aside from charges of redundancy, neurorights have provoked a range of specific legal, ethical, and conceptual challenges. Some scholars have voiced apprehension about overhasty legislative change. For instance, Fins [31] identifies aspects of Chile’s recent constitutional amendments that could have severe consequences for distinct patient populations. A danger is that if negative rights are too stringent, they will disadvantage persons with cognitive-motor dissociation (CMD) who cannot consent to treatment but stand to benefit from therapeutic neurointervention [31]. However, precipitous legislative change is not the only hazard. At the other extreme, scholars argue that certain neurorights are too permissive. They worry that a ‘right to mental augmentation’ would embed controversial transhumanist goals and create coercive social conditions that might pressure individuals to adopt neurointerventions against their will [32, 33]. Corresponding fears are that neurorights would conflict with diverse beliefs and modes of being by contravening the tenets of various cultural groups [34]. Additionally, there are obvious conceptual issues attached to philosophically complex terms such as ‘free will’ and ‘personal identity’, yet efforts to recognize such challenges and offer subsequent clarification have not been forthcoming [27, 32].

² See Lighthart et al. [24] for a relevant discussion.

³ A demonstrable case is observed in national responses to Clearview AI’s identity-matching facial-recognition application. In 2021, Clearview AI was heavily penalized for breach of privacy laws in countries including the UK, France, Greece, Italy, and Australia [30].

While the above points are noncomprehensive, they demonstrate some typical challenges to neurorights.⁴ As an interdisciplinary panel of scholars recently recognized, it is now possible to distinguish three marked positions in the debate [1]. Following their lead, these views can be described as follows:

1. **Rights-Positive view:** New rights for protecting the brain and mind (i.e., neurorights) against neurotechnology are necessary.
2. **Rights-Negative view:** No new rights, legislative reforms, or translations are necessary.⁵
3. **Rights-Conservative view:** Additional legal frameworks and interpretations of existing rights may be necessary, but new brain-based human rights are not.

While I remain non-committal to any of these for present purposes, it is worth recognizing the emerging consensus amongst many neuroethicists on the latter of these views [1, 25, 26, 28, 29, 35]. That is, that while prudent measures should involve protective efforts at various levels, that does not necessitate the creation of new brain-based human rights. In any event, doubts persist over the degree of harmony between distinct neuro-rights perspectives, with a need for greater conceptual clarity over particular rights themselves. Correspondingly, this paper turns attention to a specific concern in highlighting an underappreciated point of tension with respect to the proposed right to mental privacy.

A (Neuro)Right to Mental Privacy?

Broadly, the right to mental privacy can be viewed as a proposed legislative response to common fears that neurotechnologies will soon have the capacity to peer inside minds to reveal people's innermost thoughts.⁶ This is regularly associated with

the notion of mind reading – an act which under certain conditions could plausibly be held to violate one's mental privacy (how the term 'mind reading' should be interpreted is a question that will be examined later). But beyond general intuitions, questions surface about how a right to mental privacy ought to be formalized, whether and how it might be legally practicable, and hence what specific actions might constitute rights-breaches. Although I will not attempt to resolve these concerns here, I aim to draw out some attendant complexities and their implications for the neurorights proposal.

According to the typical view, mental privacy concerns a person's capacity to control information about their mental life. This entails the idea that individuals have the right to keep their internal experiences private and free from unwanted access by external parties such as governments, corporations, or other individuals. For instance, Wajnerman-Paz [36] describes mental privacy as: "the idea that we should have control over access to our neural data and to the information about our mental processes and states that can be obtained by analyzing it" ([36], 1). Ienca [11] proposes a corresponding right to mental privacy to "explicitly protect individuals against the unconsented intrusion by third parties into their mental information (be it inferred from their neural data or from proxy data indicative of neurological, cognitive, and/or affective information) as well as against the unauthorized collection of those data" ([11], 5). Similarly, the right to mental privacy is framed as a negative right that aims to "help protect the mind from external access and inspection" and "ensure a freedom from external coercion or interference with agents' brains and minds" ([1], 7–10). Mental privacy violations can therefore be understood as undue incursion into the *locus internus*, which extends to the undisclosed or non-consensual use, collection, or distribution of information about a person's mental activity. This includes access to processes such as thoughts, desires, intentions, memories, preferences, experiences, and so on.

However, this broad understanding conflicts with certain neurorights definitions that suggest a comparatively narrow interpretation. For instance, the Neurorights Foundation frames the right to mental privacy as follows:

⁴ For further challenges see: [25–27]

⁵ It should be pointed out that few scholars or policy elites have publicly argued that absolutely no action is necessary in response to neurotechnologies. Nonetheless, this is a possible stance on the issue, and may be seen to reflect broader views such as those typified by 'transhumanist' or 'accelerationist' communities, for example.

⁶ I use 'legislative' in the broad sense of 'legislation as setting new legal norms'; however, the proposal need not be carried out specifically by legislatures and could also involve judicial, executive, independent, or international institutions.

Any data obtained from measuring neural activity (“NeuroData”) should be kept private. If stored, there should be a right to have it deleted at the subject’s request. Moreover, the sale, commercial transfer, and use of neural data should be strictly regulated [14].

Given the referential absence of subjective mental states or properties, these statements might be taken to imply that mental privacy merely covers quantitative data that directly represents the state, function, or activity of the brain. However, Ienca [11] offers a helpful distinction here:

While “mental privacy” aims at protecting mental information, however collected or inferred, neuroprivacy relates specifically to the protection of neural data—also called neurodata or brain data ([11], 7).

By these lights, it would appear that the Neurorights Foundation is committed not to mental privacy, as per the explicit title – but to neuroprivacy. *Prima facie*, an immediate issue would be that a right to mental privacy limited to the protection of neurodata (i.e., neuroprivacy) would fail to capture the normatively relevant features of mental privacy as typically received.

In reality, it is doubtful that any neurorights advocates actually believe that a mental privacy right should apply exclusively to raw neurodata to the omission of mental content. Nonetheless, based on the primary definition provided by the Neurorights Foundation, that is one potential interpretation. If, indeed, this were to be taken literally, the obvious issue is one of generalization. As others have been quick to recognize, the basic point is that brain-based information is simply another form of information that assimilates under the blanket category of personal data [25, 35, 37].⁷ On this view, unless a compelling case is made for the specialness of brain data over other sensitive data (e.g., genomic data), then strict appeals to neurodata have little force. In any case, the presence of varied definitions speaks to the inconsistent framing of the right to mental privacy in neurorights discourse. Aside from causing basic confusion, a worry is that these neuro-specific articulations could lead many, including those from non-specialist audiences, to erroneous conclusions.

⁷ See also the discussions on p. 2, 8, and 18–24 of this paper.

In practice, the correct view is that neural measurements are indeed capable of disclosing critical details about a person’s mind.⁸ This leads to the wider understanding that includes mental content per se under the right to mental privacy. For instance, Yuste & Quadra-Salcedo [10] offer the following definition:

The right to mental privacy: the content of our mental activity should not be decoded without the consent of the person subject to these new technologies. This mental privacy includes both conscious thinking and the subconscious. Most brain activity is actually subconscious; we are not even aware of its existence, yet it determines our way of life and who we are. Despite its “hidden” nature, subconscious mental activity can be deciphered in the same way, given that it is generated by neurons ([10], 23).

Nonetheless, this description prompts an independent issue: it risks potentially excluding violations of mental privacy through non-neural means. In fact, such narrow interpretations are one problem raised by the recent Colorado bill. If the assumption is that mental privacy is solely or primarily threatened by neurodata – whether gleaned from direct neurotechnologies or from peripheral sources – then there is a danger that the corresponding right will be inadequate.⁹ In other words, the possibility of non-neural threats to mental privacy minimizes the case for a special neuroright.

What this encourages is not a ‘neuroright’ after all; rather, it calls for broad privacy measures that incorporate not just neurodata but *any* personal data that might be representative of mental features. In fact, this is what many scholars have suggested. The indication is that a distinct mental privacy right is nonessential because of the provisions made by existing rights, such as the right to private life in accordance with Article 8 ECHR, the right to freedom of expression pursuant to Article 10 ECHR, the right to privacy under Article 17 ICCPR, and the right to

⁸ Further works by Neurorights Foundation members indicate that the right in question does intend to encompass mental events as typically understood, rather than being limited to neurodata as might be construed by initial definitions [2, 9].

⁹ For example, neurodata broadly construed might include data not taken directly from the brain but from the peripheral nervous system via electromyography (EMG).

freedom of thought as set out by Article 18 ICCPR [24, 35, 38–40]. A further provision is the General Data Protection Regulation (GDPR) that affords protection for the processing of personal data including biometric and ‘special category’ data about a person’s health, genetics, physiological, and mental features [1, 25]. It is additionally worth noting that, at a more granular level, domestic privacy laws are generally framed in terms of ‘personal information’ that covers all sensitive data of individuals rather than designating specific types or technologies. The key point is that laws and rights apply to the level of the *person*, and not merely at the level of the brain.

All the same, what emerges is that neurorights discourse is marked by imprecision concerning the material focus of mental privacy. It is possible that the issue rests partly on a fundamental ambiguity between general privacy and the recent notion of mental privacy; despite an extensive record of privacy in law and philosophy, there is no universally accepted definition of mental privacy [1]. Whether the latter concept is unique or sufficiently distinct from the former so as to justify independent standing is far from certain. What can be affirmed, though, is that the debate is characterized by significant conceptual disparities, leaving the scope of mental privacy largely undetermined.

This discrepancy is not obvious given the unilateral focus on neurotechnology in the discourse. One could easily be mistaken for thinking that mental privacy is a novel concern prompted only by fears of mind-reading neurotechnologies. However, it is important to emphasize that not all neurorights advocates claim that mental privacy is exclusively threatened by neurotechnology, and similarly, that not all proponents of a right to mental privacy link it specifically to neurotechnology but define it in broader terms. For instance, Ienca’s [11] view, although oriented toward neurodata, also includes “proxy data” and “mental information, however collected or inferred” ([11] 5–7). Nonetheless, the strong association between mental privacy and neurotechnology that has been promoted in the neurorights debate is a concern. Because it is unclear what mental privacy entails and whether or how it is distinct from general privacy, the range of actions that might constitute a breach of mental privacy – and the vehicles capable of actioning such breaches – remain unfixed. The reasonable corollary is that we ought to be cautious

about the predominant identification of mental privacy with neurotechnology. Accordingly, this paper supports the case for broad privacy safeguards by underscoring modes of discerning mental activity that do not depend on neurotechnology. This is what I will refer to as ‘digital mind reading’.

Digital mind reading highlights the non-select capacities of neurotechnology, showing that the concerns raised over brain-targeted devices are at-present pervasive in assorted modes. As the name suggests, there are reasons to think that mental privacy is comparably threatened by digital modes of mental access. Although such parallels have not gone wholly unnoticed, it would be a thin claim that they have been well appreciated in the neurorights conversation.¹⁰ That does not imply that mind-reading neurotechnology is a matter of little consequence, but it does call for perspective in situating specific and emerging interventions in the wider scheme of privacy concerns. It further suggests that failure to factor relevant considerations into new rights proposals risks implementing measures that convey the promise of security, but which are functionally deficient. Before getting to the concept of digital mind reading, some points of orientation are needed with respect to the term ‘mind reading’.

Definitions of Mind Reading

It is fair to assume that under certain circumstances reading a person’s mind would breach their right to mental privacy. However, the term ‘mind reading’ is employed in various ways and contexts, such that its intended meaning in any particular encounter is not obvious. This is perhaps unaided by the fact that ‘reading’ is used as technical shorthand to describe neurotechnologies that image, record, or monitor the brain. And although in most instances this usage relates specifically to objective measurements of brain activity (or neurodata), it nonetheless provokes the sticky question of whether or to what degree brain reading can be considered mind reading. The answer, of course, turns on how exactly ‘mind reading’ is understood.

¹⁰ Some notable exceptions include: Susser & Cabrera [35]; Kellmeyer [41]; Wajnerman & López-Silva [42]; Alegre [43].

In the literal sense, reading defines the process of deriving meaning through the recognition and interpretation of textual representations. Along these lines, Haselager & Mecacci [44] adopt a strong sense of mind reading which refers to the decoding of combinatorial syntax and semantics of mental states. Similarly, Meynen [3] distinguishes ‘Type II brain-based mind reading’ as “real-time recording of a subject’s thoughts” ([3], 315). Mind reading on this view demands the capacity to reliably detect mental representations, discern the precise structural relations between them, and their associated meaning. Or, as other neurorights scholars put it: “a full, granular and real-time account of the neural patterns of specific cognitive processes” ([45], 3). While there is no doubt that such direct thought apprehension would constitute a type of mind reading, there is neither any current nor prospective technology that meets such a lofty standard [4, 37, 46, 47]. It might be evocative of sci-fi tropes, but this understanding doesn’t align with the concept as customarily used in discussions about mind-reading technologies; the threshold needn’t be so high.

At the general level, reading is a bundled term used to describe common ways of perceiving, calculating, interpreting, apprehending, and evaluating. One ‘reads the room’ at a party, reads between the lines, or reads the weather, the game, and so on. Reading in this sense conveys acts of interpreting and understanding implicit or non-textual data in various contexts, and often entails the predictions made based on such information. This can encompass a broad range of processes, such as decoding complex patterns, analyzing body language, and interpreting cultural symbols. For instance, when data analysts talk about ‘reading’ a data set, they refer not to the literal act of deriving meaning from written words, but to the capacity of identifying patterns and insights from the presented information. In these everyday settings, reading is used metaphorically to capture the process of inferring meaning beyond explicit cues, or more loosely to convey a general sense of comprehension. Such idiomatic examples clearly depart from the strict definition, and while they get us some way to what is often intended by the term mind reading, they stray too far into the realm of generality.

An intermediate usage characterizes mind reading as any process that enables insights into a person’s mental life. This includes some degree of epistemic

access to features such as mental states, activity, and functioning, and may also involve deeper observations about a person’s health, capacities, and personality. By this definition, mind reading encompasses but is not limited to the following: detecting stimulus recognition, discerning psychiatric illness, predicting neurodegenerative disorders, exposing particular biases, ascertaining moral values, reconstructing visual percepts, identifying yes/no responses, establishing cognitive aptitudes, anticipating words and phrases, apprehending desires, beliefs, and intentions, revealing preferences and emotional dispositions, and more. Clearly, this definition admits a much greater range of activities, and its boundaries may be porous and imprecise. But a key factor on this understanding is that mind reading entails the disclosure of a person’s sensitive mental content or features, absent the requirement of syntactical thought decoding. This is generally consistent with other definitions of mind reading in neuroethics [5, 48, 49] and with what has variously been referred to as “brain reading” [50], “brain-based mind reading” (*Type I*) [3], and “pseudo-mind reading” [6]. Additionally, this wide understanding is typical in much of the neuroscientific literature both scholarly and mainstream, and is therefore the version I will adopt for the remainder of the paper.¹¹

Before proceeding, a caveat is in order. I employ this intermediate notion of mind reading for present purposes only, which is to illustrate the moral correspondence between different modes of mental access. However, this usage is merely instrumental to the current argument regarding the right to mental privacy. From a normative standpoint, debate is clearly unaided by the ambiguity produced by disparate understandings. The broad usage that has become pervasive in both mainstream and scholarly discourse may be seen as somewhat hyperbolic and typical of the ‘neurohype’ that characterizes public discussion, and as an associated contributor to overblown fears and expectations around neurotechnologies [51, 52]. With that stated, the distinctions between mind reading types and the challenging of weak definitions in recent accounts are advancements that will expectantly benefit from further reinforcement. Progress in this direction invites either the establishment of

¹¹ See, for example: [2, 4, 7, 8].

regular classificatory labels for mind reading types (e.g., ‘strong’ and ‘weak’), or reservation of the term for methods that meet the strict definition earlier outlined.¹² Nonetheless, I leave the issue here to turn attention to different modes of mental access.

Three Types of Mind Reading

Natural Mind Reading

There is nothing intrinsically novel about mind reading, and much of a person’s mental life is discernible in mundane ways. As Meynen [53] recognises: “mind reading is not exceptional, it is what humans normally do, all the time” ([53], 213). Behavioral observation enables us to infer others’ inner experiences, and we regularly attribute various mental states, processes, and conditions to people in our daily interactions. These obvious, non-controversial kinds of mind reading include interpreting a person’s words, intonation, and facial expressions in everyday conversation to apprehend their thoughts, feelings, beliefs, desires, and so on. These basic forms are refined and employed by trained psychologists and interrogation experts to observe micro-expressions – small, rapid facial contractions – that assist in identifying a person’s mental content. Likewise, familiar ‘cold reading’ techniques are exploited by con artists, mentalists, salespersons, and fortune tellers. The *locus internus* of others may be further glimpsed via interview routines, logical inference, ‘poker tells’, and other behavioural cues.

It is not strictly mental states that are ascribed by ordinary means; it is also possible to ascertain deeper personal qualities and characteristics. Cognitive psychology has developed impressive environmental tools that allow researchers to detect various facets of a person’s mental life including unconscious biases, preferences and tendencies, latent dispositions, and moral orientation. These include startle response and implicit association tests for racial bias and prejudice, as well as surveys and structured interviews for moral inclinations [48, 54, 55]. Such methods often reveal information that is unknown to the subject themselves

and which they cannot control and would wish to remain private. Similarly, careful observation allows medical professionals (perhaps also laypersons) to detect early signs of specific illness and mental disorders, even when subjects themselves do not recognise them. Neurodegenerative conditions such as dementia are a typical example.

Despite common efforts, ‘natural’ mind reading is far from precise. It is a fair assessment that on the whole people are unreliable lie detectors, and mind reading by verbal and nonverbal interpersonal communication carries significant and well-established limitations [56–58]. Nevertheless, this does not dismiss our ability to accurately infer others’ mental states at least some of the time. As Shen [59] emphasizes, in some regard, we are all “natural born mind readers” ([59], 656). The universal capacity to ascribe mental states to others is what cognitive scientists characterize as a ‘theory of mind’; and while there is dispute over whether it is achieved by theoretical reference or simulated modelling, there is agreement that natural mind reading is a fundamental capacity through which humans understand each other. Social interactionist theory goes further by claiming that other minds are apprehended not via ascriptive processes but via direct and immediate perception. Whatever the case, and in spite of imperfections, it is uncontested that these basic mind reading capacities are a primary feature of human social interaction that play a critical role in enabling us to predict, interpret, and respond to others. Hence, natural mind reading is already pervasive and does not raise any concerns regarding mental privacy.

Digital Mind Reading

The most acute fears around mind-reading neurotechnologies concern the prospect of having one’s private mental content exposed by covert or otherwise non-consensual means. In such dystopian scenarios, brain-scanners pluck thoughts from brains sans the cognizance or compliance of citizens. While this is an assuredly harrowing vision, such anxieties are misguided if they are fixated only on neurotechnology. That is because increasingly powerful and potentially more sinister modes of mind reading already exist in the form of digital media. In what follows, I sketch an intervening view that recognizes a type of mind reading which is neither natural nor neurotechnological. This is what I will refer to as ‘digital mind reading’.

¹² For some compelling accounts of what should constitute genuine mind reading, see: [3, 4, 47].

Digital mind reading simply describes the disclosure of information about a person's mental life via digital technology. This involves the recording, monitoring, or apprehension of user data via a digital platform and the related practices of inference, prediction, identification, and classification of mind-sensitive content. Perhaps the most immediate example is digital phenotyping, a diagnostic approach that involves passive objective monitoring of smartphone data about a person's behaviour, cognition, mood, and mental well-being to reveal details about their psychological state [60]. Examples include early detection of Parkinson's through keystroke logging [61], diagnosis of indicators for Alzheimer's disease via mobility pattern tracking [62], relapse prediction in schizophrenia using motion-sensor data [63], and objective markers of bipolar disorder through voice analysis [64]. Other forms of digital mind reading might include the following: 'emotional AI' that classifies emotions from biometrics like face scanning [65]; psychometric profiling to determine attitudes, intelligence, and personality traits [66]; internet brain-training games that provide insight about users' cognitive functions and aptitudes [67]; machine learning algorithms that detect verbal deception [56]; task-based apps that identify autism in children [68]; auto-complete functions in emails and browsers which use word-vectors to predict users' intended words and phrases; and data-harvesting practices that allow social media companies to monitor all user interactions including screen taps and text inputs about passwords, addresses, credit card numbers, and other sensitive data, such as digitized health records.¹³

There is no doubt that considerable insights about one's mental activity can be gleaned from such informational sources. Do they offer direct, unequivocal access to a person's mind? Perhaps not. But digital platforms clearly go some way in revealing the mental activity, processes, and states of individuals. If, by accessing A's calendar, contacts,

and location, I learn of A's specific plans to meet with B, I can reasonably infer that A has particular desires, beliefs, and intentions pertaining to B. Similarly, there are obvious predictions that can be made about a user's affective states and cognitive processes during certain types of gaming or excessive smartphone use (e.g., addiction, maladaptive cognitions) [71, 72]. Even milder patterns of digital device use have been found to predict states such as stress, anxiety, and loneliness [73]. This may not be the same as 'reading thoughts' per se, but it captures the right category of concern as the broad usage of mind reading.

Neurotechnological Mind Reading

Certainly neurotechnology enables unprecedented access to the human brain. But as claims about the advent of mind-reading machines proliferate, we should remind ourselves that the application of neurotechnology to human beings is nothing new.¹⁴ Recording of bioelectrical signals via electroencephalography (EEG) dates as far back as the 1920s, and modern neuroimaging techniques such as computerized axial tomography (CT or CAT scanning) and magnetic resonance imaging (MRI) have been employed for decades. Recent years have seen advancements in brain-computer interfaces (BCIs) – systems that establish a direct communication channel between the brain and peripheral devices. These use invasive and minimally-invasive means like electrocorticography (ECoG) and intracranial EEG (iEEG) respectively to encompass cognitive, sensory, and motor functions, enabling individuals to interact with external objects using neural activity. Undeniably, powerful insights are made possible through modern methods that integrate neurotechnology with machine learning techniques to facilitate large-scale data processing that allows for the increasingly dynamic mapping of neurocognitive processes. Although they provide remarkable insights about the brain, the question is to what extent these devices are capable of revealing the workings of minds.

One research area that aims to determine relevant insights is broadly referred to as brain-based

¹³ A July 2022 report by US-Australian cybersecurity firm Internet 2.0 found that popular social media application TikTok engaged in excessive data harvesting involving the unnecessary collection of user information related to contacts, calendar, location, and device operations [69]. Similarly, in August 2022, independent security and privacy researcher Felix Krause reported that social media companies can monitor all user interactions via custom in-app browsers that inject malicious script into third-party websites [70].

¹⁴ Claims typical of popular press reporting on neurotechnology can be found here: [74–76]

lie detection.¹⁵ Neuroscientific efforts have incorporated EEG and fMRI with questioning methods like the Differentiation of Deception (DoD) paradigm in attempts to establish neural correlates of deception [77, 78]. Early studies consistently identified deception-associated brain patterns and statistically significant variation between truthful and dishonest answers across a range of settings, offering provisional evidence that neurotechnology can discern when people are lying [79]. More recent analyses reveal a substantial consensus regarding the specific brain regions that are active during deceptive conduct [80]. Despite initial promise, neuroscientific lie-detection faces many well-known obstacles. Experimental conditions are highly contrived, and participants are usually given permission to lie. This raises obvious concerns about ecological validity in terms of the relevance of deceptive behaviour in laboratory settings to real-world scenarios. Second, findings are typically based on group averages rather than on single-event trials, limiting the practical ability to tell whether an individual is lying on a particular occasion. Further problems arise in relation to reverse inference, variation of deceptive behaviours, the necessity of participant compliance, and generalizability to diverse populations [56, 81, 82]. The result is that no present neurotechnology offers a genuinely viable means of identifying deception. While it is likely that multimodal testing and high-volume data analysis will improve accuracy, the degree to which barriers might be overcome is unclear. Nonetheless, brain-based lie detection demonstrates that neuroimaging offers a preliminary glimpse into the hidden mind, albeit under tightly constrained circumstances.

Neurotechnology may further reveal latent mental dispositions in the form of social biases, preferences, and group sympathies. Studies examining the neural basis of social attitudes have been able to identify types of prejudice such as outgroup negativity and ingroup favoritism from fMRI scans where participants viewed photos of black or white faces [83]. Research has also found that brain patterns reflecting in-group bias extend beyond responses to physical attributes and can indicate general social affiliation

[84]. Imaging results suggest that specific structural and functional variations in the amygdala may be predictive of broad political orientation [85, 86], while consistent findings report that activations in the anterior temporal lobe correlate with stereotyping judgments [87]. Such insights are consequential because disclosure of personal tendencies may not be isolated to singular aspects. Rather, they can be predictive of additional kinds of discriminatory attitudes, along with dispositions towards trust, empathy, aggression, and other characteristics. However, interpretive care must be exercised. Brain-level reactions to stimuli do not always register at the mental level, nor are they necessarily conclusive about a person's true sentiments. In some cases, prejudice may be 'automatic' – a fast, nonconscious neural response triggered while more complex processing continues in other brain regions, and which precedes subjective deliberation. Similar precaution is needed towards situational factors that may influence negative responses, such as participant anxiety about appearing prejudiced to others [87]. Even so, neurotechnologically mediated insights about neural mechanisms clearly have the potential to unveil highly sensitive features about one's mental activity and deeper values.

At the forefront of mind reading fears are BCI systems that decode speech and movement intentions, sometimes referred to as 'brain-to-speech' and 'thought-to-text' technologies. These use various modalities to convert electrical patterns from the brain into commands that enable the attainment of distinct objectives from external devices.¹⁶ For instance, intracortical BCIs have allowed persons with tetraplegia to type using their neural signals to control an on-screen cursor [88]. By pairing these devices with wireless point-and-click Bluetooth it is possible for paralyzed persons to utilize email, text messaging, and web-browsing in unmodified commercial computers [89]. Speech neuroprosthetics promise to further illuminate mental features. These systems decode and synthesize speech by analyzing electrophysiological recordings from the brain's language-processing regions, and typically rely on training algorithms during data-collection phases of overt, silently mouthed,

¹⁵ Although, strictly speaking, certain experiments are more appropriately termed brain-based deception detection, given they do not aim at detecting lying per se.

¹⁶ For example: fMRI (functional magnetic resonance imaging), iEEG (intracranial electroencephalography), and fNIRS (functional near-infrared spectroscopy).

and imagined speech [90–93]. Other approaches have shown it is possible to decode words and sentences directly from cerebral cortical activity in those who have lost significant articulatory-muscle function.¹⁷

While the above applications are non-exhaustive, they demonstrate some typical capabilities of modern neurotechnology. Further powers include the reconstruction of visual stimuli [95], decoding of covert intentions [96], detection of autobiographical memories [97], identification of pain states [98], prediction of psychiatric and neurological disorders [99], tracking of processes involved in moral judgement and decision-making [55], and more. Despite an impressive array of aptitudes, neurotechnologies do not meet the threshold of strong mind reading earlier outlined. No current device is able to decipher abstract thoughts at random or faithfully decode complex semantic structures of a person's spontaneous inner dialogue.¹⁸ In fact, studies suggest in-principle limitations to modern techniques that would preclude mind reading in the strictest sense. While haemodynamic response time places fundamental limits on the temporal resolution of fMRI, recent work identifies spatially-related limits on precision and reliability that constrains fMRI decoding of mental states [101, 102]. A considerable barrier is the non-uniformity of brains, which presents a challenge for cross-subject decoding. Additionally, within-subject decoding is complicated because of the temporal mutability of individual brains due to variables such as genetics, lifestyle, and environment. For instance, there is still no definitive mapping of the neural pathways, brain regions, or networks that correspond to any particular category of emotion [103]. But even if one-to-one mapping between brain states and mental states were achievable, it is possible that an 'explanatory gap' might nonetheless persist to obstruct an adequate understanding of the qualitative nature of subjective experience [104].

¹⁷ Computational and predictive-language models have allowed researchers to decode full sentences in real time at roughly 15 words per minute with approximately 75% accuracy in a person with anarthria [94]

¹⁸ A recent experiment by Tang et al. [100] has shown it is possible to generate structured semantic chains based on neural activation data paired with predictive language models. While able to produce singular words and phrases, precision of outputs is variable. But even if predictive limitations can be overcome, this does not erase the necessity of specified experimental constraints.

A further constraint is that mind-reading experiments commonly depend on substantive training phases with willing participants. These involve the collection of neural data over several hours during which subjects must actively focus on stimuli such as visual imagery or audio recordings. Machine learning algorithms are then applied to discern correlations between points that represent patterns of brain activity and the target objects in order to make predictions at a later time. This requires close engagement from subjects under carefully structured conditions, meaning that decoding is only feasible with a person's consent and cooperation. For example, a recent study demonstrated that participants were able to consciously control decoder output by resisting target stimuli through attentional volition [100]. In other words, mind-reading neurotechnologies can only simulate the mental content that a person is actively focused on. Furthermore, even basic measures such as slight head movement during scanning produces motion artifacts that corrupt fMRI results, rendering involuntary mind reading unfeasible. Such agent-directed constraints evince critical features when turning to considerations of mental privacy.

Mental Privacy Considerations

Justifications for a neurotechnology-led right to mental privacy are hazy, and given disparate claims from various authors, it is difficult to assemble a consistent position amongst them.¹⁹ Generally speaking, though, much support for the right to mental privacy converges on the following two assumptions: 1) that current privacy laws are inadequate for the protection of mental information, and 2) that from a privacy perspective, neurotechnologies possess mind-reading properties that are, to a marked degree, epistemically and therefore normatively distinct. This rationale leads some proponents to claim the necessity of a fundamentally new brain-based privacy right. But the previous mind reading accounts pose questions about hitching a right to mental privacy directly to neurotechnology – as we have seen, sensitive information about mental states is inferable by other means.

¹⁹ Further commentary on this issue can be found in ([25], 2–3).

Additionally, without a firm grounding for the right to mental privacy, it is unsettled where current privacy protections fall short and why efforts ought to be oriented towards a singular mode of mental access. Therefore, a pivotal question for respective neuro-rights proposals is whether mind-reading neurotechnologies do in fact possess ethically salient properties that might serve to justify a neuro-specific right.

It is important to state that most neurorights advocates do acknowledge some of the parallels between neurotech and digital technologies. Nonetheless, such comparisons are ultimately dismissed or downplayed in favour of neuro-specific protections. For instance, one article states that “it is important to recognize that the threats posed by neurotechnology to our fragile senses of human identity, agency, and privacy are not unique or exceptional,” ([15], 368) before proceeding to argue the case for exclusive neurorights. Accounts that distance digital sources from mental privacy commonly deploy the pessimistic stance that ‘it’s too late/difficult to regulate big data and the infosphere, so let’s just focus on neurodata’, followed by assorted positive claims for the moral differential of neurotechnology.²⁰ This typically involves broad-ranging remarks about the specialness or importance of brains, such as the following from Neurorights Foundation Chair, Rafael Yuste: “[W]e are facing a problem that affects human rights, since the brain generates the mind, which defines us as a species. At the end of the day, it is about our essence” [2]. I will, however, omit these kinds of reductivist and neuroessentialist appeals to the brain’s importance; the heart is also vital organ, and one that can be particularly revealing of mental states, but that does not mean we need a novel set of ‘cardio’ rights [106]. Instead, I turn attention to three of the more compelling claims for privileging mind-reading neurotechnology: *sensitivity*, *directness*, and *control*.

Sensitivity

Calls for a neuroright to mental privacy often rely on the putative sensitivity of information about brain activity. The claim is that neurodata is more intimate, more precise, or able to reveal a greater degree

of detail about minds than other kinds of personal information. For instance, Goering et al. [15] claim that brain data is sensitive by nature “because it contains information about the organ that generates the mind” ([15], 371). Certainly it is true that neural data is highly sensitive and may be particularly revealing, and this of course engenders critical privacy considerations. But what is not certain is that brain data possesses qualitative properties that justify distinct normative standing in comparison to other relevant sources. As illustrated by the earlier accounts, both digital sources and neurotechnologies are capable of discerning deeply telling details about a person’s mental health, emotional state, moral and social preferences, and so on. If the concern is about mind-sensitive data, then surely this should extend to *any* source that carries applicable disclosure risks regardless of locale, proximity, or function. From this perspective, I suggest that one potential point of confusion is a perceived necessity relation between neural data and mental data.

On one view, support for a neuroright to mental privacy depends on upholding a strict connection between data derived from neurotechnology and distinctive insights about the mental realm. These are not simply unique insights, but those that exceed a reputed normative boundary of mental privacy. Accordingly, there appears to exist a widespread belief amongst many neurorights scholars that a threshold-level of mental information may be singularly breached by data that are reflective of neural properties. An associated distinction is sometimes made between neural data as “information on the physiological structure and functioning of the brain”, and mental data as “raw brain data which is processed into information on a person’s mental states” ([107], 101). Beyond its descriptive utility, this distinction is often thought to mark a crucial moral difference. For instance, Bublitz [108] argues that the bridge between neural data and mental data “crosses a normatively significant divide” such that the latter is “ethically and legally more problematic” than the former ([108] 30–31). However, while it may be beneficial to discriminate between neural data and mental data, there is a problematic assumption that the latter may be derived solely from the former.

Certainly, it is reasonable to hold that not all neural data is necessarily generative of mental data. But importantly, we should also remain open to the

²⁰ See, for instance: ([1], 467); ([42], 147); ([15], 371); [105] 2871.

possibility that not all mental data need be attributable to neural data. The suggestion is that while some neural data may yield mental data, it is also possible that mental data may be derived from alternative sources. While overlooked in much of the neurorights discourse, it should be recognized that this proposition is not entirely without precedent. One argument to this effect comes from Palermos [109], who argues that if we accept the Extended Mind Thesis (EMT), the term mental data should apply to information stored in any cognitively-integrated application [110]. However, we need not adopt as controversial a view as the EMT to endorse a version of mental data untethered from brain activity.²¹ I think it is enough to say that if a tractable distinction between data and mental data (and therefore privacy and mental privacy) can reasonably be supported, then the relevant difference, in any form, is further manifest in the applicable digital sources.

Presumably, what makes something mental data (as opposed to neural data) is that it conveys important information about a person's mind. Given this is correct, then there appears to be no reasonable basis for limiting the category to a singular conveyance when such information is transmissible in a diversity of media. But if mental data is data that extends beyond mind-related content generally to exceed a presumptive sensitivity threshold, then explication is needed to specify this limit and to justify why it is uniquely overstepped by neurotechnology. A workable neuroright to this end would require specification of relevant content in such a way that entails neurodata-based mental data while excluding other forms of mental information. Where the sensitivity claim is concerned, this requires not only descriptive grounding, but a compelling case for the normative demarcation of neurodata in comparison with other viable candidates. This would, however, raise the pressing question of what to make of alternative modes of mental information, and how they might be situated amongst broader legal frameworks. In any event, neurorights accounts face the outstanding task of establishing why mental information doesn't simply reduce to information of the person, and how it fails to cohere with comparative sensitivity accommodations at the level of existing schemes.

²¹ An alternative perspective to this effect can be found here: [111].

Directness

A typical claim is that the 'direct' access to minds granted via neurotechnology constitutes a vital difference over 'indirect' modes (i.e., natural and digital). First, we might question the directness of brain-reading devices. As Ryberg [5] recognizes, significant disclosures about the mind are not feasible with neurotechnology alone:

In so far as neuroimaging of the brain can tell us anything about what is going on in a person's mind, this is surely a route mediated by various technological and theoretical steps ([5], 203).

This generally requires the integration of computational methods of data capture and analysis, often involving complex machine learning approaches such as generative adversarial networks (GANs). Thus, on a basic level, any meaningful insights arising from neurotechnology are typically mediated by intricate digital networks, amongst a multiform array of systems, programs, and appurtenances. Nevertheless, there might remain some intuitive appeal in the notion that raw neural data is more immediate, less adulterated, than other kinds of personal data. Here, the directness claim would imply that neural devices enable a one-to-one mapping between mind states and brain states, or at the very least, establish some form of objectively discernible epistemic relation. However, neural activity must be captured, processed, analyzed, and interpreted, and this is a layered translational processes that involves correlations between diverse properties across dynamic spatial and temporal contexts. As Susser & Cabrera [35] warn, this introduces a chain of unknowns that should lead us to skepticism over claims of directness. Associated issues of brain-to-mind-state inferences have been well articulated by others and need not be rehearsed here, but in very simple terms, the point is that brain reading does not (and plausibly will not ever) provide a real-time capture of wholesale mental states [6, 37, 104, 108]. This is true both now and for the foreseeable future. Neurodata might allow piecemeal inferences about mental processes, but it cannot convey live agential experience in its attendant phenomenological fullness. Therefore, claims that neurotechnology enables 'direct access to minds' are misleading.

Second, even if we accept the notion of directness, why should this mark a morally relevant feature? As

Ryberg [5] points out, merely stating that one form of access occurs by observing brain activity while the other is obtained through digitally or psychologically-mediated experiences of mental life is not a convincing means of distinguishing a morally relevant distinction. Similar orientation is found in Levy's [48] 'ethical parity principle', which serves as a heuristic to cleave the superficial features of mind-accessing interventions from those that are of genuine moral significance:

Alterations of external props are (*ceteris paribus*) ethically on a par with alterations of the brain, to the precise extent to which our reasons for finding alterations of the brain problematic are transferable to alterations of the environment in which it is embedded ([48], 61).

Covert
Collection
of Mental
Data

Although the parity principle specifies interventions that modify the brain, the point extends to brain-reading technologies. In other words, what is critical is not the particulars of the medium utilized, but the act of violation itself and its agential impact²². In a privacy context, Zúñiga-Fajuri et al. [37] state it plainly with the observation that privacy rights remain privacy rights "whether threatened by a medieval abbot spying on his monks or by twenty-first century governments operating surveillance cameras" ([37], 170). Whether an unauthorised party accesses private information 'indirectly' via one's personal diary or 'directly' via biological data matters little for the purpose of privacy. Nonetheless, the notion of directness might offer some purchase in so far as it acts as a stand-in for the degree of control a person has over mind-reading modalities. It is true that the significance of directness is often sought in the general way direct interferences might bypass or undermine control, which is relevant for autonomy rights such as privacy and freedom of thought. It therefore makes sense to address the question of control straightforwardly rather than via secondary targets.

Control

It is widely accepted that the most critical aspect of privacy is an individual's capacity to exercise control over their personal information [113]. Concerns in this

context can be understood in reference to brain-targeted modes of data collection that could potentially bypass a person's psychology to derive mental insights straight from neural activity ([107], 102); ([42] 149). The worry is that this would circumvent the filter that we ordinarily apply to our inner dialogue, rendering individuals powerless to prevent the disclosure of private thoughts, emotions, and dispositions. Mental access might hypothetically be possible by obtaining neural data covertly, say, by applying non-invasive EEG to an unconscious subject. Obviously, this would constitute not merely a breach of privacy but would entail precursory violations of autonomy and consent. Nonetheless, it does raise the theoretical prospect that, under certain conditions, some aspects of a person's mental life might be observable without their knowledge or approval. This prompts queries about the urgency and plausibility of such scenarios, coupled with consideration of whether the accompanying interventions are indeed morally exceptional.

There is no denying neurotechnology opens new and intriguing avenues of brain exploration that warrant serious privacy concern. Nevertheless, this does not imply that such concerns are inherently dissimilar or more troubling than those raised by other data-oriented technologies with which we regularly engage. For instance, digital mind reading techniques are also capable of by-passing our everyday filtering processes. Such intrusions can occur in myriad ways: through facial recognition software, 'bossware' and automated surveillance technology, social network analysis, geo-location tracking, monitoring of spending and consumption habits, data mining systems, unwitting online data sharing, failure to opt-out of end-user licence agreements, and so on. At any rate, neurotechnology is not unique in terms of the in-principle non-externalization of mind-sensitive data. Genetic information from a person's DNA is also theoretically obtainable without the consent or awareness of subjects, and this too can be indicative of a person's mental condition. For example, large-scale studies have recently found widespread positive correlations between blood-based biomarkers and certain psychiatric disorders [114]. Strictly speaking, then, evasion of agential control is possible by other means.

Bypassing
Psychological
Filters

However, it is certainly not a foregone conclusion that neurotechnology will in actuality admit access to the minds of non-consenting persons. As earlier outlined, neuroscientific mind reading requires the attentive

²² For contrast, Bublitz [112] argues that the direct vs indirect distinction is normatively relevant owing to the extent of control admitted by different classes of neurointervention, with specific focus on mind modification rather than mind reading.

participation of volunteers, and even under ideal experimental conditions, achieves a coarse-grained patchwork of predictions and probabilities about mental content. Meaningful results depend not only on raw data about brain activity, but on a range of additional data points that can include behavioural and observational data, the subjective reports of participants, and vital contextual features. Additionally, sophisticated brain-reading devices are readily conspicuous (and hardly mobile), to the effect that covert mind reading is practically unfeasible. Although sceptical views should take care to avoid the ‘delay fallacy’ when considering present limitations, we must also be wary of speculative bids of the kind that often mark bioethical discussions around new technologies. Genetic modification has become common practice in synthetic biology, for instance, but it has not led to armies of genetically engineered supersoldiers as some scholars once feared it might [115, 116]. The sensible conclusion is that mind-reading scenarios of the type earlier described belong firmly in the realm of science fiction.

Of course, control hazards need not be as explicit as the above scenario. We can envision various situations in which neurodata might be accessed without permission, such as a company collecting a greater scope of brain data than what an employee has consented to. Similarly, there are genuine concerns about data repurposing and the prospect of re-identification from anonymized data in medical, research, and commercial settings. The connectivity of big data means that disparate components can be reconfigured like fragmented machine parts to yield fresh insights about subjects. As people consent to share their data for a particular purpose, any subsequent analysis that reveals novel patterns or perspectives within a population may generate new information about individuals that is undesired, unintentional, or non-consensual. Certainly neurodata carries such risks, not only in the context of controlled patient records but particularly when considering the comparative latitude of the direct-to-consumer commercial market. But again, these privacy dangers are not unique to this space, as they generalize to all forms of personal data. The ‘4R challenge’ as it is sometimes known, is a prominent and longstanding issue in data ethics.²³ An illustrative

example is the AOL search log release in 2006, where the internet provider published an anonymized dataset that was re-identifiable using small amounts of auxiliary data, allowing specific users to be traced by their search queries [118]. So, while neurodata may be vulnerable to subtle and sophisticated forms of privacy breach, the problem is one of a general nature. It is therefore difficult to understand why data from one particular medium should warrant unconventionally special treatment to other highly sensitive personal data from the financial, health care, and commercial domains. In this sense, there is nothing to suggest that the control elements of neurotechnology are principally atypical or distinct from comparable systems.

Conclusion

Articulations of mind reading in arguments for new rights matter. If calls for a mental privacy right presuppose a narrow construction that is partially aligned with neurotechnology then a risk is the overshadowing of comparative privacy threats from alternative mind-reading methods. Certain neurorights proposals take it as a priori that mind-reading neurotechnology presents a fundamentally new privacy issue. But any argument for rights-oriented amendments based on a novel ethical problem ought to establish what that problem is, and provide firm clarification of the conceptual and normative dimensions that prompt revision with respect to associated gaps in existing frameworks. In any case, these are critical details that cannot simply be assumed on the basis of neuroexceptionalism. This presses upon neurorights advocates to demonstrate why brain-based devices for predicting and inferring mental states are normatively discontinuous with existing techniques, and why such distinctions present a unique risk for the sake of mental privacy. It seems, then, that the justificatory basis for a neuro-specific right to mental privacy must meet two challenges: 1) to establish that there is a legally tractable difference between general privacy and mental privacy, and, pending a convincing case, then 2) to demonstrate that attendant conditions apply to neurotechnology, to the exclusion of relevant alternatives. Correspondingly, a taxonomy of mind reading that recognises ‘natural’, ‘digital’, and ‘neurotechnological’ kinds encourages caution around privileged associations between mental privacy and neurotechnology. This paper has reinforced the perspective that the

Commercial Exploitation

²³ The 4R challenge refers to ethical issues stemming from data reuse, repurposing, recombination, and reanalysis [117].

26. Hertz, N. 2022. Neurorights – Do we Need New Human Rights? A Reconsideration of the Right to Freedom of Thought. *Neuroethics* 16 (1): 5.
27. Rainey, S. 2023. Neurorights as Hohfeldian Privileges. *Neuroethics* 16 (1): 9.
28. Alegre, S. 2023. We don't need new 'neurorights' — we need to know the existing law. In *Financial Times* Nikkei Inc.
29. Borbón, D. and J.A. Ramírez-Gómez, 2024. Between politics and scholarship: the (un)settled debate over neurorights. *Frontiers in Political Science* 6: 1–5.
30. ICO, 2022. *ICO fines facial recognition database company Clearview AI Inc more than £7.5m and orders UK data to be deleted* 2022: <https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/05/ico-fines-facial-recognition-database-company-clearview-ai-inc/>
31. Fins, J.J. 2022. The Unintended Consequences of Chile's Neurorights Constitutional Reform: Moving beyond Negative Rights to Capabilities. *Neuroethics* 15 (3): 26.
32. Borbón, D. and L. Borbón, 2021. A critical perspective on NeuroRights: comments regarding ethics and law. *Frontiers in Human Neuroscience* 15: 703121.
33. Borbón Rodríguez, D.A., L.F. Borbón Rodríguez, and M.A. León Bustamante. 2021. *Neuroderecho al acceso equitativo a tecnologías de mejora: Análisis desde el posthumanismo, el derecho y la bioética*. Revista Iberoamericana de Bioética 16: 1–15.
34. Herrera-Ferrá, K., et al. 2022. *Contextual and Cultural Perspectives on Neurorights: Reflections Toward an International Consensus*. *AJOB Neuroscience*, 1–9.
35. Susser, D., and L.Y. Cabrera. 2023. Brain Data in Context: Are New Rights the Way to Mental and Brain Privacy? *AJOB Neuroscience*, 1–12.
36. Wajnerman Paz, A. 2021. Is mental privacy a component of personal identity? *Frontiers in Human Neuroscience* 15: 773441.
37. Zúñiga-Fajuri, A., et al. 2021. *Chapter Seven - Neurorights in Chile: Between neuroscience and legal science*. In *Developments in Neuroethics and Bioethics*, ed. M. Hevia, 165–179. Academic Press.
38. Ligthart, S. 2024. Mental privacy as part of the human right to freedom of thought? In *The law and ethics of freedom of thought : Cognitive liberty and privacy*, ed. J.C.B.M. Blitz. Palgrave Macmillan.
39. Bublitz, C. 2021. Freedom of Thought as an International Human Right. In *The Law and Ethics of Freedom of Thought, Volume I: Neuroscience, Autonomy, and Individual Rights*, ed. C.M.B. Blitz, 49–101. Palgrave Macmillan.
40. Alegre, S. 2022. *Freedom to Think*. London: Atlantic Books.
41. Kellmeyer, P. 2022. 'Neurorights': A Human Rights-Based Approach for Governing Neurotechnologies. In *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives*, ed. O. Mueller, et al., 412–426. Cambridge: Cambridge University Press.
42. Wajnerman, A., and P. López-Silva. 2022. Mental Privacy and Neuroprotection: An Open Debate. In *Protecting the Mind: Challenges in Law, Neuroprotection, and Neurorights*, ed. P. López-Silva and L. Valera, 141–155. Cham: Springer International Publishing.
43. Alegre, S. 2022. Toward an International Agenda to Protect the Forum Internum. CIGI Policy Brief No. 176. <https://www.cigionline.org/publications/toward-an-international-agenda-to-protect-the-forum-internum>. Accessed 6 Feb 2024.
44. Haselager, P., and G. Mecacci. 2018. Is Brain Reading Mind Reading? In *NeuroLaw and Responsibility for Action: Concepts, Crimes, and Courts*, ed. B. Donnelly-Lazarov, 182–192. Cambridge: Cambridge University Press.
45. Ienca, M., et al. 2022. Towards a Governance Framework for Brain Data. *Neuroethics* 15 (2): 20.
46. MacDuffie, K.E., and S. Goering. 2019. Neurotechnologies Cannot Seize Thoughts: A Call for Caution in Nomenclature. *AJOB Neuroscience* 10 (1): 23–25.
47. Mecacci, G., and P. Haselager. 2019. Identifying Criteria for the Evaluation of the Implications of Brain Reading for Mental Privacy. *Science and Engineering Ethics* 25 (2): 443–461.
48. Levy, N. 2007. *Neuroethics : Challenges for the 21st Century*. Cambridge, UNITED KINGDOM: Cambridge University Press.
49. Lippert-rasmussen, K. 2017. Brain Privacy, Intimacy, and Authenticity: Why a Complete Lack of the Former Might Undermine Neither of the Latter! *Res Publica* 23 (2): 227–244.
50. Haynes, J. D., 2012. Brain reading. In *I Know What You're Thinking: Brain imaging and mental privacy*, ed. Sarah D. Richmond, Geraint Rees, and Sarah J. L. Edwards, 29–40. UK: Oxford University Press.
51. Wexler, A. 2019. Separating neuroethics from neurohype. *Nature Biotechnology* 37 (9): 988–990.
52. Morse, S.J. 2019. 5 - Neurohype and the law: A cautionary tale. In *Casting Light on the Dark Side of Brain Imaging*, ed. A. Raz and R.T. Thibault, 31–35. Academic Press.
53. Meynen, G. 2018. Author's Response to Peer Commentaries: Brain-based mind reading: Conceptual clarifications and legal applications. *Journal of Law and the Biosciences* 5 (1): 212–216.
54. Greene, J.D. 2014. Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics. *Ethics* 124 (4): 695–726.
55. Greene, J.D. 2014. The cognitive neuroscience of moral judgment and decision making. In *The cognitive neurosciences*, 5th ed., 1013–1023. Cambridge, MA, US: Boston Review.
56. Brennen, T., and S. Magnussen. 2022. The Science of Lie Detection by Verbal Cues: What Are the Prospects for Its Practical Applicability? *Frontiers in Psychology* 13.
57. Vrij, A., M. Hartwig, and P.A. Granhag. 2019. Reading Lies: Nonverbal Communication and Deception. *Annual Review of Psychology* 70 (1): 295–317.
58. Bond, C.F., Jr., and B. DePaulo. 2008. Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin* 134 (4): 477–492.
59. Shen, F.X. 2013. Neuroscience, Mental Privacy, and the Law Privacy, Security, and Human Dignity in the Digital Age. *Harvard Journal of Law & Public Policy* 36 (2): 653–714.
60. Insel, T.R. 2017. Digital Phenotyping: Technology for a New Science of Behavior. *JAMA* 318 (13): 1215–1216.

61. Giancardo, L., et al. 2015. Psychomotor Impairment Detection via Finger Interactions with a Computer Keyboard During Natural Typing. *Scientific Reports* 5 (1): 9678.
62. Nieto-Reyes, A., et al. 2017. Classification of Alzheimer's Patients through Ubiquitous Computing. *Sensors* 17 (7): 1679.
63. Kalisperakis, E., et al. 2023. Smartwatch digital phenotypes predict positive and negative symptom variation in a longitudinal monitoring study of patients with psychotic disorders. *Frontiers in Psychiatry* 14.
64. Faurholt-Jepsen, M., et al. 2016. Voice analysis as an objective state marker in bipolar disorder. *Translational Psychiatry* 6 (7): e856–e856.
65. McStay, A. 2020. Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. *Big Data & Society* 7 (1): 2053951720904386.
66. Rust, J., M. Kosinski, and D. Stillwell. 2020. *Modern Psychometrics : The Science of Psychological Assessment*. Milton, UNITED KINGDOM: Taylor & Francis Group.
67. Purcell, Ryan H., and Karen S. Rommelfanger. 2015. Internet-Based Brain Training Games, Citizen Scientists, and Big Data: Ethical Issues in Unprecedented Virtual Territories. *Neuron* 86 (2): 356–359.
68. Chang, Z., et al. 2021. Computational Methods to Measure Patterns of Gaze in Toddlers With Autism Spectrum Disorder. *JAMA Pediatrics* 175 (8): 827–836.
69. Perkins, T. 2022. Internet 2.0 TikTok Analysis. <https://internet2-0.com/tiktok/>. Accessed 21 Jan 2024.
70. Krause, F. 2022. iOS Privacy: Instagram and Facebook can track anything you do on any website in their in-app browser. KrauseFX. <https://krausefx.com/blog/ios-privacy-instagram-and-facebook-can-track-anything-you-do-on-any-website-in-their-in-app-browser>. Accessed 21 Jan 2024.
71. Ryding, F.C., and D.J. Kuss. 2020. Passive objective measures in the assessment of problematic smartphone use: A systematic review. *Addictive Behaviors Reports* 11: 100257.
72. Fioravanti, G., et al. 2020. How maladaptive cognitions contribute to the development of problematic social media use. *Addictive Behaviors Reports* 11: 100267.
73. Cornet, V.P., and R.J. Holden. 2018. Systematic review of smartphone-based passive sensing for health and well-being. *Journal of Biomedical Informatics* 77: 120–132.
74. Reardon, S. 2023. Mind-reading machines are here: is it time to worry? *Nature* 17: 236
75. Samuel, S. 2023. *Mind-reading technology has arrived*. 2023; Available from: <https://www.vox.com/future-perfect/2023/5/4/23708162/neurotechnology-mind-reading-brain-neuralink-brain-computer-interface>.
76. Futurism. 2016. *Scientists Have Invented a Mind-Reading Machine That Visualises Your Thoughts*. 2016; Available from: <https://futurism.com/scientists-have-invented-a-mind-reading-machine-that-visualises-your-thoughts>.
77. Hu, X., H. Wu, and G. Fu. 2011. Temporal course of executive control when lying about self and other-referential information: An ERP study. *Brain research* 1369: 149–157.
78. Meijer, E.H., et al. 2016. Deception detection with behavioral, autonomic, and neural measures: Conceptual and methodological considerations that warrant modesty. *Psychophysiology* 53 (5): 593–604.
79. Sip, K.E., et al. 2008. Detecting deception: The scope and limits. *Trends in Cognitive Sciences* 12 (2): 48–53.
80. Farah, M.J., et al. 2014. Functional MRI-based lie detection: Scientific and societal challenges. *Nature Reviews Neuroscience* 15 (2): 123–131.
81. Jones, O.D., and A.D. Wagner. 2020. Law and Neuroscience: Progress, Promise, and Pitfalls. In *The Cognitive Neurosciences*. The MIT Press.
82. Seaman, J.A. 2018. Your brain on lies: Deception detection in court. In *The Routledge handbook of neuroethics*, 198–213. New York, NY, US: Routledge/Taylor & Francis Group.
83. Molenberghs, P. and W.R. Louis. 2018. Insights from fMRI studies into ingroup bias. *Frontiers in Psychology* 9: 1868.
84. Cikara, M., et al. 2017. Decoding “us” and “them”: Neural representations of generalized group concepts. *Journal of Experimental Psychology: General* 146 (5): 621–631.
85. Kanai, R., et al. 2011. Political Orientations Are Correlated with Brain Structure in Young Adults. *Current Biology* 21 (8): 677–680.
86. Schreiber, D., et al. 2013. Red Brain, Blue Brain: Evaluative Processes Differ in Democrats and Republicans. *PLoS ONE* 8 (2): e52970.
87. Amodio, D.M., and M. Cikara. 2021. The Social Neuroscience of Prejudice. *Annual Review of Psychology* 72 (1): 439–469.
88. Pandarinath, C., et al. 2017. High performance communication by people with paralysis using an intracortical brain-computer interface. *eLife* 6: e18554.
89. Nuyujukian, P., et al. 2018. Cortical control of a tablet computer by people with paralysis. *PLoS ONE* 13 (11): e0204566.
90. Soroush, P.Z., et al. 2023. The nested hierarchy of overt, mouthed, and imagined speech activity evident in intracranial recordings. *NeuroImage* 269: 119913.
91. Proix, T., et al. 2022. Imagined speech can be decoded from low- and cross-frequency intracranial EEG features. *Nature Communications* 13 (1): 48.
92. Moses, D.A., et al. 2019. Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nature Communications* 10 (1): 3096.
93. Anumanchipalli, G.K., J. Chartier, and E.F. Chang. 2019. Speech synthesis from neural decoding of spoken sentences. *Nature* 568 (7753): 493–498.
94. Moses, D.A., et al. 2021. Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria. *New England Journal of Medicine* 385 (3): 217–227.
95. Kamitani, Y., and F. Tong. 2005. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience* 8 (5): 679–685.
96. Haynes, J.-D., et al. 2007. Reading Hidden Intentions in the Human Brain. *Current Biology* 17 (4): 323–328.
97. Rissman, J., et al. 2016. Decoding fMRI Signatures of Real-world Autobiographical Memory Retrieval. *Journal of Cognitive Neuroscience* 28 (4): 604–620.
98. van der Miesen, M.M., M.A. Lindquist, and T.D. Wager. 2019. Neuroimaging-based biomarkers for pain: state of the field and current directions. *Pain Reports* 4 (4): e751.
99. Mateos-Pérez, J.M., et al. 2008. Structural neuroimaging as clinical predictor: A review of machine learning applications. *NeuroImage: Clinical* 20: 506–522.

100. Tang, J., et al. 2023. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience* 26 (5): 858–866.
101. Jabakhanji, R., et al. 2022. Limits of decoding mental states with fMRI. *Cortex* 149: 101–122.
102. Elliott, M.L., et al. 2020. What Is the Test-Retest Reliability of Common Task-Functional MRI Measures? New Empirical Evidence and a Meta-Analysis. *Psychological Science* 31 (7): 792–806.
103. Barrett, L.F., and A.B. Satpute. 2019. Historical pitfalls and new directions in the neuroscience of emotion. *Neuroscience Letters* 693: 9–18.
104. Levine, J. 1983. MATERIALISM AND QUALIA: THE EXPLANATORY GAP. *Pacific Philosophical Quarterly* 64 (4): 354–361.
105. Yuste, R. 2023. Advocating for neurodata privacy and neurotechnology regulation. *Nature Protocols* 18 (10): 2869–2875.
106. Zang, X., et al. 2022. End-to-End Depression Recognition Based on a One-Dimensional Convolution Neural Network Model Using Two-Lead ECG Signal. *Journal of Medical and Biological Engineering* 42 (2): 225–233.
107. Istace, T. 2022. Neurorights: The Debate about New Legal Safeguards to Protect the Mind. *Issues in Law & Medicine* 37 (1): 95–114.
108. Bublitz, J.C. 2019. Privacy Concerns in Brain-Computer Interfaces. *AJOB Neuroscience* 10 (1): 30–32.
109. Palermos, S.O. 2023. Data, Metadata, Mental Data? Privacy and the Extended Mind. *AJOB Neuroscience* 14 (2): 84–96.
110. Clark, A., and D. Chalmers. 1998. The Extended Mind. *Analysis* 58 (1): 7–19.
111. Marcus Lionel Brown, C. 2023. C Extended Mind Over Matter: Privacy Protection Is the Sine Qua Non. *AJOB Neuroscience* 14 (2): 97–99.
112. Bublitz, J.C. 2020. Why means matter: legally relevant differences between direct and indirect interventions into other minds. In *Neurointerventions and the law: Regulating human mental capacity*, ed. N.A. Vincent, T. Nadelhoffer, and A. McCay, 49–88. UK: Oxford University Press.
113. Westin, A.F. 1968. Privacy and Freedom. *Washington & Lee Law Review* 25 (1): 166–170.
114. Reay, W.R., et al. 2022. Genetic estimates of correlation and causality between blood-based biomarkers and psychiatric disorders. *Science Advances* 8 (14): eabj896.
115. Goodman, K.W. 2004. WONDERGENES: GENETIC ENHANCEMENT AND THE FUTURE OF SOCIETY. *Journal of Legal Medicine* 25 (2): 257–265.
116. Burkhardt, M.T.A. 2007. Operation Rebirth: Captain America and the Ethics of Enhancement. *Philosophy Now* 64: 8–10.
117. Steinmann, M., S.A. Matei, and J. Collmann. 2016. A Theoretical Framework for Ethical Reflection in Big Data Research. In *Ethical Reasoning in Big Data: An Exploratory Analysis*, ed. J. Collmann and S.A. Matei, 11–27. Cham: Springer International Publishing.
118. Porter, C. C. 2008. De-identified data and third party data mining: the risk of re-identification of personal information. *Washington Journal of Law, Technology & Arts* 5 (1): 1–9.

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