



# Is Your Neural Data Part of Your Mind? Exploring the Conceptual Basis of Mental Privacy

Abel Wajnerman Paz<sup>1</sup>

Received: 27 February 2021 / Accepted: 14 September 2021 / Published online: 23 September 2021  
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

## Abstract

It has been argued that neural data (ND) are an especially sensitive kind of personal information that could be used to undermine the control we should have over access to our mental states (i.e. our mental privacy), and therefore need a stronger legal protection than other kinds of personal data. The Morningside Group, a global consortium of interdisciplinary experts advocating for the ethical use of neurotechnology, suggests achieving this by treating legally ND as a body organ (i.e. protecting them through bodily integrity). Although the proposal is currently shaping ND-related policies (most notably, a Neuroprotection Bill of Law being discussed by the Chilean Senate), it is not clear what its conceptual and legal basis is. Treating legally something as something else requires some kind of analogical reasoning, which is not provided by the authors of the proposal. In this paper, I will try to fill this gap by addressing ontological issues related to neurocognitive processes. The substantial differences between ND and body organs or organic tissue cast doubt on the idea that the former should be covered by bodily integrity. Crucially, ND are not constituted by organic material. Nevertheless, I argue that the ND of a subject *s* are analogous to neurocognitive properties of her brain. I claim that (i) *s'* ND are a ‘medium independent’ property that can be characterized as natural semantic personal information about her brain and that (ii) *s'* brain not only instantiates this property but also has an exclusive ontological relationship with it: This information constitutes a domain that is unique to her neurocognitive architecture.

**Keywords** Mental privacy · Neurorights · Neural data · Semantic information · Cognitive domain

---

✉ Abel Wajnerman Paz  
abelwajnerman@gmail.com; awajnerman@uahurtado.cl

<sup>1</sup> Universidad Alberto Hurtado, Almirante Barroso 10, 8340575 Santiago, Chile

## 1 Introduction

During the last decade, government agencies around the world led a neurotechnological revolution. Spurred by the creation of the U.S. BRAIN Initiative in 2013, which funded public research in neurotechnology and artificial intelligence, China, Korea, the European Union, Japan, Canada and Australia, among others, have created similar research projects. These initiatives contributed to the development of neurotechnologies and techniques that have an unprecedented ability (both in terms of scope and reliability) to “read” mental states, in the sense of decoding information about mental states or processes by analyzing data about neural activity patterns, and “transcribe” mental states by modulating neural computation. Crucially, these advancements have sparked the rapid development of non-invasive and potentially ubiquitous consumer neurotechnologies having various non-clinical (educational, entertainment-related, work-related and military) applications, which are not fully explored or regulated by either national laws or international treaties (Fernandez et al. 2015). Thus, the development of a regulatory framework has become a global priority.

During 2020 we witnessed in Chile the emergence of a pioneering regulatory framework that seeks to regulate neurotechnology development and applications. The proposal of the “Future Challenges, Science, Technology and Innovation” Commission of the Chilean Senate consists of a constitutional reform bill (Bulletin 13.827-19) and a bill on neuro-protection (Bulletin 13.828-19). These two bills were drafted during 2020 by the Senate’s Committee and are being discussed by the Senate and the Chamber of Deputies of Chile during 2021. In turn, Article 24 of the Spanish Charter of Digital Rights, recently announced by the Secretary of State for Digitalization and Artificial Intelligence of the Government of Spain, represents a second pioneering effort in establishing specific rights for the regulation of neurotechnology, also known as “*neurorights*”. Both proposals, inspired by the framework developed by the *Morningside Group*, an interdisciplinary group led by the neuroscientist Rafael Yuste, safeguard five key neurorights: the right to personal identity, the right to free will, the right to mental privacy, the right to equal access to cognitive enhancement technologies, and the right to protection against algorithmic bias. These rights build on, expand and/or specify existing international human rights for the protection of human dignity, liberty and security of persons, non-discrimination, equal protection, and privacy. The underlying idea is that, in their previous versions, these rights address certain ethically relevant dimensions of human life in very generic terms, often subject to interpretation, and regulating the ramifications of neurotechnology requires greater specificity (Yuste et al. 2017, 2021).

In this paper, I will discuss the conceptual basis of the right to mental privacy. This right protects the control over access to our neural data (ND) and to the information about our mental processes and states that can be obtained by analyzing it. The *Morningside Group’s proposal* includes an innovative interpretation of this right. Rather than describing it in broad-brush terms, it proposes to treat ND as a special kind of information that is intimately related to who we are and

that partly defines our identity. This would be accomplished by legally considering ND as organic tissue and therefore applying to them the laws for organ transplantation and donation. This entails that people not only have a right to not be compelled to give up ND but, crucially, ND collection requires explicit "opt-in" authorization. Additionally, ND cannot be commercially transferred and used but only donated for altruistic purposes. That is, ND commercialization is prohibited regardless of consent status (Goering & Yuste, 2016; Goering et al., 2021; Yuste et al., 2017). As the first country with a Neuroprotection Bill that has taken up this proposal, Chile has become a pilot case.<sup>1</sup>

However, to the best of my knowledge, this proposal has not yet been subjected to philosophical discussion. Instead of discussing how ND should be regulated, a substantial part of the recent debate on mental privacy has been concerned with whether we need new ND-focused regulations at all. It has been suggested current technological mind reading has substantial limitations and therefore it poses no real threat to mental privacy. These applications can often only decode a very limited set of predetermined mental states from neural activity, lacking unlimited real time access to just any content of the mind (e.g. see Meynen 2019 and its Open Peer Commentaries). However, the refusal to legislate on the basis of current technological limitations is arguably an instance of the so-called "delay fallacy" (Mecacci & Haselager, 2019). If we wait for this technology to be fully developed before deciding how to regulate it, by the time it is already developed, the technical features and social practices associated with it may become too culturally entrenched to be easily modified.

A prime example of the failure to anticipate the consequences of technological innovation is given by the rise of powerful algorithms that obtain sensitive information about our psychological traits and states by analyzing our digital footprint (e.g. Facebook likes, posts, photos, etc.). By the time the risks of these techniques (such as "psychological targeting") were brought to public attention by the Facebook–Cambridge Analytica data scandal, they had already blurred the line between private and public information, to the point that users have no control over what mental information about them can be digitally gathered (Matz et al., 2020). This may force governments to regulate only how this information is applied (e.g. preventing unethical manipulations of behavior), rather than how or whether it is collected.

This is why anticipation is necessary in thinking about how ND should be regulated, and specifically in discussing the Morningside Group's proposal, which is based on conceptual assumptions that have concrete practical implications (e.g. the prohibition or restriction of ND commercialization), but may not be appropriately grounded. From a legal standpoint, treating something as something else requires some kind of analogical reasoning. In legal reasoning, analogy involves an earlier decision being followed in a later case provided both cases are sufficiently similar (Lamond, 2006). Thus, in order to do the job, the organic approach must be based

<sup>1</sup> <https://nri.ntc.columbia.edu/news/chile-neurorights-international-lab-originally-spanish>, see also <https://nri.ntc.columbia.edu/projects>

on an analogy showing similarities between ND and body organs or organic tissue. Nevertheless, an argument supporting an analogy this kind has not been provided. In this paper, I will ground the conclusion that ND deserves the special protection provided by bodily integrity by using a different analogy.

After presenting different views on ND protection (Sect. 2) I describe substantial disanalogies between ND and body organs, which cast doubt on the possibility of protecting the former through bodily integrity (Sect. 3). Crucially, ND are not constituted by organic material.

Nevertheless, in Sect. 4 I argue that the ND of a subject *s* are analogous to neurocognitive properties of her brain. I claim that *s*' ND are a ‘medium independent’ property (Sect. 4.1) that can be characterized as natural semantic personal information about her brain (Sect. 4.2) and that *s*' brain not only instantiates this property but also has an exclusive ontological relation with it: This information constitutes a domain that is unique to its cognitive architecture (Sects. 4.3, 4.4, 4.5, 4.6).

## 2 Mental Privacy and Privacy Dimensions

The basic principles related to research with human beings (i.e. autonomy, integrity, beneficence and justice) ground different privacy interests or dimensions (Salles et al., 2017). Following Laurie et al. (2010), privacy can be analyzed in terms of physical privacy, information privacy, decisional privacy and proprietary privacy. In turn, mental privacy (the idea that we should have control over informational access to our mental/neural states) is often not presented as a fifth privacy dimension but rather as part of one of the other four. In what follows I will briefly characterize the dimensions that are relevant for the present debate.

A possible approach to ND protection is to say that, given that they satisfy the definition of ‘personal information’ (see Sect. 4.2), they are covered by information privacy. Information privacy is the idea that we should be protected against the different kinds of damage (e.g. discrimination) that can result from the diffusion of our personal data and is therefore based on the principles of beneficence and justice (Salles et al., 2017). This entails, for instance, the obligation to anonymize the data employed in research contexts. According to this view, ND is already protected by the regulations that are applied to other kinds of personal information (such as the Fourth Amendment to the United States Constitution). Thus, if one has a “reasonable expectation of privacy” regarding the identifying information derived from one’s blood or saliva samples, one also has a reasonable expectation of privacy regarding the data decoded from one’s own brain (Shen, 2013).

By contrast, members of the Morningside Group suggest that, given the biological nature of the signals that carry ND, they should be protected by *physical privacy* (Goering & Yuste, 2016; Goering et al., 2021; Yuste et al., 2017). This privacy dimension is related to the access to our organic samples (the fact that these cannot be gathered and stored without consent) and is therefore grounded in bodily integrity. As I mentioned, Chile is on its way to implement this idea. On October 7th, 2020 the “Future Challenges, Science, Technology and Innovation” Commission of the Chilean Senate has introduced a *Constitutional Reform Bill (Bulletin 13.827-19)*

and the **Neuroprotection Bill of Law** (Bulletin 13.828-19). Both bills were approved by the Committee on October 30<sup>th</sup> and then approved in general by the Chilean Senate<sup>2</sup> on December 16th, 2020. The specific details of both bills are being actively discussed (the final version of the Constitutional Reform was already approved by the Senate on April 21st 2021 and is being addressed by the Chamber of Deputies). Following the original proposal advanced by the Morningside Group, the Neuroprotection Bill (explicitly or implicitly, and directly or indirectly) safeguards the five ‘neurorights’: The Right to Personal Identity (Article 4), The Right to Free-Will (Articles 3 and 4), The Right to Mental Privacy (Articles 6 and 7), The Right to Equal Access to Mental Augmentation (Article 10) and The Right to Protection from Algorithmic Bias (Article 9). Regarding mental privacy, the Bill establishes that neural data is a *special category* of sensitive health-data. Specifically, Article 7 states that:

the collection, storage, treatment, and dissemination of neuronal data and the neuronal activity of individuals will comply with the provisions contained in Law No. 19.451 regarding transplantation and organ donation, as applicable, and the provisions of the respective health code.<sup>3</sup>

As other similar legislations regarding organ transplant (such as the 1984 US National Organ Transplant Act), this determines that we cannot consent to ND commercialization but only to its donation for altruistic purposes.

In the following section, I will emphasize relevant disanalogies between ND and organic material. However, I also reject the idea that mental privacy is simply a case of information privacy. In Sect. 4, I will claim that mental privacy is part of *psychological integrity*. We can define psychological integrity as the idea that no one can alter or manipulate the mind of an individual (e.g. modulate her neural computation or information through electrical or magnetic brain stimulation) without her consent (e.g. Ienca & Haselager, 2016; Lavazza, 2018). I will defend a version of the idea that access to information about our mind can sometimes be identical with *access to the mind itself* or, more specifically, to its own informational properties (what Ienca & Andorno, 2017 call ‘the inception problem’). If this is so, then mental privacy is part of psychological integrity, it is related to the control over the different aspects (including, but not limited to, the informational properties) that constitute our minds.

### **3 The Disembodied Nature of Neural Data**

The idea of protecting ND through bodily integrity is plausibly undermined by substantial disanalogies between ND and the kind of objects covered by this right (i.e. body organs or components and organic tissue). The general disanalogies between

<sup>2</sup> <https://nri.ntc.columbia.edu/news/unanimously-chilean-senate-approves-regular-neurorights>

<sup>3</sup> This is part of the original text, the bulletin in the Chilean senate’s website ([https://www.senado.cl/appsenado/templates/tramitacion/index.php?boletin\\_ini=13826-13](https://www.senado.cl/appsenado/templates/tramitacion/index.php?boletin_ini=13826-13)) includes an English version.

health data extraction and body organ harvesting have been addressed. A key observation is that our bodies are not the only source of health-data. As Montgomery (2017) has pointed out, there are two ‘external’ sources of health-care related data: (1) analyses performed by clinicians/researchers and (2) information about other patients’ bodies.

In this section, I would like to emphasize another difference between ND extraction and body organ harvesting, namely, that the former does not involve any *organic material transfer*. This can be explained by using Borgatti’s distinction between information replication and transfer (Borgatti, 2005). Communication via transfer is simply moving an information carrying object from a source to a receiver (e.g. mailing a letter to another person). In turn, communication via replication is reproducing a new (materially different) copy of a message at another point of a network, i.e. the original material message does not leave the source (e.g. sending a file from one computer to another or the spread of a viral infection through direct contact).

Unlike a biopsy, in which the tissue containing medical data is extracted and preserved for analysis, ‘harvesting’ ND is often similar to information replication. This can be exemplified by EEG recording. The basic components of an EEG system include electrodes or voltage sensors, amplifiers, and output devices. First, the data contained in neural waves of ions is replicated by a materially different signal constituted by electrons in EEG electrodes, which are affected by (or correlated with) those waves. After this second signal is amplified, data is replicated again in a new material format by the output devices. In classic analog EEG this involves a galvanometer-driven pen-writing system which offloads ND on paper. In current digital EEG the amplified signal is sent to an ADC (analog-to-digital converter) circuit that produces a digital signal (a high-resolution sampling of the original analog signal) constituted by strings of digits that can be stored in a variety of physical media (Yeh, 2012). This means that the physical ND register that is kept by clinicians or researchers for analysis is often not constituted by any organic material at all.

The key question that will guide what follows is whether the non-organic properties that ND *do* possess could be considered analogous to something that is constitutive of ourselves. In the next section, I will suggest that some specific neurocognitive properties of our brains are plausible candidates.

## 4 A Neurocognitive Approach to Neural Data Protection

In what follows, I trace the properties required for building the ND analogy. I will tackle this task by addressing three key questions: (i) ‘are there brain properties that can be implemented by inorganic material?’ (4.1), (ii) ‘are the specific properties that define our ND properties of this kind?’ (4.2) and (iii) ‘is there a sense in which the specific properties that define the ND of a subject *s* are *unique* to her brain?’ (4.3–4.6). The upshot of this section is that natural semantic personal information about *s*’ neural processes is (i) a kind of medium independent property implemented by *s*’ brain that (ii) defines her ND and (iii) constitutes the informational domain of a key neural mechanism in *s*’ brain but not in other brains.

#### 4.1 Neurocognitive Properties and Medium Independence

The absence of material or organic transfer in ND extraction suggests that the ND analogy requires properties that are constitutive of ourselves but whose (type) identity is relatively *independent* from (or is not constituted by) the physical or material medium implementing them. Functional properties are a kind of property paradigmatically characterized by this ontological autonomy. This idea has often been grounded on classic arguments for multiple realizability and, crucially, it has been argued that many psychological states are autonomous in this sense (e.g. Putnam, 1967). Therefore, although ND are not analogous to purely physical states of our body, they may be analogous to psychological states.

I will argue that this idea is on the right track. However, a caveat is in order. Functionalism about mental properties has been challenged by a dominant view in the philosophy of cognitive science. Aligned with the development of cognitive neuroscience at the beginning of the XXI century, advocates of the ‘new mechanism’ suggest that the functional properties studied in the biological sciences cannot be understood as ontologically independent from their underlying structural properties (e.g. Bechtel, 2008; Craver, 2007). Specifically, research in neuroscience reveals that cognitive mechanisms are constituted by both structural and functional properties that constrain each other (Bechtel, 2008). The conceptual framework underlying much of cognitive neuroscience is inconsistent with the distinction between (and the ontological autonomy of) a functional and a structural or implementation level (Boone & Piccinini, 2016). This seems to entail that mental mechanisms are constituted by organic structural properties which, as we saw, are not part of what ND are.

Nevertheless, mechanistic philosophers (perhaps most notably Gualtiero Piccinini and Marcin Milkowski) have pointed out that some cognitive mechanisms are constituted by kinds of (functional and *structural*) properties that exhibit some degree of independence from their physical medium. For instance, Piccinini (2015) has suggested that neural information and neural computation are constituted by ‘medium independent’ properties.<sup>4</sup> Computational and informational processes in both neural and artificial systems are medium independent in the sense that the rule (i.e. the input–output maps) that define a process of this kind,

is sensitive only to differences between portions of the vehicles along specific dimensions of variation—it is insensitive to any more concrete physical properties of the vehicles. Put yet another way, the rules are functions of state variables associated with a set of functionally relevant degrees of freedom, which can be implemented differently in different physical media. Thus, a given computation can be implemented in multiple physical media (e.g. mechanical, electro-mechanical, electronic, and magnetic), provided that the media possess a sufficient number of dimensions of variation (or degrees of freedom) that can be appropriately accessed and manipulated and that the components

<sup>4</sup> See also Milkowski (2013, p. 70) for the related idea that computational and informational properties are ‘substrate neutral’.

of the mechanism are functionally organized in the appropriate way. (Piccinini & Bahar, 2013, p. 458).<sup>5</sup>

Thus, although ND are independent from organic material, they may be defined by some of the medium independent or substrate neutral (computational and/or informational) properties that constitute our minds *as neurocognitive systems* (Boone & Piccinini, 2016).

## 4.2 ND as Natural Semantic Personal Information ND作为自然语义个人信息

As I mentioned, ND is fundamentally *personal information about neural states, processes and structures*. They are information about properties within this domain. In the philosophy of neuroscience, the idea of carrying information about something else is often understood through the notion of *semantic information*. This notion is different from Shannon's mathematical notion of mutual information, which does not determine the specific content of the (random variable defining a) signal but rather only the (average) amount of information it carries about a given source (i.e. how much it minimizes the uncertainty regarding the source variable).

Piccinini and Scarantino (2011) distinguish between non-natural semantic information (nNSI) and natural semantic information (NSI). Although there is an ongoing debate regarding its ultimate nature,<sup>6</sup> NSI is sometimes understood in terms of reliable correlations or co-variation between event types (Dretske, 1981). An event token *a* of type A carries natural information about an event token *b* of type B just in case A reliably correlates with B. Tree-rings (together with other variables) carry information about tree age, acute loss of smell or taste carries NSI about Coronavirus disease 2019, and so on. This is a notion often employed for characterizing neural signals: these are 'about' the properties that define their receptive fields. The receptive field of a neuron (or neural population) is characterized in terms of the specific environmental properties (e.g. bars, edges, borders, etc.) whose instantiation is reliably correlated with the instantiation of specific neural response properties (e.g. variations in spike rate). By contrast, nNSI does not require signals to be reliably correlated with what they are about. The aboutness of signals carrying this kind of information relies on alternative processes, such as social conventions. For instance, although the hour hand of a broken clock is not correlated with the hours of the day, it can carry non-natural (and most often incorrect) information about time in virtue of a social convention connecting its positions to different hours.

Although some pieces of ND may carry nNSI (see below), they are only defined by their NSI. The different ND production processes entail that ND are always correlated with properties of neural signals and therefore carry NSI about them. For instance, we saw that properties of neural waves of ions are correlated with

<sup>5</sup> This paragraph does not mention information, it only characterizes the medium-independence of computational processes. However, Piccinini & Bahar (2013) argue that information processing implies computation precisely because it is medium-independent in the same sense.

<sup>6</sup> e.g., see Scarantino (2015) and Martínez (2019) for interesting discussions.

properties of electrons in EEG electrodes, which in turn are correlated with the strings of digits produced by the analog-to-digital converter. Similarly, the typical systems employed in single cell recordings are based on correlations between voltage changes in neurons, states of microelectrodes, amplifiers and recording devices. Neuroimaging techniques such as PET or fMRI also rely on correlations between output signals received by digital devices and neurophysiological variables (e.g. blood flow or oxygen level) that are in turn correlated with brain activity patterns and neurocognitive processes.

Recall, however, that ND are defined as *personal* information about neural processes. This is not merely NSI about properties of a subject *s*' brain but, crucially, information that can be *traced back* to *s*. In other words, a state of a structure carries personal neural information only if it is correlated with the instantiation of a given neural property *P* in a particular subject *s*. Information about the brain is not always personal in this sense. Some structures have states that are reliably caused by the instantiation of *P* in different subjects. Therefore, these states only provide 'existential' information about *P*'s instantiation (i.e. the information that there is *a* brain in which *P* was instantiated). For instance, a specific configuration of electrons in EEG electrodes could be caused by the instantiation of a specific kind of neural activity pattern in *any* of many different brains. Therefore, this signal only tells us that the pattern was instantiated in *some* brain attached to the EEG electrodes, without being able to tell to whom the brain belongs.<sup>7</sup>

The main reason why we focus on *personal* neural data protection is that the most prominent contexts in which neural data can be used to harm us, mostly by enabling discriminatory or stigmatizing behavior, are contexts in which our information can be linked specifically to us. It is for this reason that one of the key technologies that has been developed for data protection are *anonymization* methods. These aim at transforming data in such a way that they cannot be traced back to the individuals that originated them, that is, in a way such that subjects cannot be re-identified.

Identification, the process of producing personal data, requires to put a piece of 'existential' data *d*<sup>l</sup> (i.e. information about the instantiation of a given property *P* in *some* subject) such as most of the digital records produced by EEG devices or neuroimaging technologies, together with other pieces of information, *d*<sup>2</sup>, *d*<sup>3</sup>, ..., *d<sup>n</sup>*, such that a structure that carries the set {*d<sup>l</sup>*, *d<sup>2</sup>*, *d<sup>3</sup>*, ..., *d<sup>n</sup>*} would be specifically correlated with the instantiation of *P* in a subject *s*. Identification is often achieved through what are known as 'identifiers', which are pieces of information uniquely related to *s*, such as a name or passport number. When these identifiers are removed, we have de-identified data. However, given that there are alternative ways to achieve identification, de-identified data is not equivalent to anonymized data. Anonymization requires removing or distorting '*quasi-identifiers*', which are data not uniquely

<sup>7</sup> It must be noticed that, exceptionally, some specific activity patterns do carry personal information. There are several studies that explore how to trace an individual's identity through specific properties of brain waves produced during, for instance, resting state with closed eyes (Rocca et al., 2012) or imagined speech (Brigham & Kumar 2010) and also through properties of Visual Evoked Potential (Palaniappan & Mandic 2007).

related to a particular subject (e.g. civil status, age, gender, etc.), but which can be used together with other information for reliably identifying her (i.e. an identifier can be produced by putting together different quasi-identifiers) (Salles et al., 2017). Notice that some common identifiers, such as names, are typical examples of items that can carry non-natural information (in the sense that their aboutness could be at least partially determined by social conventions). However, these items will function as *identifiers* and therefore constitute personal data only if they can be used to track an individual, what requires a reliable correlation between the item and the individual. This means that only the natural information carried by them constitutes personal data.

ND is then the NSI carried by a state  $n$  of a given structure  $N$  that is reliably correlated with the instantiation of a given property  $P$  in a subject  $s'$  brain but not with the instantiation of  $P$  in a different brain.

An implication of this characterization of ND that will be crucial for the present proposal is that NSI can include what we can call an ‘extensional’ or ‘referential’ dimension. If we understand NSI in terms of co-variation between event types, we can not only distinguish between the NSI that a property  $P$  was instantiated and the NSI that a property  $Q$  was instantiated, but also between the NSI that  $P$  was instantiated in an object  $s$  and the NSI that  $P$  was instantiated in an object  $t$ . Being about the particular subject instantiating a given property, personal information always includes this extensional dimension. This will be relevant in the following sections for determining how ND it is related to cognitive processing.

### **4.3 Psychological Integrity, Information and Exclusivity**

Having determined what kind of NSI defines the ND of a subject  $s$ , we have to assess whether it is one of the states, processes or structures covered by her psychological integrity. We saw that psychological integrity is the idea that a subject’s informed consent is necessary for the alteration or manipulation her mind’s components. These components are often simply tokens of mental properties that happen to be instantiated in her brain. A token  $m$  of a mental state type  $M$  (e.g. a particular perceptual experience) belongs to (or is part of)  $s'$  mind because this is the mind in which it is instantiated. Being particulars with specific spatio-temporal properties, these tokens are not instantiated simultaneously in different brains and therefore this condition may be sufficient for explaining why  $m$  belongs *exclusively* to  $s$ .

$s'$  ND seem to satisfy this condition. Her ability to think and reason about the mental states of her brain (see Sects. 4.4, 4.5, 4.6) entails that her brain actually instantiates and processes the specific kind of information that defines her ND. However, this is insufficient for determining that this information belongs to her in the relevant sense. The problem is that when we say that some piece of personal information  $I$  about  $s$  belongs to  $s$ , i.e. that she has the *exclusive* right to control this information, we are not talking about a particular instantiation  $i$  of  $I$  but rather about the *type I* itself. Any instantiation of  $I$  belongs to her in this sense (i.e. she has the right to control any physical copy of her personal information). Nevertheless, a given information type  $I$  is most often not instantiated *exclusively* in a particular

object. Specifically, many instantiations of  $s'$  ND are realized in physical structures that are different from her brain/mind, such as digital registers in recording devices and activity patterns in other brains (see Sect. 4.4). Thus, given that ‘being instantiated in’ is not an exclusive relation between  $s$  and her ND, it cannot ground her exclusive control over this information.

For most kinds of personal information (those protected by information privacy) the exclusive connection between  $I$  and  $s$  is determined simply by *semantic content*: any token of  $I$  belongs to  $s$  because it is about  $s$  and not about any other subject.<sup>8</sup> By contrast, I will argue that  $s'$  ownership of her ND can be grounded on psychological integrity because there is another ontological connection between this kind of information and her brain that satisfies the exclusivity condition.

#### 4.3.1 Informational Domains and Psychological Integrity

There is a second sense in which information can be part of our brain systems, a sense that involves more than the mere instantiation of information. Different kinds of information can shape our cognitive architecture, that is, the set of relatively fixed or stable structures through which mental capacities are implemented (Pylyshyn, 1998). I will suggest that although most of the properties that define cognitive architectures are shared by different brains, neural data about a particular brain is a unique aspect of that brain’s architecture.

A notion often employed in characterizing how different types of information mold our cognitive systems is that of ‘domain specificity’. Some kinds of mental mechanisms are said to constitute modules, which are systems specialized for realizing specific cognitive functions (Fodor, 1983) and are often defined by a set of special features such as informational encapsulation and inaccessibility, fast and mandatory processing and fixed neural architecture, among others. One of the key features implied by their functional specialization is their domain specificity. A system is domain specific to the extent there is a kind of information it is dedicated to process (Robbins, 2017, see also Carruthers, 2006; Samuels, 2000). These often imply a more fine-grained distinction than sensory modalities. Some typical examples include systems for color perception, visual shape analysis, sentence parsing, and face and voice recognition (Fodor, 1983, p. 47). The domain of a mechanism  $m$  is defined by a general type of information  $I$  only if all the particular pieces of information processed by  $m$  fall under  $I$ .

Of course, defining a domain in  $s'$  brain is often not an exclusive relation between a kind of information  $I$  and her brain. The domains of many cognitive systems are widely shared by different brains (e.g. information about visual shape, motion, color, etc.). This relation can ground  $s'$  exclusive rights regarding  $I$  only if  $I$  defines a domain that is unique to her cognitive architecture. In the following sections, I will try to determine whether personal information about  $s'$  brain satisfies this condition.

<sup>8</sup> For instance, the EU Charter of Fundamental Rights describes privacy as the general right to the protection of ‘personal data concerning him or her’ (Article 8).

### 4.3.2 Intensional and Extensional Domains

What does it mean for a domain to be shared by mechanisms in different brains? Most often, informational domains are understood in an ‘intensional’ sense. When we say that different cognitive systems share a domain this often simply means that they are dedicated to process information about the same *kinds of properties*. However, we could also say that domains are shared in the sense that they process information about the instantiation of those properties *in the same set of objects*. In this case, they share what we can call the ‘extensional’ dimension of a domain. For instance, visual sub-systems of different brains can process information about a common set of properties, such as shape, position, motion, color, etc. (shared intensional domains), instantiated in a common set of objects that are available for all of them in the external environment (shared extensional domains). By contrast, interoceptive mechanisms in different brains process information about the instantiation of the same properties (e.g. states of cardiovascular, respiratory or gastrointestinal systems) in *different* objects, i.e. the body to which each mechanism belongs. That is, the domains of these mechanisms are extensionally unique (more on this in Sect. 4.5).

Why should we take this extensional dimension to be part of what defines a cognitive domain? We saw that a domain is defined by a particular kind of informational content and that the main (or most common) characterization of content in neural mechanisms is in terms of semantic information. As I showed in Sect. 4.2, semantic information (or specifically NSI) is sensitive to differences in both intensional and extensional aspects of content. We can distinguish in co-variational terms information about the instantiation of a property in a given subject from information about the instantiation of the same property in a different subject. Therefore, given that it is embedded in the very notion of semantic information, it is reasonable to include this extensional dimension as part of the content that defines a given cognitive domain.

In what follows, I will argue that whereas the different instantiations of the brain system dedicated to process information about *other* brains/minds share intensional and extensional domains (4.4), the different instantiations of a system dedicated to process personal information about our own brain have extensionally unique domains and therefore satisfy the exclusivity condition (4.5 and 4.6).

## 4.4 The Domain of the Modular Mindreading System

Before characterizing the relationship between a subject and her ND, I will make a brief point that is critical to explaining (through the criterion proposed in 4.3.1) the fact that this information is not owned by *other* subjects. How is *s'* ND processed in *other* brains? A widely studied mechanism that processes information about other brains or, more specifically, about their mental states, such as thoughts, feelings, goals, etc., is the so called ‘theory of mind’ or ‘mind reading’ system (Mahy et al., 2014). Interestingly, one of the prominent mindreading models posits that this faculty constitutes a mental module, characterized by (among other features) a specific

informational domain. The key observation is that this domain is shared by the instantiations of this system in different brains. Thus, although  $s'$  ND is part of an informational domain in other subjects' brains, this domain is not unique to any of them.

Modularity theories (e.g. Baron-Cohen, 1995, 1998; Leslie et al., 2004; Scholl & Leslie, 1999) postulate that the theory of mind (ToM) development is driven by an innate neural mechanism dedicated to mental state reasoning. Leslie and colleagues have proposed the most fully articulated modularity theory of ToM (Leslie et al., 2004, Scholl & Leslie, 2001, German and Hehman, 2006). Crucially, Leslie claims that this innate theory of mind mechanism (ToMM) is characterized by a proprietary system of representation, the M-representation, that is constituted by very specific kinds of information. The M-representation provides descriptions of three-place relations involving four kinds of information: the agent involved, her attitude, an aspect of the world that anchors this attitude and the attitude's content. In a typical example, the M-representation represents that mother (the agent) pretends-true (attitude) that a banana (anchor) is a telephone (content) (Leslie, 2000).

It is clear that ToMM instantiations in different brains attribute the same kind of properties (i.e. mental states or processes) to other brains, that is, they share an intensional domain. We must determine whether they also share an extensional domain (i.e. whether they attribute these properties to the same set of objects). The internal structure of ToMM's representations makes it easy to determine this, as these representations explicitly represent the subject of a mental state through the 'agent' variable. Leslie (2000) affirms that all the parameters of the M-representation system are highly flexible. The system is not only able to represent different mental attitudes, anchors and contents but also any agent, including ourselves ('I pretend-true of the banana that it is a telephone'). Thus, ToMM is a capacity for processing information about the mental states of just *any* mind. This means that even if we suppose that mindreading is constituted by a very narrow informational domain, as modular theories do, its extensional and intensional dimensions are shared by the instantiations of the system in different brains. Given that this is the domain of the system in the brain of a subject  $t$  that processes the ND of another subject  $s$ , according to the criterion in 4.3.1 this information is not owned by  $t$ .

#### 4.5 Interoceptive Mechanisms Have Unique Informational Domains

After considering how  $s'$  ND is related to other brains, we can move on to characterizing its relation to her own brain. We have to determine whether personal information about  $s'$  brain constitutes a domain in her brain and whether this domain is unique to her cognitive architecture. In this section, I will use another kind of information (namely, personal information *about s' body*) as a paradigmatic example of how these two conditions can be satisfied and, in the next section, I will argue that the same line of reasoning can be applied to ND.

The system (or set of mechanisms) dedicated to process information about our bodies is interoception. Interoception can be regarded as a general sense of the internal states of the body. This system is constituted by different brain regions, such

as the brainstem, thalamus, insula, somatosensory, and anterior cingulate cortex, that are in charge of sensing and integrating signals originating from different body systems, such as cardiovascular, respiratory, gastrointestinal, genitourinary, thermoregulatory, endocrine, immune and nociceptive systems. The interoceptive system relies on structures that relay signals from body systems to the central nervous system, consisting in two main interconnected pathways of ‘sympathetic afferents’ that provide input to lamina I and ‘parasympathetic afferents’ that provide input to the nucleus of the solitary tract (Khalsa & Lapudus 2016).

In these mechanisms, the signals that neural structures produce respond selectively to the instantiations of a specific property in the body of a subject  $s^1$  and, crucially, they would not respond to the instantiation of that property in another subject  $s^2$ . According to the definition of personal information in Sect. 4.2, this implies that the NSI that these afferents carry is personal. Their signals are correlated with the instantiation of a given property in a particular subject. For instance, cardioperception depends on sensory neurons in our heart that detect pressure, heart rate, heart rhythm and hormones, which serve as inputs to ascending pathways in both the spinal column and vagus nerves, travelling from our heart to the medulla, hypothalamus, thalamus and amygdala and then to the cerebral cortex. Given that a signal carrying information about a particular heart rate  $r$  depends exclusively on the described pathway connecting my heart to my brain, the signal would not be activated if  $r$  were not instantiated by my heart but rather by the heart of another person. More generally, the information that these mechanisms process is personal because their characteristic wiring pattern determines that there is *only one object at the source of the neural channel* which can instantiate the target property.

Also, the object at the source of interoceptive mechanisms in different bodies will be different. In contrast with ToMM, the domains of the interoceptive mechanisms of different brains are extensionally different. Although these mechanisms carry information about the same kinds of properties, each mechanism will only process information about the instantiation of those properties in the particular subject to which it belongs. My cardioperceptive mechanism is dedicated to gather and process information about (pressure, rate, rhythm and hormones) of my heart whereas your cardioperceptive mechanism is dedicated to gather and process information about yours. It is in this sense that the domains of interoceptive mechanisms are unique to each organism. Is there a similar mechanism dedicated to process information about our brains?

#### 4.6 Neural Data as the Global Neuronal Workspace’s Domain

It has been suggested that there are interoceptive mechanisms dedicated to sense the brain’s internal ‘microenvironment’. Brain interoception may be implemented by astrocytes, a ‘star shaped’ glial cell has been found to sense changes in the brain parenchymal levels of metabolic substrates (oxygen and glucose), metabolic waste products (carbon dioxide), endocrine signals and even blood flow, and contributes thereby to the adaptive homeostatic responses coordinated by neuronal networks (Marina et al., 2018; Teschemacher et al., 2015; Turovsky et al., 2020).

Brain interoception entails the implementation mechanisms dedicated to manipulate personal information about our brains. This means that, according to the criterion discussed in Sect. 4.3.1, this information will be protected by our right to psychological integrity. However, in the same way as the interoceptive system described in the previous section, this kind of information is purely physiological. We may still wonder whether a highly sensitive kind of neural data, namely personal data about our *mental* or neurocognitive states, is equally protected, that is, whether there is a mechanism dedicated to manipulate this kind of information.

Dehaene, Changeux and colleagues posit that there are two ‘computational spaces’ in the brain. On the one side, there is a set of cortical and subcortical processors or modules characterized by narrow informational domains (e.g. line segments in V1, motion information in area MT, visual word form in the human fusiform gyrus, etc.). On the other side, they postulate the existence of a ‘global neuronal workspace’ (GNW), which is constituted by a set of cortical neurons (mostly pyramidal cells) that send projections to many distant areas through long-range excitatory axons and break the modularity of the cortex by maximizing the ability to exchange information between processors.

More specifically, through an ‘avalanche’ in which local modular signals pick up strength and are finally spread throughout parietal and prefrontal lobes, the GNW produces a sustained global or large-scale signal (ignition) reaching and connecting distant processors (Dehaene, 2014, pp. 223–225) characterized by high-frequency (gamma-band) oscillations and a massive long-distance phase synchrony (Dehaene & Naccache, 2001; Dehaene, 2014, pp. 216–262). Any given processor can have access, through this global signal, to information about the local signal that originated the avalanche. Information about any cognitive state, say, a perceptual state in my visual system, can be sent through the GNW to, for instance, language, long-term memory, attention, and intention systems and become “the subject of a sentence, the crux of a memory, the focus of our attention, or the core of our next voluntary act” (Dehaene, 2014, p. 312). Dehaene hypothesizes that the entry of inputs into this workspace constitutes the neural basis of access to consciousness.

Thus, just as personal information about our body constitutes the domain of interoceptive mechanisms, it seems that personal information about our neurocognitive states constitutes the domain of the GNW. There are, however, two apparent (and related) differences that might give a different impression. In the first place, the GNW broadcasts local signals that very often carry semantic information about features of the *external* environment (e.g. visual shape information or information about bodily states) rather than about internal properties of the brain. In the second place, the GNW seems to be anything but domain specific. Its key function is precisely to minimize modularity by broadcasting the different kinds of information processed by modular systems. It seems odd to say that it has a proprietary informational domain.

Regarding the first problem, we should notice that the GNW provides global access to information about our environment *through* access to information about the internal states that carry this environmental information. We have conscious access to the fact that we are seeing, hearing, remembering, etc., a given piece of external information. In other words, the GNW signals carry semantic information about

local neural signals that happen to be ‘vehicles’ of environmental information (i.e. that carry in turn semantic information about the environment). This also provides an answer to the second problem. The different kinds of information broadcasted by the GNW do not form an unruly conjunction but rather share informational properties that can define a domain in the sense specified in Sect. 4.3. The GNW provides access to very different kinds of external information *as* the informational contents of internal states. Information about some kind of internal (visual, auditory, motor, somatosensory, etc.) neural signal is always nested in GNW signals. They share the property of being about (i.e. carrying NSI about) internal neural states that are about something else. That is, GNW signals are about other neural signals, i.e. activity patterns that carry semantic information.<sup>9</sup>

More importantly, these signals carry *personal* information about *my* neurocognitive states. A GNW signal that responds to a local (e.g. visual) signal  $\sigma^1$  in my brain would not be activated by the presence of a signal  $\sigma^2$  of the same kind (e.g. a local signal with the same informational content and coding properties) in *other* brain. This information is personal because the GNW’s pyramidal neurons constitute exclusive channels between the different systems of my own brain. Also, the source of each instantiation of the GNW will be different. Although they all carry information about neural signals, each GNW will only process information about the instantiation of these signals in the particular brain to which it belongs. That is, each GNW instantiation has an extensionally unique domain.

What does this entail regarding ND’s special ontological status? Ienca & Andorno (2017) suggest that what is special about ND is that they are not merely information about our neural or neurocognitive states. Unlike other kinds of information, they have a strong ontological connection with its source in the sense that they cannot be easily distinguishable or even separated from mental/neural states. This is what they call the “inception problem”. I interpret their insight as articulating the fact that neural information is sometimes not merely information about the brain, but also information *in* the brain, that is, a component part of neurocognitive processes. As I argued in Sect. 4.3, this is not sufficient for supporting the idea that  $s'$  ND are part of her mind. When we say that we own some piece of ND we are not talking about a particular token  $i$  of information  $I$  that happens to be instantiated in my brain, but about the very information type  $I$ , which can also be instantiated outside my brain. Therefore, we needed to determine whether information *types* (and specifically ND) can bear some kind of exclusive ontological relation with a given subject.

The argument in this section aims at showing precisely this.  $s'$  ND are a kind of information that constitutes her neurocognitive identity by shaping unique aspects of her cognitive architecture, not unlike a mental fingerprint. It is in this sense that  $s'$  ND are not simply information about  $s'$  brain/mind but part what  $s'$  brain/mind

<sup>9</sup> Notice that I do not claim that the GNW has a metacognitive component, at least not in the sense that metacognition requires meta-representations. Semantic information does not necessarily entail representation (Piccinini & Scarantino 2011). My point is simply that GNW signals track (and therefore carry semantic information about) other neural signals (which in turn track environmental properties).

is made of. If s' ND constitute a unique informational domain and informational domains are part of the very make up of cognitive agents, then collecting, analyzing and applying s' ND is indeed analogous dealing with the very mental architecture that constitutes her as a person.<sup>10</sup> Thus, we can conclude that this information is one of the properties to which her right to psychological integrity applies.

A final worry I would like to address is related to the fact that non-consensual ND manipulation seems substantially different from paradigmatic violations of psychological integrity. Although I argued that there is a sense in which accessing our ND is equivalent to accessing our minds, there may be a relevant disanalogy between these two situations. Psychological integrity concerns the protection from non-consensual intrusions into our minds/brains, which typically involve the alteration or modulation of neurocognitive processes through some kind of technological intervention (such as in cases of “brain-hacking”, see Ienca & Haselager, 2016). However, it seems that collecting, analyzing and applying s' ND does not involve any intrusion of this kind. Manipulating pieces of s' ND that are instantiated outside her brain (e.g. in a digital register) may not affect her mental processes, as this information seems to be *functionally or causally disconnected* from them.<sup>11</sup> As we saw, GNW signals are fed into language, long-term memory, attention, and intention systems (i.e. they are functionally integrated with them) and therefore modulating or manipulating these signals would plausibly affect the cognitive processes implemented by these systems. By contrast, manipulating a piece of ND in a digital register may have no such effect on these processes.

Nevertheless, interventions on ND do have an effect on cognition, albeit a subtler one. The key function of the GNW is to select and amplify specific local signals, thus broadcasting them to other brain systems and (through our language system) to other agents (i.e. we can only report the mental information we have conscious access to). By having and providing access to information about s' neurocognitive processes and states, an agent manipulating an ND external register bypasses s' GNW, *overruling its control over which pieces of s' ND are shared and which are not*. It is in this sense that non-consensually accessing and sharing ND is analogous to a psychological integrity violation even in this more “brain-hacking” oriented sense.

This connection between mental privacy and psychological integrity is related to the idea that privacy essentially the cognitive ability to express ourselves selectively, which could be undermined by technological mind-reading (Ienca & Andorno, 2017). The present observation entails that this ability, which is often associated

<sup>10</sup> Notice that what makes this information worthy of special protection is not merely the fact that it constitutes a unique domain within an information processing system. There is nothing *ethically* special regarding, for example, information about the temperature of an individual fridge, despite the fact that the fridge has a thermostat dedicated to process this ‘personal’ information. This information may be part of the fridge’s identity as an artifact or information processing system, but this does not have any ethical implication regarding how this information should be manipulated (I thank an anonymous reviewer for this example and observation). It is the fact that ND is part of a *person’s* identity what makes it normatively special. This makes it part of what would be covered by her psychological integrity.

<sup>11</sup> I thank an anonymous reviewer for this observation.

with conscious decision-making and planning, is actually more deeply rooted in our cognitive architecture, involving the selection and amplification of pre-conscious signals by the very mechanism underlying consciousness. The connection between mental privacy and the GNW shows that the former is aimed at protecting a very fundamental aspect of our subjectivity.

## 5 Conclusion

In this paper I tried to supply the analogy (missing from the Morningside Group's proposal) required for making ND protection more stringent than the protection of other kinds of personal information. The basic strategy was to show that although ND are *prima facie* very different from body organs and organic tissue, they are analogous to properties of our minds and are therefore covered by psychological integrity. Following Piccinini (2015), I argued that *s'* ND constitute a kind of medium independent property that can be instantiated in her brain but also in non-organic material, and that can be characterized as natural semantic personal information about *s'* brain. I claimed that despite its multiple realizability, *s'* ND have an exclusive ontological relation with her brain. This information constitutes an informational domain that is unique to *s'* brain. All interoceptive mechanisms in *s'* brain have domains that are extensionally unique in the sense of being about the instantiation of properties in *s*. I suggested that the GNW is one of such interoceptive mechanisms, broadcasting signals that carry information about the neurocognitive states of the subject in which it is instantiated. If ND is part of our minds in this sense, then they could be protected not merely by the regulations concerned with information privacy but also by those addressing psychological integrity.

It is worth emphasizing that this connection between mental privacy and psychological integrity lines up with a version of the Chilean Constitutional Reform Bill being currently discussed by the Chamber of Deputies. This texts explicitly includes ND as part of psychological integrity by affirming that "Scientific and technological development shall be at the service of people and shall protect their life and physical and mental integrity, *including brain activity and information derived from it.*" (emphasis added). Also, Article 1.a of the Neuroprotection Bill affirms that the law aims to "[p]rotect the physical and mental integrity of individuals, *through the protection of the privacy of neuronal data [...]*" (emphasis added).

A final question, which is beyond the scope of this paper, is what specific ND regulations would entail psychological integrity. It seems reasonable to assume that these would be more stringent than those provided by information privacy because (in line with the Morningside Group's proposal) the present approach entails treating ND as if it were part of ourselves. However, one may still wonder whether psychological integrity would ground the same ND regulations as bodily integrity. Although I will not address this issue here, I would like to suggest why I think this is probably the case.

In a thorough discussion on the nature and scope of bodily integrity, Herring & Wall (2017) argue that our intimate relation to our bodies depends on the fact that it is the point of implementation or realization of morally relevant aspects of our mind,

subjectivity and experience. For instance, states of well-being, pain and pleasure, states of flourishing, communing and relating are “all states that are located somewhere in the chain of physiological systems [of our bodies]” (p.13). The right to bodily integrity gives a person exclusive use of, and control over, their body on the basis that the body is the site, location, or focal point of their subjectivity.

If this idea is on the right track, then the relation between bodily and psychological integrity is straightforward. If the rights and regulations related to our body are grounded on the relation that the body has with some morally valuable aspects of the mind, then it seems that they should also be applied to the components of the mind itself (our neurocognitive states, processes and structures), which are (at least) as intimately related to those aspects as our body. The inviolability of our minds (i.e. psychological integrity) should ground the same restrictions that are applied to the manipulation of our bodies (e.g. the prohibition to commercialize its components). Specifically, the present proposal could ground the regulations mentioned in Article 7 of the Chilean bulletin No 13.828-19.

Ultimately, the present discussion aims at fostering an urgent international debate surrounding the regulation of mental privacy. Among other reasons, an international consensus (which will be the focus of the UN neurorights agenda, see Yuste et al. 2021), may be necessary for shielding countries like Chile and Spain from losing potential investors in “neuro-rights havens” with lax regulations. Also, I tried to show that the concepts and frameworks developed around central issues in the philosophy of mind and cognitive neuroscience have a window of opportunity for shaping the upcoming and urgent legislations required for regulating the rapid development and increasing applications of neurotechnologies. Philosophy could contribute to building an ethical relationship with them before they become an integral part of our societies, when substantial amendments will be plausibly difficult to implement.

**Acknowledgments** I would like to thank the members and friends of Neuroética UAH (specially, Roberto Rubio, Francisco Pereira and Juan Manuel Garrido) and the Neuroethics Program at CIF (specially Arleen Salles, Nicolás Serrano and Paula Castelli) for early discussions of this material and the organization of academic events that contributed to its development. I am also grateful to the organizers of the International Neuroethics Society 2020 meeting and editors of the Neuroethics Blog from the Neuroethics Program at Emory University (specially Karen Rommelfanger), and anonymous reviewers from Minds & Machines for comments on the main ideas presented in this manuscript. Also, for their relentless work on neurorights and the main ideas on which my view on mental privacy was inspired, many thanks to Sara Goering, Rafael Yuste, Marcello Ienca and Roberto Andorno. Finally, I am indebted to Julieta Picasso Cazón for ongoing support.

**Funding** Funding was provided by fondo nacional de desarrollo científico y tecnológico (Proyecto FONDECYT Regular #1210091, Proyecto Fondecyt Regular #1200197).

## References

- Baron-Cohen, S. (1995). *Mind Blindness: An Essay on Autism and Theory of Mind*. MIT Press.  
Baron-Cohen, S. (1998). Does the study of autism justify minimalist innate modularity? *Cognitive Development*, 10, 179–191.

- Bechtel, W. (2008). *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. Routledge.
- Boone, W., & Piccinini, G. (2016). The cognitive Neuroscience Revolution. *Synthese*, 193(5), 1509–1534.
- Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, 27(1), 55–71.
- Brigham, K., & Kumar, B. V. (2010). Subject identification from electroencephalogram (EEG) signals during imagined speech. *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)* (pp. 1–8). IEEE.
- Carruthers, P. (2006). *The architecture of the mind*. Oxford University Press.
- Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Clarendon Press.
- Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin.
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79, 1–37.
- Dretske, F. I. (1981). *Knowledge and the flow of information*. MIT Press.
- Fernandez, A., Sriramam, N., Gurewitz, B., & Oullier, O. (2015). *Pervasive Neurotechnology: A Ground-breaking Analysis of 10,000+ Patent Filings Transforming Medicine, Health, Entertainment, and Business*. SharpBrains.
- Fodor, J. A. (1983). *The modularity of mind*. MIT Press.
- Goering, S., Klein, E., Sullivan, L. S., Wexler, A., Arcas, B. A., Bi, G., & Yuste, R. (2021). Recommendations for responsible development and application of neurotechnologies. *Neuroethics*. <https://doi.org/10.1007/s12152-021-09468-6>
- Goering, S., & Yuste, R. (2016). On the necessity of ethical guidelines for novel neurotechnologies. *Cell*, 167(4), 882–885.
- Herring, J., & Wall, J. (2017). The nature and significance of the right to bodily integrity. *The Cambridge Law Journal*, 76(3), 566–588.
- Ienca, M., & Andorno, R. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy*, 13(1), 1–27.
- Ienca, M., & Haselager, P. (2016). Hacking the brain: Brain–computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology*, 18(2), 117–129.
- Khalsa, S. S., & Lapidus, R. C. (2016). Can interoception improve the pragmatic search for biomarkers in psychiatry? *Frontiers in Psychiatry*, 7, 121.
- Lamond, G. (2006). Precedent and Analogy in Legal Reasoning, in Edward N. Zalta (ed.) The Stanford Encyclopedia of Philosophy. <https://plato.stanford.edu/archives/spr2016/entries/legal-reas-prec/>.
- Lavazza, A. (2018). Freedom of thought and mental integrity: The moral requirements for any neural prosthesis. *Frontiers in Neuroscience*, 12, 82.
- Leslie, A. M. (2000). ‘Theory of mind’ as a mechanism of selective attention. *The New Cognitive Neurosciences*, 2, 1235–1247.
- Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in ‘theory of mind.’ *Trends in Cognitive Sciences*, 8, 528–533.
- Mahy, C. E., Moses, L. J., & Pfeifer, J. H. (2014). How and where: Theory-of-mind in the brain. *Developmental Cognitive Neuroscience*, 9, 68–81.
- Marina, N., Turovsky, E., Christie, I. N., Hosford, P. S., Hadjihambis, A., Korsak, A., & Gourine, A. V. (2018). Brain metabolic sensing and metabolic signaling at the level of an astrocyte. *Glia*, 66(6), 1185–1199.
- Martínez, M. (2019). Representations are Rate-Distortion Sweet Spots. *Philosophy of Science*, 86(5), 1214–1226.
- Matz, S. C., Appel, R. E., & Kosinski, M. (2020). Privacy in the age of psychological targeting. *Current Opinion in Psychology*, 31, 116–121.
- Mecacci, G., & Haselager, P. (2019). Identifying criteria for the evaluation of the implications of brain reading for mental privacy. *Science and Engineering Ethics*, 25(2), 443–461.
- Meynen, G. (2019). Ethical issues to consider before introducing neurotechnological thought apprehension in psychiatry. *AJOB neuroscience*, 10(1), 5–14.
- Milkowski, M. (2013). *Explaining the computational mind*. MIT Press.
- Montgomery, J. (2017). Data sharing and the idea of ownership. *The New Bioethics*, 23(1), 81–86.
- Palaniappan, R., & Mandic, D. P. (2007). Biometrics from brain electrical activity: A machine learning approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4), 738–742.
- Piccinini, G. (2015). *Physical Computation, a Mechanistic Account*. Oxford University Press.

- Piccinini, G., & Bahar, S. (2013). Neural computation and the computational theory of cognition. *Cognitive Science*, 34, 453–488.
- Piccinini, G., & Scarantino, A. (2011). Computation vs. information processing: Why their difference matters to cognitive science. *Studies in History and Philosophy of Science*, 41, 237–246.
- Putnam, H. (1967). Psychological Predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, Mind, and Religion* (pp. 37–48). University of Pittsburgh Press.
- Pylyshyn, Z. W. (1998). Cognitive architecture. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. Taylor and Francis.
- Robbins, Philip, "Modularity of Mind", The Stanford Encyclopedia of Philosophy (Winter 2017 Edition), Edward N. Zalta (ed.). <https://plato.stanford.edu/archives/win2017/entries/modularity-mind/>.
- La Rocca, D., Campisi, P., & Scarano, G. (2012). EEG biometrics for individual recognition in resting state with closed eyes. In: 2012 BIOSIG-Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG). IEEE, pp. 1–12
- Salles, A., Stahl, B., Bjaalie, J., Domingo-Ferrer, J., Rose, N., Rainey, S., & Spranger, T. (2017). *Opinion and Action Plan on “Data Protection and Privacy.”* Danish Board of Technology Foundation.
- Samuels, R. (2000). Massively modular minds: Evolutionary psychology and cognitive architecture. *Evolution and the human mind: Modularity, language and meta-cognition*, 2, 13–46.
- Scarnitino, A. (2015). Information as a probabilistic difference maker. *Australasian Journal of Philosophy*, 93(3), 419–443.
- Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and ‘theory of mind.’ *Mind and Language*, 14, 131–153.
- Scholl, B. J., & Leslie, A. M. (2001). Minds, modules, and meta-analysis. *Child Development*, 72, 696–701.
- Shen, F. X. (2013). Neuroscience, mental privacy, and the law. *Harvard Journal of Law and Public Policy* 36, 653.
- Teschemacher, A. G., Gourine, A. V., & Kasparov, S. (2015). A role for astrocytes in sensing the brain microenvironment and neuro-metabolic integration. *Neurochemical Research*, 40(12), 2386–2393.
- Turovsky, E. A., Braga, A., Yu, Y., Esteras, N., Korsak, A., Theparambil, S. M., & Gourine, A. V. (2020). Mechanosensory signaling in astrocytes. *Journal of Neuroscience*, 40(49), 9364–9371.
- Yeh, M. (2012). Digital EEG. In T. Yamada & E. Meng (Eds.), *Practical guide for clinical neurophysiologic testing: EEG* (pp. 54–62). Lippincott Williams & Wilkins.
- Yuste, R., Goering, S., Bi, G., Carmena, J. M., Carter, A., Fins, J. J., & Kellmeyer, P. (2017). Four ethical priorities for neurotechnologies and AI. *Nature News*, 551(7679), 159.
- Yuste, R., Genser, J., & Herrmann, S. (2021). It’s Time for Neuro-Rights. *Horizons: Journal of International Relations and Sustainable Development*, 18, 154–165.

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