



EPISTEMIC CHALLENGES OF DIGITAL TWINS & VIRTUAL BRAINS PERSPECTIVES FROM FUNDAMENTAL NEUROETHICS

DESAFÍOS EPISTÉMICOS DE LOS GEMELOS DIGITALES Y LOS CEREBROS VIRTUALES: PERSPECTIVAS DESDE LA NEUROÉTICA FUNDAMENTAL

Kathinka Evers^{a} and Arleen Salles^{a,b}*

Fechas de recepción y aceptación: 29 de marzo de 2021 y 20 de octubre de 2021

DOI: https://doi.org/10.46583/scio_2021.21.846

Abstract: In this article, we present and analyse the concept of Digital Twin (DT) linked to distinct types of objects (artefacts, natural, inanimate or living) and examine the challenges involved in creating them from a fundamental neuroethics approach that emphasises conceptual analyses. We begin by providing a brief description of DTs and their initial development as models of artefacts and physical inanimate objects, identifying core challenges in building these tools and noting their intended benefits. Next, we describe attempts to build DTs of model living entities, such as hearts, highlighting the novel challenges raised by this shift from DTs of inanimate objects to DTs of living objects. Against that background, we give an account of contemporary research aiming to develop DTs of the human brain by building “virtual brains”, e.g. the simulation engine The Virtual Brain (TVB) as it is carried out in the European Human Brain Project. Since the brain is structurally and functionally the most complex organ in the human body, and our integrated knowledge of its

^a Centre for Research Ethics and Bioethics (CRB). Uppsala University, Sweden.

^{*} Correspondencia: Uppsala University. Centre for Research Ethics & Bioethics (CRB). P.O. Box 564, SE-751 22, Uppsala, Sweden.

E-mail: kathinka.evers@crb.uu.se

^b Programa de Neuroética, Centro de Investigaciones Filosóficas. Buenos Aires. Argentina.



functional architecture remains limited in spite of recent neuroscientific advances, the attempts to create virtual copies of the human brain are correspondingly challenging. We suggest that a clear scientific theoretical structure, conceptual clarity and transparency regarding the methods and goals of this technological development are necessary prerequisites in order to make the project of constructing virtual brains a theoretically promising and socially beneficial scientific, technological and philosophical enterprise.

Keywords: Digital Twin, Virtual Brain, fundamental neuroethics, conceptual analysis, ontological complexity, epistemic transparency.

Resumen: En este artículo, presentamos y analizamos el concepto de gemelo digital vinculado a distintos tipos de objetos (artefactos, objetos naturales, animados e inanimados) y examinamos los desafíos que presenta su creación utilizando la perspectiva de la neuroética fundamental que enfatiza el análisis conceptual. Comenzamos con una breve descripción de los gemelos digitales y de su desarrollo inicial como modelos de artefactos y objetos físicos no animados, identificando los desafíos centrales que presenta su construcción y destacando sus beneficios. Luego describimos intentos de construir gemelos digitales de entes vivos, como el corazón, identificando los desafíos novedosos que se plantean en este caso. A continuación describimos estudios contemporáneos que tienen como objeto desarrollar gemelos digitales del cerebro humano por medio de la construcción de “cerebros virtuales”, tal como se lleva a cabo en el *Human Brain Project* europeo por medio del motor de simulación The Virtual Brain (TVB). Si consideramos que el cerebro es el órgano más complejo del cuerpo humano, tanto estructural como funcionalmente, y teniendo en cuenta que nuestro conocimiento integral de su arquitectura funcional sigue siendo limitado, los intentos de crear copias virtuales del cerebro humano constituyen un reto significativo. Sugerimos que una estructura científicamente clara y una transparencia conceptual sobre los métodos y fines de este desarrollo tecnológico son requisitos necesarios para lograr que el proyecto de construir cerebros virtuales se convierta en una iniciativa teóricamente prometedora, así como científica, social y filosóficamente beneficiosa.

Palabras clave: gemelo digital, cerebro virtual, neuroética fundamental, análisis conceptual, complejidad ontológica, transparencia epistémica.



§1. INTRODUCTION: THE CONCEPT OF DIGITAL TWIN

One of the latest advances in the field of technology is the development of the Digital Twin (DT), a computational model or digital replica of a living or non-living physical object or process. At the forefront of this development is the creation of DTs of human brains, a central line of research in the European Human Brain Project (HBP). Before discussing HBP's work on the virtual brain further below, we shall begin by describing what a DT is, why such models are developed, and what challenges they meet in distinct areas of application. Our aim in this article is to examine the different challenges involved in creating DTs linked to distinct types of objects (artefacts, natural, inanimate or living) from a fundamental neuroethics approach that emphasises conceptual analyses. We argue that a clear scientific theoretical structure, conceptual clarity and transparency regarding the methods and goals of these technological developments are necessary prerequisites in order to make the project of constructing virtual brains a theoretically promising and socially beneficial scientific, technological and philosophical enterprise.

A DT must be able to project a digital reality to provide information that, due to physical constraints, would not be available otherwise. Recent advances make them actionable: they can make the physical product more “intelligent” so that it adjusts its behaviour according to recommendations provided by the virtual twin, and make the virtual twin more fact-based so that it reflects its physical counterpart more accurately (F. Tao et al., 2019). The concept was initially used in the context of manufacturing and industry (Grieves, 2014) and since then it has been variously applied to fields such as aerospace and aviation research and increasingly in the health care context and medicine (<https://www.sdte.se/#concept>) with a number of aims (Jones, Snider, Nassehi, Yon, & Hicks, 2020).

There are different conceptions of DT, and diverse views on which main characteristics they must possess in order to qualify as such, in part depending on the context in which the DT is developed and applied (Barricelli, 2019; Jones et al., 2020). One of the first formulations of the concept describes it as “a set of virtual information constructs that fully describes a potential or actual manufactured product from the micro atomic level to the macro geometrical



level” (Grieves & Vickers, 2017). A more recent one sees it as “a real mapping of all components in the product lifecycle using physical data, virtual data and interaction between them” (F. Tao et al., 2019). DTs provide “the potential of understanding changes in the status of the physical entity through sensing data, to analyze, predict, estimate and optimize changes” (Barricelli et al., 2019: 10). Definitions of DT technology emphasise two important aspects: (1) the presence of a seamless connection¹ between the physical model and the corresponding virtual model or virtual counterpart; and (2) the establishment of this connection by generating real-time data using sensors (Barricelli, 2019).

The immediate goal of the DT is, by and large, to synchronise part of the physical world (e.g., an object or a place) with its cyber representation (which can be an abstraction of some aspects of the physical world). The ultimate goal is often to use information generated by the DT to treat (benefit, or improve) the entity that the DT is a copy of. For example, the ultimate goal of having a DT of somebody’s heart or brain is likely to be to benefit/treat the person whose heart or brain it is (cf. below) in some way. Linked with its physical twin over time, the DT is intended to be used not just in conceptualisation, testing and design but also during the whole life cycle of the replicated object and even possibly beyond (Rasheed, San, & Kvamsdal, 2020). Because the virtual-physical coupling needs to be able to identify the physical entity uniquely there has to be a one-one connection between the DT and the physical twin.² Thus, there are at least three conditions (Figure 1) that need to be met by the physical-virtual bridge in a DT:

- a seamless connection in the physical-virtual bridge
- a real-time data exchange enabling the “twinning” of this bridge
- a unique identifier allowing the necessary bijective connection between the digital and physical twins.

¹ We understand this “seamless connection” to mean that the difference, or the transition, between the object and the digital copy cannot easily be detected. For example, in the virtual brain described below, the output that is the simulated brain signal cannot be easily differentiated from experimental recordings

² Parameters of the twinning process are well reviewed in Jones (2020).

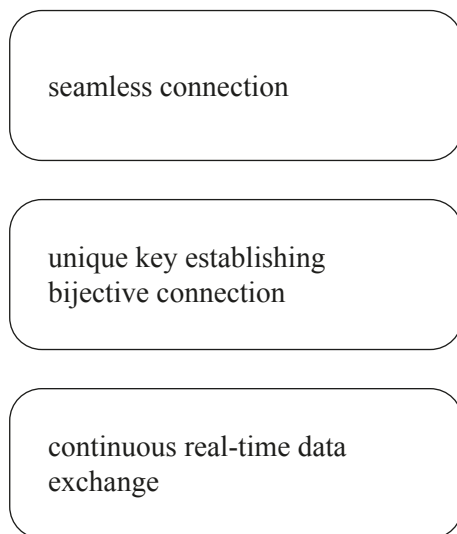


Figure 1. Necessary features of a DT's physical-virtual bridge

Recently, DTs have been described as “a living, intelligent and evolving model...always aware of what is happening in the physical world” where its ability of “recording, controlling, and monitoring the conditions and changes of the physical system enables applying AI predictive and prescriptive techniques for forecasting failures, testing the outcome of possible solutions, and activating self-healing mechanisms.” (Barricelli, 2019): 6)

From a philosophical perspective, the above formulation might seem somewhat careless, since a digital model is neither alive nor sentient and therefore cannot properly be said to “be aware” of anything. A description of DTs that uses human-centred terms might reflect the all too common tendency to project human agency to technological products (Salles, Evers, & Farisco, 2020). However, it might also be a way to emphasise the dynamic and responsive, reactive nature of the DT in its relations to the physical counterpart, as illustrated by Figure 2:

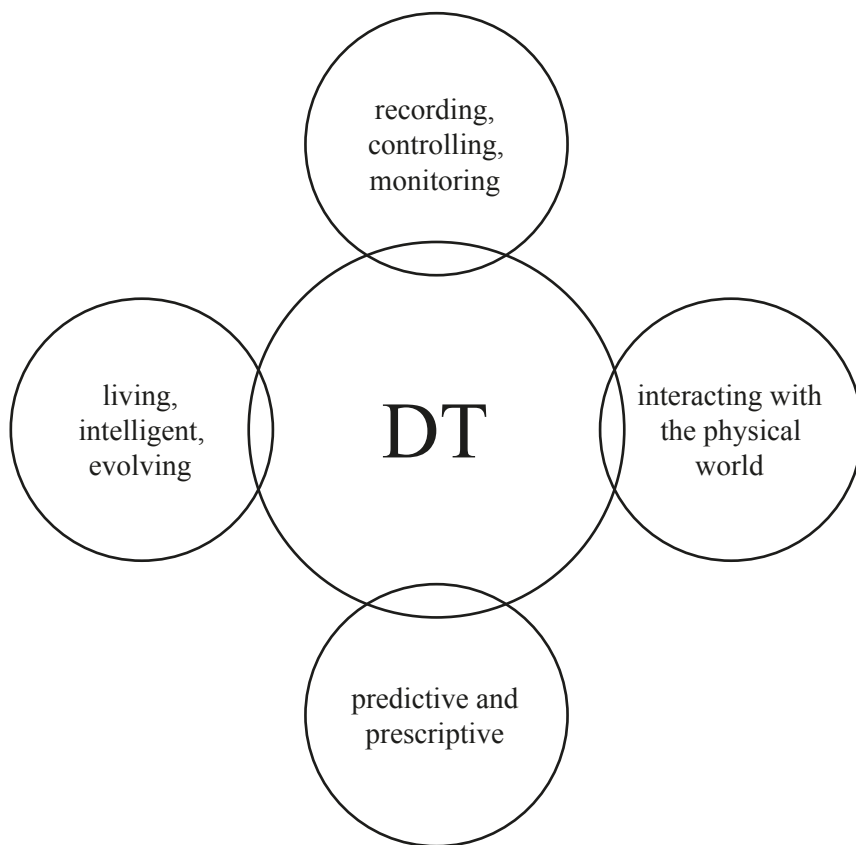


Figure 2. Features distinguishing a DT from a simple model or simulation; and from the type of “Product Avatar” (PA) that is limited to replication

§2. DIGITAL TWINS, APPLICATIONS, AND CHALLENGES

During the first development of DTs, the digitally replicated objects were physical and inanimate and the applications focused on manufacturing and industry. More recently, natural physical objects have become of interest for DT development, e.g. in innovative conservation of nature. Of special interest is the expansion of the DT framework to also include living objects.

Whether DTs replicate artificial or natural and even living objects, they must meet some minimal conditions to qualify as such. And regardless of how DTs are used, all DT endeavours have one main challenge in common, namely, the actual synchronization of the virtual and the physical (twinning) so as to make the DT accurate and reliable. This is about levels of achievable fidelity, that is, the degree of compatibility between physical and virtual parameters (Jones et al., 2020). However, whether a high-fidelity DT is achievable, whether it is sought, the extent to which it is necessary, and the potential implications depend to a great extent on the DT's physical counterpart and the contexts and goals of application. How difficult it is to achieve a digital replica may depend on the nature of the modelled object, e.g. its complexity (structural, functional, and developmental), dynamics or ontological status, which in turn depends on application domains and goals. Another important and related variable in assessing the challenges involved in building DTs is epistemic: how well do we know and understand the physical object that is to be twinned? Generally, the simpler the object, the easier it is to know and might therefore also be easier to replicate. Simple versus complex objects, artefacts versus natural objects, living organs such as hearts and brains do not pose identical challenges for DT construction and application, as we shall discuss further below.

2.1 *DT of inanimate objects: application in industry*

The virtual-physical integration enabled by the DT plays an important role in industry, as evidenced by publications, patented developments, and surveys of leading companies that point to numerous applications of DT technology in that area (F. Tao, Zhang, H., Liu, Ang, Nee AYC., 2019) (Escorsa, 2018).

When used in the context of production systems and smart manufacturing the goal of DTs is to make the product production process more reliable and predictable via monitoring, which allows for the necessary adjustments and thus facilitates control. The continuous back and forth interaction and synchronization of the digital twin, its physical twin, and its environment (DT-PT closed loop) leads to the possibility of improving the performance of the product/process in the physical space and thus to considerably improving the



manufacturing process (Kiritsis, 2011). This is key to smart manufacturing, a widely shared approach that consists precisely in the use of DTs to optimise the manufacturing process through the use of autonomous modules that execute high-level tasks without direct human intervention (Rosen, Von Wichert, Lo, & Bettenhausen, 2015).

More recently, it has been argued that the area of application of DTs should be widened, to include other life cycle processes such as design. The argument is that the use of DTs at different early stages of the design process will improve product quality and innovation (F. Tao et al., 2019; F. Tao, Zhang, H., Liu, Ang, Nee AYC., 2019).

In addition to production, DTs have typically played a significant role in product prognostics. In the aviation industry, the use of DTs to simulate material changes plays an important role in analysing and predicting the operational state of the aircraft's structure and probability of failure. Similarly, in the healthcare context, DTs have been used for predictive equipment maintenance and performance optimization (Barricelli, 2019).

An important feature of all these DTs is that their physical counterparts (that is, the physical objects that they replicate) are *manufactured* objects or spaces (i.e. DT of a radiology room): they are designed and built by humans. As such, they have a high degree of transparency; we know their functions in corresponding detail. In modelling, scientists often either explicitly or implicitly assume that, to genuinely understand a system, one should be able to reconstruct it in detail from its components. This assumption resonates with a classical maxim of scholastic philosophy, resurging in Vico (Vico, 1710/1988): only the one who makes something can fully understand it (Dudai & Evers, 2014).

That the physical counterparts of the DTs mentioned above are manufactured objects may make accurate DTs easier to achieve, and this should make an impact on their predictive and monitoring power. However, even when initially defined as virtual counterparts of manufactured objects (Grieves, 2014), the use of the concept DT and the relevant engineering framework has been expanded and it now includes *natural* objects (not designed by humans) whose structure and functions we might still be in the process of discovering.

2.2 *Natural inanimate objects: application in natural systems*

In addition to the above-mentioned applications in industry that entail the creation of DTs of manufactured objects, there are attempts to use the DT technology and build DTs of natural systems, either to better understand such systems (e.g., meteorology, (Rasheed et al., 2020)) or to monitor and predict how they interact with engineered systems (“digital mining” (Hodgkinson & Emouttie, 2020)). Not surprisingly, considering the many known unknowns of a natural system, whether a digitised representation of a natural system can be described as a digital twin is controversial.

To illustrate, digital mining, which attempts to improve productivity and safety by employing digital models, simulations, and feedbacks, requires DTs both of engineered systems and of the geological and hydrogeological systems that interact with them. However, geological and hydrogeological natural systems are epistemically opaque. Their structural and physical complexities make it extremely difficult (if at all possible) to fully know them. Of course, it could be argued that in fact such a level of detail is not really necessary to build a DT. However, if so, the notion of DT may need to be rethought, e.g., defined in more modest terms. It has been argued that, at present, considering the level of accuracy that by definition is required in a digital twin, it is not clear that a digitised copy of a geological system qualifies as one (Hodgkinson & Emouttie, 2020). This does not rule out the possibility of some type of virtual representation of the natural system, nor does it rule out the potential utility of such a representation. Yet, it does present a challenge to the idea that such a representation is a digital twin of the relevant natural system and raises the possibility of conceptualizing those replicas not as digital twins but rather as digital cousins that, while related to the real physical object are not the geological model’s digital counterpart.

2.3 *Digital twins of living objects: application in personalized and precision medicine*

Beyond inanimate natural objects, the next potentially impactful step is represented by recent attempts to borrow the engineering concept of DT and



expand it to apply it in healthcare. Because of their capacity to take inter-subject variability into account, models that computationally integrate detailed data are expected to significantly enhance clinical practice, drawing better diagnosis, prognosis, and providing patient-tailored treatments (Bruynseels, de Sio, & van den Hoven, 2018). The application of the DT paradigm to living objects represents a promising development in personalized and precision medicine. However, in this context DT development and application stands confronted with important challenges over and above those that DTs of manufactured objects encounter.

From an ontological perspective, in a very basic sense, life is the condition that distinguishes animals and plants from inorganic matter. Although notoriously difficult to define (Machery, 2012; Macklem & Seely, 2010), the presence of life (with all the biological processes that it entails) affects the dynamics of objects. Living objects' wider range of internal and external/contextual interaction and sensitivity to epigenetic mechanisms results in dynamics that are far greater than the dynamics of inanimate objects. This suggests that in building a DT of a living object it is necessary to attend to the additional dimension of complexity introduced by its dynamics, noting that it may make the DT's goals of accurately recording, predicting and monitoring even more challenging.

To illustrate, at present we are witnessing the early steps of a DT of the heart, considered a virtual tool intended to integrate the clinical data acquired over time from patient observation coherently and dynamically into a predictive framework (Corral-Acero et al., 2020). Whereas limited when taken by themselves, the combination of statistical (that inductively associates data) and mechanistic (that deductively integrates knowledge and the associated data) modelling is considered to show promising results: their synergy allows enhanced diagnosis, treatment guidance and prognosis assessment.

A fully developed DT of the heart would combine general population data with individual data to optimally inform clinical decisions. It would "follow the life journey of each person and harness both data collected by wearable sensors and lifestyle information that patients may register, shifting the clinical approach towards preventive healthcare" (Corral-Acero et al., 2020). However, as yet, DTs in precision cardiology have not reached wide clinical translation

for a number of scientific reasons including, notably, that they raise issues of whether they can be validated beyond the initial concept, that they suffer a lack of clinical interpretability (i.e., lack of understanding with respect to how they provide clinical predictions) and that models might fail. Furthermore, DT technology in precision cardiology is beset by additional technical challenges, such as data fragmentation and the fact that specific skills and supercomputers are required, and ethical concerns e.g., in terms of confidentiality and privacy, that we discuss further below. Nevertheless, DTs are seen as highly promising in leading to better predictions of causes of disease and to treatments for restoring health.

The expectations that DTs will actually improve health diagnosis and care point to a more theoretical issue: how accurate must the twin of a living organ be in order to still be a “twin” and have the necessary explanatory, predictive, and monitoring power? The questions raised above concerning DTs of natural systems seem equally appropriate in this context: when would the digital “twin” of the heart turn into a digital “sibling”, or even a digital “cousin”? (Hodgkinson & Elmouttie, 2019, Rasheed et al., 2020). The two main issues here that involve both descriptive and normative components are: how similar is the DT to its physical counterpart? And what level or type of similarity is a necessary condition for the DT to be (a) a twin rather than a more distant relation, and (b) useful in the relevant context of application? In short, how similar need digital twins be to their physical counterparts to be adequate and useful?

We suggested above that similarity is arguably easier to achieve in such manufactured objects whose structure and functioning we generally know than in those natural objects whose structure and functions we may still be in the process of discovering. We have identified two aspects, one ontological and one epistemological that shape the challenge of building DTs: ontological complexity (that we will henceforth label simply “complexity”) referring to the complexity of the physical object to be twinned, and epistemic transparency referring to our knowledge of the physical object to be twinned (that we will label simply “transparency”).



§3. COMPLEXITY AND TRANSPARENCY

The challenges for the DT to be an adequate twin of a physical object (i.e. having a high degree of fidelity and mirroring it well), and to monitor and predict the behaviour of the physical object increase, not only in terms of the physical counterpart's complexity, but also in terms of its epistemic transparency (i.e. how well we know and understand it). In principle, an object may combine high complexity with high transparency (be complex but well known and understood); or low complexity with low transparency (be simple but less well known and understood).

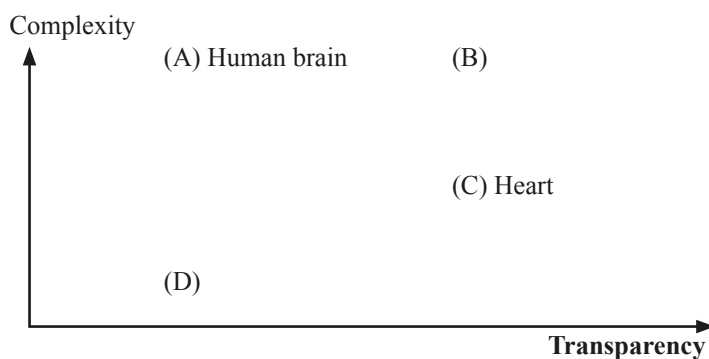


Figure 3a. Complexity & transparency: natural objects

A-D show four logically possible combinations (that may or may not be actual):

(A) The human brain has a high level of complexity (structural and functional) and a low level of transparency: we do not (yet) know or understand it well, nor do we have a well-integrated theory of its functional architecture. For now, in spite of increasing data and attempts at theories, the brain remains quite epistemically opaque.

(B) A natural object that is as complex as, but more transparent than, the human brain.

(C) The heart is notably in terms of function a simpler organ than the brain and better understood (lower complexity, higher transparency).

(D) A simple natural object that is epistemically opaque, an elusive entity.

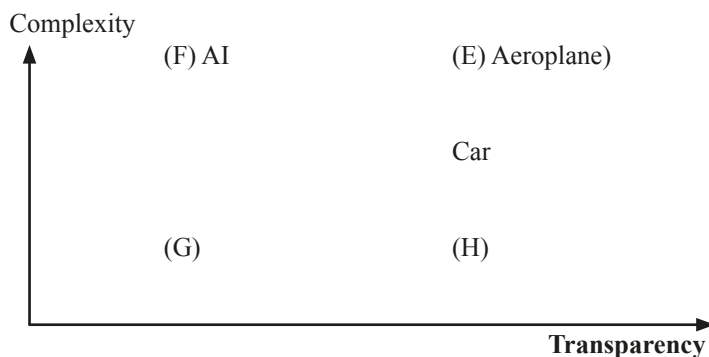


Figure 3b. Complexity & transparency: artefacts

E-H show four logically possible combinations (that may or may not be actual):

(E) An aeroplane and a car are transparent, possibly equally so, although their relative complexity is different.

(F) A complex artefact that is epistemically opaque. For example, one that is not constructed directly by humans. If, say, a very sophisticated AI would itself construct another still more sophisticated AI, this new AI could be (almost or even entirely) beyond human comprehension. That would amount to an artefact that is not only structurally complex but also dynamic.

(G) A simple artefact that is epistemically opaque. Probably, this could not be one constructed by humans (since then it would be well understood). But one could use a thought-experiment as in F: if a highly sophisticated AI created - not a very structurally complex object this time, but a profoundly different one with fundamentally new and unknown structure - then it could conceivably be (almost or even entirely) beyond human comprehension.

(H) A simpler artefact than aeroplanes and cars that is transparent.

Now, if the challenges for a DT to be an adequate twin of a physical object and to monitor and predict its behaviour increase both in terms of the physical counterpart's complexity and in terms of its epistemic transparency, then, *prima facie*, building a DT of a human brain might seem a pharaonic task. In view of the outstanding structural and functional complexity of the brain, and our presently limited integrated knowledge of its functional architecture, how could one even hope to understand it well enough to construct its virtual twin? Is the attempt to build a virtual twin of an object that combines such high levels of ontological complexity and epistemic opacity at all rational?



The brain is the most complex organ in the human body. The estimated number of nerve cells is about 86 billion, approximately the same number of glial cells, and about 10,000 synapses per neuron. For comparison, a galaxy has about 100 billion stars. The type of signal transduction³ is electro-chemical, while recent computers use electrical signals. The total length of connections is 2-3 million kilometres of fibres – this is more than the diameter of the sun at 1.4 million kilometres. But the complexity does not stop here, it is not exclusively a matter of internal structures and interactions but also external interactions, with the rest of the body and the environments in which this body lives and operates. In real life, brains do not live in isolation: brains are complex adaptive systems nested in larger complex adaptive systems. They reside in bodies. The interaction between the brain and the other bodily systems is, in reality, impossible to disentangle. This complexity in turn affects how much we can know about this organ. Our brain gets and sends information to all other bodily systems, and its state at any given point in time is determined to a substantial degree by this interaction. That the brain is a *brain-in-a-body* cannot be ignored in considering the goal to model the realistic brain (Dudai & Evers, 2014). As Dudai & Evers also point out, the brain-in-a-body at any given point in time is in fact the outcome of the individual experience accumulated over the period preceding this specific point in time. The brain is subject to important contextual and cultural epigenetic influences: it is “culture-bound” (Evers, 2015, 2020; Evers & Changeux, 2016). In trying to understand the brain’s functional architecture (a prerequisite for simulating it) one has therefore to consider the *experienced-brain-in-a-body*. Neglecting experience sets a severe limit on what we can know about the brain and thus on the construction of a virtual brain. On the other hand, taking experience into account would necessitate including real-life contexts and the brain’s dynamic connection with its many environments, a daunting task per se, specifically given that part of the real-life experience is the interaction over time with the functioning body.

The question of how these limitations may affect the adequacy and fidelity of a virtual brain to its physical counterpart must be borne in mind when assessing

³ Transduction occurs when a sensory receptor converts a type of stimulus energy (e.g. photon, sound wave) into an electro-chemical impulse that can be interpreted by the brain.



such attempts. We suggest below that with a clear philosophical framework and scientific modesty in how goals are set and results are interpreted; the project of constructing virtual brains is not necessarily pharaonic, but can be a theoretically exciting and socially beneficial scientific, technological and philosophical enterprise.

§4. THE VIRTUAL BRAIN⁴

The creation of DTs of human brains is central to the European Human Brain Project (HBP).

The HBP is one of the three FET (Future and Emerging Technology) European Flagship projects and aims to put in place a cutting-edge research infrastructure that will allow scientific and industrial researchers to advance our knowledge in the fields of neuroscience, computing, and brain-related medicine. Started in 2013, the HBP is one of the largest research projects in the world involving more than 500 scientists and engineers at over 140 universities, teaching hospitals, and research centres across Europe. To address brain complexity, the project is building research infrastructure to help advance neuroscience, medicine, computing and brain-inspired technologies - EBRAINS. The HBP is developing EBRAINS to create lasting research platforms that benefit the wider community (<https://www.humanbrainproject.eu/en/>).

The HBP's aims in creating DTs of human brains are above all clinical; contributing to progress in personalised and precision medicine for brain diseases. As stated in the project's proposal, "...a "digital twin" brain can be used clinically for patient-specific hypothesis testing and treatment discovery. It provides a qualitative advance beyond the state of the art and opens up novel avenues in research and innovation (e.g. early detection of trajectories of brain disease manifesting on different levels of brain organisation, personalised tracking of brain health and better stratification of patients) (SPECIFIC AGREEMENT 945539 – HBP SGA3, p. 232).

⁴ We owe the scientific contents in this section to Maxime Guye & Viktor Jirsa who develop virtual brain models for clinical applications in the HBP.



The search for “patient-specificity” is central in this context, because, in addition to being highly interactive (both internally and externally) and subject to important contextual and cultural epigenetic influences, the brain is pronouncedly variable: each brain is unique. This variability strongly impacts the outcome of treatments for a number of conditions, since distinct individuals may react quite differently to the same kind of intervention.

Precision medicine attempts to address this problem by customising treatment. As illustrated by the case of precision cardiology summarised above, personalised or precision medicine proposes that big data-driven mathematical models of patients can be the basis for more effective interventions that are tailored to meet their individual needs. This idea importantly motivates research into the creation of neurocomputational models of human brain networks within the HBP. Maximising specificity to the individual level (by incorporating data sets recorded non-invasively from individual human subjects’ specific brain connectivity) would enable the development of virtual brains of individual healthy subjects or patients which allows for testing clinical hypotheses (Jirsa et al., 2017).

The computational “multiscale brain connectome” (i.e. a computational and comprehensive multiscale map of neural connections in the brain) developed in the HBP attempts to demonstrate the predictive and explanatory power of mechanistic human brain network models built to generate functional brain signals that can be linked to behaviour indicators. In particular, the Virtual Brain (TVB) is a simulation engine that aims to significantly reduce the gap between modelled brain activity data and empirically measured sensor data (Sanz Leon et al., 2013; Sanz-Leon, Knock, Spiegler, & Jirsa, 2015).

Brain models can attain different levels of specificity. They can become more “personalised” when individual data are used as constraints (Jirsa, Sporns, Breakspear, Deco, & McIntosh, 2010). The brain network model in TVB integrates structural empirical data, foremost connectivity, into its otherwise computationally defined architecture. Broadly speaking, a network is a system consisting of many parts connected together to allow communication when operating together, and structural connectivity is the set of physically existing interconnecting anatomical links (axons).



In TVB, hypotheses on structural connectivity and regional variability are typically derived from the interpretation of the empirical data produced by a number of distinct sources.⁵ These data can be organised along two dimensions, spatial resolution and biological realism. It is worth noting that these two dimensions are related: the higher the required degree of biological realism, the higher the necessary spatial resolution. The reason is that maximal biological realism of an in-silico twin would require a description of the position and momentum of every molecule as present in the real counterpart brain. This is unattainable. The options, therefore, are, either to increase the spatial scale to single cells or even clusters of cells that can be more easily measured using non-invasive brain imaging and mathematically formalised; or to substitute some of this missing high-resolution information from other sources, such as post-mortem brains. But this entails a reduction of the biological realism of the description as in this case we deal with a different brain and not even a living brain.

A virtual brain comprising only subject-specific data would be very poor in spatial resolution, missing a large amount of relevant data. On the other hand, a digital high-resolution brain model would lack the personalising data features that are crucial for rendering the model useful for clinical applications. TVB research in the HBP aims to advance a hybrid approach by constraining high-resolution virtual brain models through individual data to maximise specificity to the individual level and develop high-resolution virtual brains of individual healthy subjects or patients (Jirsa et al., 2017) (Melozzi et al., 2019).⁶

Two steps are needed to personalise virtual brain models in a biologically realistic way.

The first step consists in inferring the connectome (the whole set of connectivity between the nodes) from real data by white matter tracts reconstruction from diffusion-weighted MRI (dMRI). This technique allows for quantification of the strength of the connection (e.g. number of fibres) thus leading to weighted connectivity as opposed to binary connectivity (i.e. presence or

⁵ E.g., diffusion-weighted magnetic resonance imaging (dMRI), fibre tracing (Allen atlas), histology (Big Brain), and polarized light imaging (PLI), among others.

⁶ The methods to achieve this goal are described in (Ref to our other article, forthcoming).



absence of connection). Individualisation can be improved further by adding quantitative information probing several processes (such as metabolism or neurotransmission, for example) occurring at multiple temporal and spatial scales at the nodal level. Therefore, the model can be constrained by multiple information that can be chosen depending on the goal pursued.

The second step to personalise virtual brain models consists in fitting the model parameters with empirical functional data (i.e. real brain signals) to generate accurate and meaningful simulated signals.

To the extent that TVB attempts this type of specificity derived from patients' data, it is in between two different objects. At one end, there is the traditional neural network model i.e. algorithms with no empirical input. At the other end, there is the organic brain in all its biological complexity. The hybridity of TVB is a novel aspect that plays the important role in informing the personalization of the modelling of the brain by integrating empirical information at the individual level.

Studies suggest that virtual brain models have been successfully used in pathology and clinical care (Falcon, Jirsa, & Solodkin, 2016). This can be illustrated by their application in drug-resistant focal epilepsy (Proix, Bartolomei, Guye, & Jirsa, 2017). Focal epilepsy is characterized by recurrent transient changes of cerebral activity (i.e. seizure) that lead to significant behavioural changes (i.e. ictal symptoms). The term “focal” refers to localised alterations involving networks at different spatial scales. When focal epilepsy is refractory to medical treatment, surgery is the only curative treatment. It consists in the localization and removal (when possible) of the regions responsible for seizure generation. Because of the variability between individual brains, surgery requires a comprehensive presurgical evaluation that sometimes includes invasive techniques for recording electrical activity.

In its application to epilepsy, what is known as the virtual epileptic patient (VEP) is both informed by real structural data (containing relevant pathological and patient-specific information (Besson et al., 2017)) and fitted with real electrophysiological data demonstrating features of the disease (e.g. seizures) (Jirsa et al., 2017). It was introduced in the clinic in a first cohort of patients (Proix et al., 2017). The authors suggested that this introduction was successful by showing: first, that the use of real individual connectome improves seizure



simulation; second, that patients in whom simulated data showed supplementary pathological regions (compared to the clinical hypotheses) had worse postsurgical outcomes. This was followed by an ongoing clinical trial to show the clinical value of the VEP in the presurgical evaluation of focal epilepsy patients (EPINOV <https://www.3ds.com/fr/recits/living-brain/>). Further uses of the VEP approach have paved the way for improving surgical strategies by limiting the invasiveness of surgery by following model indications (Olmi, Petkoski, Guye, Bartolomei, & Jirsa, 2019).⁷

§5. TWINS OR COUSINS? NAVIGATING BETWEEN SCYLLA AND CHARYBDIS

The description above provides a general overview of research into creating individual brain network models (derived from non-invasive data of people). Virtual brain models are arguably valuable in purely epistemic terms – to the extent that they can be expected to further our understanding of the functional architecture of the human brain – but they might also be valuable in practical terms. One of the main goals for developing them in the HBP is to advance precision medicine; specifically, to offer more precise and effective interventions in brain pathologies and to improve clinical care.

Even so, a set of interrelated issues concerning terminology that previously arose in connection with other applications, e.g., hearts, re-emerge: is it adequate to describe these brain models as “twins” when in fact their level of fidelity appears to be more limited than the term might suggest? Does the term used to refer to this tool in this context really matter? Might not the word “twin” be merely a light metaphor that shouldn’t be taken too seriously?

For reasons explained below, we suggest that the term “digital twin” is not appropriate in descriptions of HBP research on virtual brains. In those contexts, the term should neither be used metaphorically, nor should the significance of

⁷ VEP illustrates well the seamless connection posited as a necessary condition for a model to qualify as a DT (Figure 1). When simulated EEG signals are analysed both at rest and during seizures, several data features cannot be differentiated from VEP and those from the empirical recording.



its use be minimized. Our reasons for making these claims combine conceptual, ethical and social (interconnected) considerations.

If we focus on (a) the conditions that a DT must meet to qualify as such (a seamless connection in the physical-virtual bridge, a real-time data exchange enabling the “twinning” of this bridge, and a unique identifier allowing the necessary bijective connection between the digital and physical twins, cf. Figure 1) and (b) the goal of DTs in general (predictive and monitoring activity) it would seem that the term “digital twin” is appropriate to refer to computational brain models. After all, the virtual brain is minimally a synchronized connection of a computational and a physical domain intended to measure and predict certain conditions and behaviours. Notably, however, even if meeting what we identified as the main conditions for qualifying as a digital twin, it is not clear that the virtual brain can easily fit some extant descriptions of DTs as replicas of a physical object, mostly because virtual brains are not digital replicas of brains. If anything, they attempt to replicate very specific and targeted brain functions.

And yet, the understanding of a digital twin of the brain as a replica is quite likely to resonate with non-experts and is likely to play a dominant role in the perception of the general public of this tool as it relates to the brain, even if such understanding contrasts with experts’ views of what TVB is and does.

If so, conceptually and potentially ethically, reference to these brain models’ “twinness” is arguably problematic even when they meet the requirements for being a digital twin as specified above. It is conceptually problematic because, despite the fact that a certain degree of accuracy and fidelity to the physical object is necessary for the virtual model to be successfully used in the clinic, this is probably not the degree of accuracy and fidelity that the use of the term “twin” typically suggests. In view of this, one might ask: would it be better to use other “family resemblance” terms, such as “digital sibling” or “digital cousin” to refer to virtual brains? Answering this question would require the identification of the conditions necessary for a digital “twin” to turn into a digital “sibling”, or even a digital “cousin” (Hodgkinson & Elmouttie 2019, Rasheed et al. 2020). It is not clear, however, to what extent such a task would be either theoretically or practically useful in this context.



Relating TVB to twinness is also potentially ethically problematic because conceptualizing the virtual brain as a “twin” calls for addressing the ethical issue of whether cerebral twinness is desirable in this context and if so why. Indeed, it could be argued that the lack of similarity of TVB to an actual human brain might actually be a virtue⁸: its limited fidelity allows replication of the functions relevant to intervene in certain pathologies without necessarily raising the types of issues that would have to be confronted if this model were an alleged full replica of the whole brain. A full replica would require a discussion of its ontological and moral status vis-a-vis an actual human brain, and an examination of the extent to which replicas of human brains would challenge the generally recognized ontological boundary between people and objects. Therefore, it would seem that using a term that avoids assumptions of family-likeness altogether might be preferable. The question arises: Is “virtual brain” such a term?

On the one hand, it is true that the term “virtual brain” (more commonly used within HBP research) does avoid assumptions about family likeness to a certain extent. On the other hand, in practice, it is not clear that it fares much better. This is due to the fact that the term “virtual brain” suggests a specific referent (the brain) when in reality it refers to a much more limited object: it is a computational model of a particular connective or functional relationship in the brain (certainly not the whole brain in every detail, and not all the connective or functional relationships within the brain) in order to understand the brain in general and to develop more personalized brain interventions in particular. This point is not meant to minimize its importance. Indeed, the tool that we know as the virtual brain includes many factors that are available and, importantly, necessary for more reliable predictions and treatments, and it can be argued that its very simplicity in that respect is welcome. Conceptually, however, the term’s lack of precision is arguably problematic. Since the expression “virtual brain” may wrongly suggest to the non-expert that there is such a thing as a virtual whole brain-copy, it may be preferable to stick to using the more traditional and modest expression “brain model”, e.g., “computational brain model”, or maybe “targeted brain model” if one wants to emphasise that it is

⁸ A point suggested to us by Katharina Dornenzweig in discussion.



not a model of the whole brain. Still, this terminology is arguably so generic that it can apply to almost anything, including the unicellular brain, and can as such be unpractical, even borderline vacuous.⁹ From that perspective, a terminology is needed that captures the aforementioned criteria, use of individual structural imaging data and the capacity of generating and reproducing individual functional brain imaging data. Such capacity is suggested by the term “virtual” as this expression is commonly used in equivalent technical applications with this property such as virtual reality, etc. Thus the challenge is to choose terminology that is both informative (not vacuous) and avoids or mitigates confusions.

§6. CONCLUSION: WHY CONCEPTS MATTER

If the analysis above is correct, we are in the midst of developing and applying a technology that promises to revolutionize how brain disorders are diagnosed and treated, and yet we are questioning the terms used to refer to such technology. Does this really matter?

Using a fundamental neuroethics approach, we have tried to show that it does. Fundamental neuroethics (Evers, 2007, 2009, forthcoming) gives a key role to conceptual analysis in clarifying fundamental (and often unexamined) scientific and philosophical notions used in research (e.g., brain model, consciousness, human identity, etc.) and in exploring issues such as how neuroscientific knowledge is constructed, what its underlying assumptions are and how they are justified, how results may be interpreted, and why or how empirical knowledge of the brain can be relevant to philosophical, social, and ethical concerns (Farisco, Salles, & Evers, 2018; Salles & Evers, 2017). The general relevance of examining these issues in relation to the development of digital twins of the human brain cannot be minimized.

In particular, concepts and the choice of terminology matter for a number of reasons. To begin with, terms shape conceptualisations, and conceptualisations are not innocuous; the use of concepts and the meaning we attach to

⁹ As suggested by Viktor Jirsa in discussion.



them may carry considerable normative as well as theoretical weight, and can accordingly have important social consequences.

Notably, unclear conceptualisations increase the risk of hype, whether in the form of inspiring unrealistic expectations or unfounded worries. Conceptualisations can also be inherently normative and suggest values either implicitly or explicitly. These values may be more hidden if the concepts are unclear which in turn can make their social consequences more insidious. History offers numerous illustrations from normative as well as scientific discourses of how language and the meaning assigned to terms influence the contexts in which they are used. The meanings assigned to concepts may change (and sometimes they are changed deliberately) to mirror or drive social changes, as for example, in the quests to abolish or reduce racism, or misogyny. To illustrate, in Sweden, non-white or female humans are not conceptualised in the same way in the 21st century as they were in the 19th century, when white male superiority was considered to be a biological fact by the predominantly racist and misogynistic scientific community of that time (attitudes that reflected those dominating society as a whole). Whence the need for conceptual clarity to bring inherent norms to the surface then to decide how to deal with them.

Conceptual clarity is both intrinsically and instrumentally valuable. Since one of the main concerns of science (not least in this area) is to further human understanding as well as well-being, we need conceptual clarity to understand the human problems that science aims to solve, how scientists are framing them in their search for solutions, and how such framing shapes their findings. Only a certain level of conceptual clarity will allow the different societies to assess the proposed technologies, ask the right questions, and make the right decisions.

In the case of virtual brains, in the relative absence of a deeper and more integrated understanding of the brain and of which of its functions are replicated and why, the terminology and suggested conceptualisation can be misleading. The fact that some research shows that lay groups find neuroscientific terminology particularly compelling (Weisberg, Keil, Goodstein, Rawson, & Gray, 2008) should provide an additional reason to be cautious. Misleading conceptualisations will distort citizens' perceptions of neuroscience and emerging neurotechnologies, shape people's attitudes, reactions, and willingness to use the technology, and ultimately hinder the promotion of trust that is required



for productive evaluation and public acceptance and support of science. In this sense, we can say that conceptual clarity (or lack thereof) has implications at the micro as well as the macro levels.

A clear scientific theoretical structure, conceptual clarity and transparency regarding the methods and goals of this technological development are necessary prerequisites in order to make the project of constructing “virtual brains” a theoretically promising and socially beneficial scientific, technological and philosophical enterprise.

ACKNOWLEDGEMENTS

We thank Viktor Jirsa, Maxime Guye, and our colleagues at CRB for valuable comments to an earlier version of this manuscript and the reviewers for useful editorial suggestions.

This project/research has received funding from the European Union’s Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3).

REFERENCES

- Barricelli, B. R., Casiraghi, E., Fogly, D. (2019). A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. *IEEE Access*, 7, 167653-167671.
- Besson, P., Bandt, S. K., Proix, T., Lagarde, S., Jirsa, V. K., Ranjeva, J. P., ... Guye, M. (2017). Anatomic consistencies across epilepsies: a stereotactic-EEG informed high-resolution structural connectivity study. *Brain*, 140(10), 2639-2652. DOI: <https://doi.org/10.1093/brain/awx181>
- Bruynseels, K., de Sio, F., & van den Hoven, J. (2018). Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm. *Frontiers in Genetics*, 9.



- Corral-Acero, J., Margara, F., Marciniak, M., Rodero, C., Loncaric, F., Feng, Y., ... Lamata, P. (2020). The 'Digital Twin' to enable the vision of precision cardiology. *Eur Heart J*. DOI: <https://doi.org/10.1093/eurheartj/ehaa159>
- Dudai, Y., & Evers, K. (2014). To simulate or not to simulate: what are the questions? *Neuron*, 84(2), 254-261. DOI: <https://doi.org/10.1016/j.neuron.2014.09.031>
- Escorsa, E. (2018). Digital Twins: A Glimpse at the Main Patented Developments.
- Evers, K. (2007). Towards a philosophy for neuroethics. An informed materialist view of the brain might help to develop theoretical frameworks for applied neuroethics. *EMBO Rep*, 8 Spec No, S48-51. doi:10.1038/sj.embor.7401014
- Evers, K. (2009). *Neuroetique. Quand la matière s'éveille*. Paris: Odile Jacob.
- Evers, K. (2015). Can we be epigenetically proactive? In W. Metzinger T., J. (Ed.), *Open Mind: Philosophy and the mind sciences in the 21st century*. Cambridge, MA: MIT Press.
- Evers, K. (2020). The Culture Bound Brain: Epigenetic Proaction Revisited. *Theoria*, <https://doi.org/10.1111/theo.12264>
- Evers, K. (forthcoming). Fundamental neuroethics. In M. Farisco (Ed.), *Neuroethics and cultural diversity*: ISTE-Wiley.
- Evers, K., & Changeux, J. P. (2016). Proactive epigenesis and ethical innovation: A neuronal hypothesis for the genesis of ethical rules. *EMBO Rep*, 17(10), 1361-1364. DOI: <https://doi.org/10.15252/embr.201642783>
- Falcon, M. I., Jirsa, V., & Solodkin, A. (2016). A new neuroinformatics approach to personalized medicine in neurology: The Virtual Brain. *Curr Opin Neurol*, 29(4), 429-436. DOI: <https://doi.org/10.1097/WCO.0000000000000344>
- Farisco, M., Salles, A., & Evers, K. (2018). Neuroethics: A Conceptual Approach. *Camb Q Healthc Ethics*, 27(4), 717-727. DOI: <https://doi.org/10.1017/S0963180118000208>
- Grieves, M. (2014). Digital Twin: Manufacturing excellence through virtual factory replication.



- Grieves, M., & Vickers, J. (2017). Digital Twin: Mitigting Unpredictable, Undesirable Emergent Behavior in Complex Systems. In F. Kahlen, Flumerfelt, S., Alves, A. (Ed.), *Transdisciplinary Perspectives on Complex Systems* (pp. 85-113). Switzerland: Springer International Publishing.
- Hodgkinson, J., & Emouttie, M. (2020). Cousins, Siblings and Twins: A Review of the Geological Model's Place in the Digital Mine. *Resources*, 9(24).
- Jirsa, V. K., Proix, T., Perdikis, D., Woodman, M. M., Wang, H., Gonzalez-Martinez, J., ... Bartolomei, F. (2017). The Virtual Epileptic Patient: Individualized whole-brain models of epilepsy spread. *Neuroimage*, 145(Pt B), 377-388. DOI: <https://doi.org/10.1016/j.neuroimage.2016.04.049>
- Jirsa, V. K., Sporns, O., Breakspear, M., Deco, G., & McIntosh, A. R. (2010). Towards the virtual brain: network modeling of the intact and the damaged brain. *Arch Ital Biol*, 148(3), 189-205. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/21175008>
- Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterizing the Digital Twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, 29, 36-52.
- Kiritsis, D. (2011). Closed-loop PLM for intelligent products in the era of the Internet of Things. *Comput.-Aided Des*, 43, 479-501.
- Machery, E. (2012). Why I stopped worrying about the definition of life... and why you should as well. *Synthese*, 185, 145-154.
- Macklem, P. T., & Seely, A. (2010). Towards a definition of life. *Perspect Biol Med*, 53(3), 330-340. DOI: <https://doi.org/10.1353/pbm.0.0167>
- Melozzi, F., Bergmann, E., Harris, J. A., Kahn, I., Jirsa, V., & Bernard, C. (2019). Individual structural features constrain the mouse functional connectome. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1906694116
- Olmi, S., Petkoski, S., Guye, M., Bartolomei, F., & Jirsa, V. (2019). Controlling seizure propagation in large-scale brain networks. *PLoS Comput Biol*, 15(2), e1006805. DOI: <https://doi.org/10.1371/journal.pcbi.1006805>
- Proix, T., Bartolomei, F., Guye, M., & Jirsa, V. K. (2017). Individual brain structure and modelling predict seizure propagation. *Brain*, 140(3), 641-654. DOI: <https://doi.org/10.1093/brain/awx004>



- Rasheed, A., San, O., & Kvamsdal, T. (2020). Digital Twin: Values, Challenges and Enablers from a Modeling Perspective. *IEEE Access*, 8, 21980-22012.
- Rosen, R., Von Wichert, G., Lo, G., & Bettenhausen, K. D. (2015). About the importance of autonomy and digital twins for the future of manufacturing. *IFAC-PapersOnLine*, 48(3), 567-572.
- Salles, A., & Evers, K. (2017). Social neuroscience and Neuroethics: A Fruitful Synergy. In A. Ibanez, Sedeno, L., Garcia, A. (Ed.), *Social Neuroscience and Social Science: The Missing Link* (pp. 531-546): Springer.
- Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. *AJOB Neuroscience*, 11(2), 88-95.
- Sanz Leon, P., Knock, S. A., Woodman, M. M., Domide, L., Mersmann, J., McIntosh, A. R., & Jirsa, V. (2013). The Virtual Brain: a simulator of primate brain network dynamics. *Front Neuroinform*, 7, 10. doi:10.3389/fninf.2013.00010
- Sanz-Leon, P., Knock, S. A., Spiegler, A., & Jirsa, V. K. (2015). Mathematical framework for large-scale brain network modeling in The Virtual Brain. *Neuroimage*, 111, 385-430. doi:10.1016/j.neuroimage.2015.01.002
- Tao, F., SUI, F., QI, Q., ZHANG, M., Song, B., Guo, Z., ... A., N. (2019). Digital Twin-Driven Product Design Framework. *International Journal of Production Research*, 57(12), 3935-3953.
- Tao, F., Zhang, H., Liu, Ang, Nee AYC., (2019). Digital Twin and Industry: State of the Art. *IEEE Transactions on Industrial Informatics*, 15(4), 2405-2415.
- Vico, G. (1710/1988). *On the Most Ancient Wisdoms of the Italians Unearthed from the Origins of the Latin Language. The Book of Metaphysics*. Ithaca: Cornell University Press.
- Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *J Cogn Neurosci*, 20(3), 470-477. DOI: <https://doi.org/10.1162/jocn.2008.20040>



