

Rethinking the Right to Freedom of Thought: A Multidisciplinary Analysis

Sjors Ligthart^{*} , Christoph Bublitz^{**}, Thomas Douglas^{***},
Lisa Forsberg[#] and Gerben Meynen^{##}

ABSTRACT

In recent years, there has been increased academic interest in the human right to freedom of thought (RFoT). Scholars from various disciplines are currently debating the content and scope of this right. In his annual thematic report of 2021, the United Nations Special Rapporteur on Freedom of Religion or Belief paid explicit and comprehensive attention to the RFoT, encouraging further clarification of the content and scope of the right. This paper aims to contribute to this end, setting the stage for further research, by offering a multidisciplinary analysis of the RFoT's scope, relation to other rights, practical significance and moral foundations.

KEYWORDS: Article 18 ICCPR, Article 9 ECHR, freedom of expression, mental integrity, mental privacy, moral foundations, (neuro)technology

1. INTRODUCTION

In recent years, there has been increased academic interest in the right to freedom of thought (RFoT), guaranteed *inter alia* by Article 18(1) of the International Covenant on Civil and Political Rights (ICCPR) and by most regional instruments, such as Article 9(1) of the European Convention on Human Rights. Scholars from various disciplines, including law, philosophy and psychiatry, are currently discussing the content and scope of this right.¹ Moreover, in his annual thematic report of 2021, the United Nations (UN) Special Rapporteur on Freedom of Religion

^{*} Postdoctoral researcher at Utrecht University and assistant Professor of Criminal Law at Tilburg University, the Netherlands. Email: s.l.t.j.ligthart@tilburguniversity.edu.

^{**} Legal scholar at the Faculty of Law, Universität Hamburg, Germany. Email: christoph.bublitz@uni-hamburg.de.

^{***} Professor of Applied Philosophy at the Oxford Uehiro Centre for Practical Ethics and Senior Research Fellow at Jesus College, University of Oxford, UK. Email: thomas.douglas@philosophy.ox.ac.uk.

[#] Uehiro Research Fellow, Oxford Uehiro Centre for Practical Ethics and Fulford Junior Research Fellow, Somerville College, University of Oxford, UK. Email: lisa.forsberg@philosophy.ox.ac.uk.

^{##} Professor of Forensic Psychiatry at Utrecht University and Professor of Ethics, in particular Bioethics, at VU University Amsterdam, the Netherlands. Email: g.meynen@uu.nl.

¹ See, for example, Blitz and Bublitz (eds), *The Law and Ethics of Freedom of Thought*, Volume 1 (2022); Alegre, *Freedom to Think* (2022); O'Callaghan and Shiner, 'The Right to Freedom of Thought in the European Convention of Human Rights' (2021) 8 *The European Journal of Clinical Pharmacology* 112; Ligthart, 'Freedom of Thought in Europe' (2020) 7 *The Journal of Law and the Biosciences* 1; Alegre, 'Rethinking Freedom of Thought for the 21st Century', 2017 (3) *European Human Rights Law Review* 221; McCarthy-Jones, 'The Autonomous Mind' (2019) 2 *Frontiers in Artificial Intelligence* 19.

Received: May 20, 2022. Accepted: July 20, 2022

© The Author(s) [2022]. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

or Belief, Ahmed Shaheed, paid explicit and comprehensive attention to the RFoT. The Rapporteur mentions several possible violations of the RFoT which call for urgent attention from policymakers. These include psychological torture, re-education of political prisoners, involuntary treatment for mental disorders and non-consensual conversion practices. He also draws special attention to new challenges created by modern digital and neuroscientific technologies. These, he notes, 'pose dilemmas about how to protect mental privacy, how to protect thoughts from impermissible manipulation and modification, and how to prevent these technologies from being used and abused to punish real or inferred thoughts, rather than an individual's conduct'.²

These challenges relate to a broader debate on how human rights should protect the mental aspects of our lives. Traditionally, human rights, taken together, have provided greater protection to human bodies and to visible objects in the outside world than to mental states such as emotions and intentions. Technological developments have motivated a reconsideration of this arguable imbalance in legal protection.³ Several actors, such as the Committee on Bioethics of the Council of Europe and the International Bioethics Committee of United Nations Educational, Scientific and Cultural Organisation (UNESCO), are currently examining whether the normative issues raised by emerging technologies can be sufficiently addressed by existing human rights or whether new rights are needed.⁴ In this regard, particular attention is paid to the existing RFoT.⁵

The historical development of Article 18 of the Universal Declaration of Human Rights (UDHR) and Article 18 ICCPR is almost exclusively focused on 'religion' and, to a lesser degree, on 'conscience'. Recent literature and the reports for the UN, the Council of Europe and UNESCO, bring increased salience to the third element: 'thought'. This is important, as the RFoT is as yet poorly defined in theory and in practice and faces serious challenges. To better address these challenges, the UN Special Rapporteur has encouraged clarification of the content and scope of the RFoT both by individual States and through a General Comment of the Human Rights Committee.⁶ This paper aims to contribute to this end and to set the stage for further research by offering a multidisciplinary analysis of the RFoT's scope, relation to other rights, practical significance and moral foundations. The focus of this contribution is on Article 18 ICCPR and 9 ECHR, but the arguments may equally apply to the RFoT pursuant to other provisions.

2. SCOPE OF THE RIGHT

The RFoT consists of at least three substantive negative elements: (1) that persons are not forced to *reveal* their thoughts; (2) that persons are not *sanctioned* for their thoughts and (3) that thoughts are not impermissibly *altered*. In addition, the RFoT may include a positive element: (4) that States *foster* an enabling environment for freedom of thought.⁷ Importantly, on the dominant current understanding, and in accordance with the *travaux préparatoires* and the

² Shaheed, UN Special Rapporteur on Freedom of Religion or Belief, Report on the Freedom of Thought, 5 October 2021, A/76/380, at 94.

³ Bublitz and Merkel, 'Crimes against Minds' (2014) 8 *Criminal Law and Philosophy* 51; Ienca and Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotechnology' (2017) 13 *LSP* 1; Yuste et al., 'Four Ethical Priorities for Neurotechnologies and AI' (2017) *Nature* 551; Ligthart et al., 'Forensic Brain-Reading and Mental Privacy in European Human Rights Law' (2021) 14 *Neuroethics* 191; Bublitz, 'Novel Neurorights: From Nonsense to Substance' (2022) 15 *Neuroethics* 7.

⁴ Ienca, Common human rights challenges raised by different applications of neurotechnologies in the biomedical field (2021); Report of the International Bioethics Committee of UNESCO, Ethical Issues of Neurotechnology, SHS/BIO/IBC28/2021/3Rev. (15 December 2021).

⁵ *Supra* n 1.

⁶ Shaheed, *supra* n 2 at 96.

⁷ Shaheed, *supra* n 2 at 25–28, 40. Slightly different: Bublitz, 'Cognitive Liberty or the International Human Right to Freedom of Thought' in Clausen and Levy (eds), *Handbook of Neuroethics* (2014).

analysis of the Human Rights Committee, the negative dimension of the RFoT is an absolute right. Infringements cannot be justified. As Shaheed puts it: ‘unlike *forum externum* (external realm) freedoms that are subject to State limitations, if prescribed by law and necessary to protect public safety, order, health or morals, or the rights of others, States legally cannot ever interfere with freedom of thought.’⁸

In recent debates, a broad understanding of the notion of ‘thought’ has been suggested. The Special Rapporteur observes that ‘[f]reedom of thought is simultaneously “profound and far-reaching”. It protects thoughts on “all matters”, whether about conscience, religion or belief or other topics’.⁹ This understanding of the RFoT echoes a previous statement by the Human Rights Committee¹⁰ and is in line with some recent contributions in the academic literature.¹¹ By contrast, some other scholars take a more restrictive view. For example, according to Partsch, ‘thought’ includes only political and social thought, while ‘conscience’ includes only moral attitudes and moral beliefs.¹² De Jong understands ‘thought’ as referring primarily to personal views that have a major impact on a person’s way of living, such as views concerning philosophy, politics and science.¹³ And Evans equates ‘thought’ with ‘conscience’. Since freedom of thought is an absolute right, Evans warns that, if the right were interpreted as having a broad scope, this could have profound legal and practical implications, as advertising and various other means of changing people’s mental states may then be covered by the right and therefore be prohibited in absolute terms by Article 18(1) ICCPR.¹⁴ This worry was also noted by the drafters of the Covenant, who shared the understanding that ordinary forms of persuasion should not fall under its ambit.¹⁵

A broad understanding of ‘thought’ may indeed bring a wide range of interventions under the scope of the RFoT. One could think of neurotechnologies (such as memory detection in criminal justice and neuroimaging to infer suicidal thoughts) and digital techniques that influence emotions, desires and opinions (such as microtargeting in advertisement and political campaigns).¹⁶ Neurotechnologies might in the future have the potential to identify thoughts based on brain activity and to modify or manipulate thoughts through brain stimulation or optogenetics. Whereas most neurotechnologies have so far only been used in medicine or research settings, some applications, such as portable electroencephalograms that monitor brain activity, are also available on the consumer market.¹⁷ Consumer use of neurotechnology may well increase when companies, such as Elon Musk’s *Neuralink*, succeed in establishing a direct interface between the brain and computers. Novel neurotechnologies also have (potential) applications in criminal justice, such as risk assessment using neuroimaging.¹⁸ Furthermore, digital technologies may enable states and private companies to monitor a considerable part of our daily activities, revealing information about emotions, sexuality, political preference and

⁸ Shaheed, *supra* n 2 at 8; UN Doc. A/2929, at 51; General Comment No. 22, CCPR/C/21/Rev.1/Add.4, at 1.

⁹ Shaheed, *supra* n 2 at 91.

¹⁰ General Comment No. 22, CCPR/C/21/Rev.1/Add.4, at 1.

¹¹ Bublitz, ‘Freedom of Thought as an International Human Right’ in Blitz and Bublitz (eds), *The Law and Ethics of Freedom of Thought*, Volume 1 (2022) at 73–74; McCarthy-Jones, ‘Freedom of Thought: Who, What, and Why?’ in Blitz and Bublitz *ibid.*; Alegre, *supra* n 1.

¹² Partsch, ‘Freedom of Conscience and Expression, and Political Freedoms’ in Henkin (ed.), *The International Bill of Rights: The Covenant on Civil and Political Rights* (1981) at 213–214.

¹³ De Jong, *The Freedom of Thought, Conscience and Religion or belief in the United Nations* (2000) at 23–33. Cf. Ligthart *supra* n 1.

¹⁴ Evans, *Religious Liberty and International Law in Europe* (1997) at 289, 295.

¹⁵ UN Secretary General, Annotations to the draft of the ICCPR, UN. Doc A/2929 (1955) at 110; Cf. Nowak, *UN Covenant on Civil and Political Rights: Commentary* 2nd ed. (2005) at 412.

¹⁶ Shaheed *supra* n 2 at 68–82.

¹⁷ Ienca, Haselager and Emanuel, ‘Brain Leaks and Consumer Terotechnology’ (2018) 36 *Nature Biotechnology* 805.

¹⁸ For example: Zijlmans et al., ‘The predictive value of neurobiological measures for recidivism in delinquent male young adults’ (2021) 46 *J of Psychiatry and Neuroscience* 271.

criminal proclivities.¹⁹ As our online lives further expand, our vulnerability to digital means of mental influence will likely also increase.

Developments in both digital and neurotechnologies strengthen the need to clarify the scope of the RFoT.²⁰ But how may this be achieved? At present, the notion of ‘thought’ not only lacks legal precision but also scientific and philosophical consensus. This issue is explicitly raised by the UN Special Rapporteur on Freedom of Religion or Belief, who emphasizes that different stakeholders, including neuroscientists, psychologists and linguists, use different definitions that often diverge within one single discipline.²¹ Though lawmakers must be attentive to these disagreements, we contend they should not rely on science to deliver a conception of ‘thought’ that is fit for legal purposes. Different scientific fields have developed different conceptions of thought, with different scientific purposes in mind, so these conceptual differences across the sciences can be expected to persist. Moreover, the scientific purposes served by scientific conceptions of thought may not be relevant to the law. Thus, rather than waiting for scientific consensus to emerge, the law should stipulate a certain *legal meaning* of ‘thought’ with an eye to ethical and legal considerations. This is a common approach in the law, where subjective mental features are often objectively defined—sometimes in ways that conflict with non-legal definitions—to enable effective enforcement in legal practice and the development of distinctive doctrines in legal scholarship. Well-known examples are mental elements of a crime, such as intent, recklessness and negligence. A similar approach is used to define notions and concepts in human rights law. Think, for example, of ‘fair trial’, ‘torture’ and ‘inhuman and degrading treatment.’²² Likewise, to qualify as a ‘belief’ in the meaning of Article 18 ICCPR, the law requires some specified conditions to be met: beliefs are understood to be formed by a system of principles or philosophical considerations of life.²³ Under Article 9 ECHR, it is generally accepted that for a ‘belief’ to be protected, it must attain a certain level of cogency, seriousness, cohesion and importance.²⁴ Similarly, ‘religion’ and ‘conscience’ in Article 18 ICCPR and Article 9 ECHR are defined in a narrow way that does not clearly correspond to their uses outside the law.²⁵

In our view, a similar *legal* approach should be followed to clarify the scope of ‘thought’. Of course, science should inform the law by providing evidence on how different types of interventions could access or interfere with different mental states. Lawyers and ethicists can then consider which applications of these interventions should count as infringements of the right by taking account of the theoretical rationales of the RFoT. The role of science here, then, is to provide evidence on the effects of different interventions and not to provide us with the basic legal concepts.

Scientific research may be particularly helpful in determining the extent to which, and manner in which, different interventions interfere with people’s inner thoughts. For example, whereas neurotechnologies have the ability to *directly*—that is, without psychological mediation—alter a person’s thoughts, certain digital technologies alter thoughts through *indirect* mechanisms that involve psychological processing, such as perceiving images or reading information. There is a debate on the ethical and legal significance of the distinction between direct and indirect forms

¹⁹ Shaheed *supra* n 2 at 68–75.

²⁰ Bublitz, ‘Freedom of Thought in the Age of Neuroscience’, (2014) 100 *ARSP* 1; McCarthy-Jones, *supra* n 1; O’Callaghan and Shiner, *supra* n 1.

²¹ Shaheed, *supra* n 2 at 15–16.

²² Cf. Bjorge, *Domestic Application of the ECHR: Courts as Faithful Trustees* (2015) Ch. 7.

²³ Taylor, *A Commentary on the International Covenant on Civil and Political Right* (2021) at 505–507.

²⁴ Harris, O’Boyle, and Warbrick, *Law of the European Convention on Human Rights*, 4th edn (2018) at 572.

²⁵ Further delineations might be necessary with respect to freedom of opinion pursuant to Article 19(1) ICCPR. ‘Opinion’ usually denotes value judgements, a subset of thought. The Human Rights Committee emphasizes the closeness between ‘opinion’ and ‘thought’ and stresses that freedom of opinion ‘permits no exception or restriction’ (CCPR/C/GC/34, at 9). The level of protection afforded to both is thus the same, which may make detailed delineations largely dispensable.

of mental influence.²⁶ This issue may prove crucial to determining which interventions infringe the RFoT.²⁷ The UN Special Rapporteur touches upon these debates too by distinguishing three forms of interference: coercion, modification and manipulation. Though these concepts require further clarification, this is a helpful first approximation and one that highlights the important point that infringements of the RFoT are not limited to cases of coercion. Manipulation and undue influence can, for example, plausibly infringe the right as well.

More generally, the scope of the RFoT is yet to be established. By performing three tasks, this paper lays some of the groundwork necessary for future academic inquiry and exploration to ultimately arrive at a clear and defensible specification of the RFoT's scope. Sections 3–5 situate the RFoT in relation to nearby rights, focussing on the three rights that have been considered most relevant in debates on the human rights protection of the mind—namely, rights to freedom of expression and opinion, to mental integrity and to mental privacy. Section 6 examines, as a case study, one type of intervention that involves a *prima facie* interference with the RFoT, at least when the right is broadly defined: involuntary treatment in psychiatry. Finally, Section 7 looks at the moral foundations of the RFoT since clarifying these may help to set the right's outer contours and to resolve some of its inner tensions.

3. FREEDOM OF THOUGHT AND FREEDOM OF EXPRESSION

Freedom of thought is historically, substantively and systematically linked to freedom of expression. But the relationship between these rights needs clarification. When held by the same person, the two rights support and strengthen each other. They are sometimes even referred to as if they were identical, but the structure of the international instruments clearly indicates that they are two distinct rights: Article 18(1) ICCPR protects the forming and having of thoughts within the *forum internum*, while Article 19(2) ICCPR protects their expression; the manifestation of thoughts in the *forum externum*.²⁸ Many issues that are publicly addressed as matters of freedom of thought ('intellectual freedom' and 'cancel culture') legally pertain to Article 19 ICCPR. Moreover, the freedom of expression of a speaker relates to the freedom of thought of (potential) listeners in at least two ways.

In the first, they are mutually reinforcing: a person's freedom of thought requires receiving expressions from others. Thinking and reasoning are not solitary activities but gain sustenance from dialogical exchanges with others. Free thinking requires questioning and interlocutors. In this way, censoring speakers or the press may affect the freedom of thought of actual and potential listeners. However, as Article 19 ICCPR pertains, explicitly and specifically, to freedom of the press and to freely receiving and imparting information, opinion and ideas, these issues are, from a systematic perspective, generally best addressed as matters of the right to freedom of expression.²⁹

In the second, one's expressions could negatively interfere with the freedom of thought of others. Examples are microtargeted marketing and psychologically tailored advertising, which work not only via the content it transmits but also by playing on individual desires and weaknesses of recipients. Such forms of influence are subtle, by themselves often ineffective and trivial and they are elements of ordinary forms of communication. However, modern communication technologies allow them to be individualized and more finely adjusted. Such developments

²⁶ Focquaert and Schermer, 'Moral Enhancement: Do Means Matter Morally?' (2015) 8 *Neuroethics* 139; Douglas, 'Neural and Environmental Modulation of Motivation' in Birks and Douglas (eds), *Treatment for Crime* (2018); Levy, 'Cognitive Enhancement in Neurointerventions and the Law' in Vincent, Nadelhoffer, and McCay (eds), *Neurointerventions and the Law* (2020); Bublitz, 'Why Means Matter' in Vincent, Nadelhoffer and McCay (2020), *ibid*.

²⁷ Bublitz, *supra* n 11. Cf. Taylor, *supra* n 23 at 505–507.

²⁸ Taylor, *supra* n 23 at 501; Evans, *supra* n 14 at 285.

²⁹ Shaheed, *supra* n 2 at 19.

require the law to draw a line between permissible and undue forms of influence.³⁰ This is a difficult task that may be best addressed in a context-specific way rather than at the level of human rights. Nevertheless, some guiding criteria may be found in the jurisprudence of the ECtHR on permissible and impermissible forms of proselytism. The latter include ‘exerting improper pressure on people in distress or in need’ as well as ‘the use of violence or brainwashing.’³¹ By contrast, merely discussing beliefs and teachings with others is not improper. At any rate, it is important for the law to recognize tensions between the freedoms of thought and expression and the open challenge to reconcile them.

Finally, it remains to be seen whether there is room for the positive obligation to foster an ‘enabling environment’ for both freedom of thought and freedom of expression.³² Providing access to education is a paradigm example of an intervention that might fulfil this enabling role, but it has—rightfully—been codified in a separate norm (e.g. Article 26 UDHR). Other possible enabling interventions would include measures to prevent the development, or perpetuation, of a heavily one-sided content media landscape, perhaps due to algorithmic content curation and monopolies, which does not leave much room for plurality of views and opinions.³³

4. FREEDOM OF THOUGHT AND MENTAL INTEGRITY

Another right that closely relates to the RFoT is the right to mental integrity,³⁴ guaranteed, for example, by Article 3(1) of the Charter of Fundamental Rights of the European Union (CFR), Article 8 ECHR and Article 5(1) of the American Convention on Human Rights.³⁵ Although the exact scope of this right is yet to be determined, it is typically taken to include at least a negative right against (certain kinds of) non-consensual interference with a person’s mind.³⁶ As yet, the relationship between the RFoT and the right to mental integrity has largely been left unexplored. It deserves closer attention. Although some scholars suggest that freedom of thought and mental integrity are interchangeable terms,³⁷ in human rights law, it is clear that these are two separate rights, with distinctive scopes, significantly different limitations and probably diverging moral justifications. It is true that, as with the RFoT, the content and scope of the right to mental integrity have not yet been fully clarified in case law.³⁸ But the contours of this right have been sketched by commentators and in the case law of human rights courts. For example, in the jurisprudence of the European Court of Human Rights, the right to mental integrity (sometimes also referred to as a right to ‘psychological integrity’) complements the right to bodily integrity under Article 8 ECHR, encompassing setbacks to mental health as well as failures to treat mental disorders in accordance with positive obligations that could be derived from the right.³⁹ This reasoning is supported by Article 3(1) CFR, which designates both bodily

³⁰ Cf. Evans, *supra* n 14.

³¹ *Kokkinakis v Greece*, Application no. 14307/88, Merits and Just Satisfaction, 25 May 1993, at para 48.

³² Shaheed, *supra* n 2 at 40–46.

³³ *Id.*, at 67.

³⁴ Bublitz, *supra* n 11 at 93; Shaheed, *supra* n 2 at 93; Mendlow, ‘Why Is It Wrong to Punish Thought?’ (2018) 127 *The Yale Law Journal* at 2376.

³⁵ See also General Comment No. 35 (2014) on liberty and security of person (CCPR/C/GC/35), at 3, 9.

³⁶ See, for example, Douglas and Forsberg, ‘Three Rationales for a Legal Right to Mental Integrity’ in Ligthart et al. (eds), *Neurolaw* (2021); Bublitz, ‘The Nascent Right to Psychological Integrity and Mental Self-Determination’ in Von Arnault, Von der Decken and Susi (eds), *The Cambridge Handbook of New Human Rights* (2020); Michalowski, ‘Critical Reflections on the Need for a Right to Mental Self-Determination’ in Von Arnault, Von der Decken and M. Susi, *ibid* (2020).

³⁷ Lavazza, ‘Freedom of Thought and Mental Integrity’ (2018) 12 *Frontiers in Neuroscience* 82; O’Callaghan and Shiner, *supra* n 1.

³⁸ Shaheed, *supra* n 2 at 39.

³⁹ De Vries, ‘Right to Respect for Private and Family Life’ in Van Dijk et al. (eds), *Theory and Practice of the European Convention on Human Rights* (2018) at 690–692; Bublitz *supra* n 36; *Guide on Article 8 of the European Convention on Human Rights*, August 2021, at 129.

and mental integrity as forms of the integrity of the person, aiming to provide holistic protection, which is not only centred on the body.

Less clear is the extent to which mental integrity covers different interferences with a person's mind that do not threaten the person's mental health. The European Union Network of Independent Experts on Fundamental Rights considers that Article 3 CFR covers a broad range of more or less serious interferences that have traditionally been covered by the right to privacy, including non-consensual psychiatric interventions and brainwashing.⁴⁰ Under Article 8 ECHR, mental and physical integrity together embrace multiple aspects of a person's identity, such as gender identity and sexual orientation.⁴¹ In addition, like the RFoT, the right to mental integrity is currently under debate in view of emerging (neuro)technologies that could interfere with mental states.⁴² In that regard, the Committee on Bioethics of the Council of Europe has recently defined mental integrity as the 'ability of individuals to exercise control over what happens to them with regard to (...) their mental state, and the related personal data.'⁴³ The ability to *exercise control* over mental states seems to fit within the conception of mental integrity in Article 3 CFR, while data protection issues are better placed under privacy rights such as covered by Article 8 ECHR.

Interestingly, the recent report by the UN Special Rapporteur considers a variety of issues as (potential) threats to the RFoT, which would also plausibly infringe a right to mental integrity. Examples are the electrical stimulation of brain functions in the treatment of depression, non-consensual medication in treating mental health conditions and non-consensual conversion practices.⁴⁴ To what extent the two rights exactly overlap is as yet unclear since a considerable part of the content and scope of these rights is still to be clarified in the literature and case law. Nonetheless, insofar as the RFoT and the right to mental integrity (appear to) protect against similar interferences with the mind, criteria to delimit their scopes are required. One relevant aspect could be the *lex specialis* rule. In a common understanding, 'thought' is more specific than 'mental'. This may suggest a need to draw distinctions between thoughts and other mental states. Another important aspect here is that, unlike the RFoT, the right to mental integrity is typically a qualified right, allowing for certain interferences with a person's mental states and capacities. The absolute protection by the RFoT goes beyond the qualified protection provided by the right to mental integrity.

A similar issue has arisen with respect to the absolute prohibition of torture, inhuman and degrading treatment, which covers cases that also fall under the right to bodily and mental integrity. In that case, scholars and courts have identified considerations that bring an intervention within the scope of the absolute protection.⁴⁵ A common formula is that an interference must rise to a certain level of seriousness to trigger absolute protection. Clarifications along those lines are required for the RFoT and a right to mental integrity too. The significance of the mental state that is interfered with could be a relevant factor that may justify stronger legal protection by the RFoT compared with the right to mental integrity. Think of political preferences and sexual orientation vis a vis more trivial thoughts such as which socks to wear or what fruit to buy. Also

⁴⁰ Nowak, 'Article 3 CFR' in Commentary of the Charter of Fundamental Rights of the European Union (2006) at 36. Cf. Article 17 ICCPR, see Taylor, *supra* n 23 at 480.

⁴¹ *Bédat v Switzerland*, Application No 56925/08, Merits, 29 March 2016 at para. 72.

⁴² Ienca, *Common Human Rights Challenges Raised by Different Applications of Neurotechnologies in the Biomedical Field* (2021); Bublitz and Merkel, *supra* n 3; Douglas and Forsberg, *supra* n 36.

⁴³ DH-BIO, Strategic Action Plan on Human Rights and Technologies in Biomedicine (2020–2025) (November 2019) at 21–22.

⁴⁴ Shaheed, *supra* n 2 at 78, 80–83. 'Conversion practices' sometimes also referred to as 'sexual orientation and gender identity change efforts'. These efforts raise concerns regarding other human rights too, such as the prohibition of ill-treatment: Trispiotis and Purshouse, 'Conversion Therapy' As Degrading Treatment' (2021) 42 *Legal Studies* 104.

⁴⁵ Taylor, *supra* n 23 at 189–190, 480–481; Harris et al., *supra* n 24 at 238–239, 519–522.

relevant may be the avoidability of the interference, the magnitude of the changes that it causes to the person's mental states and the duration of its effects.

Importantly, both the RFoT and to mental integrity are currently evolving. Together, they have the potential to form the basis for a comprehensive and nuanced framework of human rights protection against a variety of interferences with mental states and capacities. However, this potential will be realised only if the two rights are developed together and in light of an understanding of the relationship between them.

5. FREEDOM OF THOUGHT AND MENTAL PRIVACY

Similar to the right to mental integrity, the right to mental privacy is currently under debate, predominantly in view of emerging neuroimaging technologies that may enable so-called 'mind-reading'—the identification of a person's thoughts, feelings or motivations.⁴⁶ For example, in a recent report for the Committee on Bioethics of the Council of Europe, Ienca argued that '[c]ollecting and processing human brain data [. . .] allows a certain degree of access to mental information even in the absence of observable behaviour, hence may challenge the privacy of the mind or mental privacy.'⁴⁷ Likewise, Shaheed highlights that, although one might think that thoughts are free before being expressed, emerging technologies are increasingly challenging this understanding, posing dilemmas about how to protect mental privacy.⁴⁸

The first thing to note here is that a right to mental privacy is entailed by the general, qualified right to privacy pursuant to Article 17 ICCPR and, in the European context, Article 8 ECHR.⁴⁹ Likewise, in the United States law, courts have recognised that the right to privacy also encompasses a right to mental privacy.⁵⁰ In addition to privacy rights, UN Special Rapporteur Shaheed emphasizes that the first element of the RFoT—that nobody should be forced to reveal their thoughts—implies that mental privacy is also 'a core attribute' of freedom of thought, which arguably includes a general right to remain silent too.⁵¹ Privacy rights pursuant to Article 17 ICCPR and 8 ECHR are typically qualified rights that allow for certain exceptions. How would this qualified protection of mental privacy relate to the absolute RFoT?

To properly address this question, we need a definition of "mental privacy". Neither the literature nor the reports for the UN, the Council of Europe and UNESCO provide clarity on this issue. For instance, some define mental privacy as 'people's right against the unconsented intrusion by third parties into their brain data as well as against the unauthorized collection of those data',⁵² while others relate mental privacy to personal identity as it is intimately and psychologically linked to who we are.⁵³ In the absence of an accepted definition, mental privacy will be understood here broadly as a right not to reveal any unexpressed mental state.⁵⁴

As previously mentioned, the "manifestation" of thoughts is covered by the right to freedom of expression. This right also has a negative aspect; the freedom *not* to express oneself. According to the General Comment on Article 19 ICCPR: 'Freedom to express one's opinion necessarily includes freedom not to express one's opinion'.⁵⁵ Likewise, Harris and collaborators note that 'Article 10 [ECHR] guarantees to some extent the negative aspect of freedom of expression, namely the right not to be compelled to express oneself. One notable example is the right to

⁴⁶ Ienca and Andorno, *supra* n 3; Ligthart et al., *supra* n 3; Bublit, *supra* n 3.

⁴⁷ Ienca, *supra* n 4, at 31. See also, International Bioethics Committee of UNESCO, *supra* n 4 at 72–77.

⁴⁸ Shaheed, *supra* n 2, at 96.

⁴⁹ Bublit, *supra* n 3, at 7; Ligthart, 'Coercive Neuroimaging, Criminal Law and Privacy' (2019) 6 *JLB* 289.

⁵⁰ *Long Beach City Employees Assn. v City of Long Beach* (1986); *Stanley v Georgia* (1969).

⁵¹ Shaheed, *supra* n 2 at 26.

⁵² Ienca, *supra* n 4 at 5.

⁵³ Wajnerman Paz, 'Is Mental Privacy a Component of Personal Identity?' (2021) *Frontiers in Human Neuroscience*.

⁵⁴ Cf. Bublit, *supra* n 36, at 400.

⁵⁵ General comment No. 34, CCPR/C/GC/34 at 10.

remain silent.⁵⁶ As such, mental privacy is at least in part protected by the qualified right to freedom of expression alongside the right to privacy.⁵⁷

There may well be compelling reasons to protect the secrecy of certain types of mental content through the absolute RFoT over and above the qualified protection offered by the rights to privacy and freedom of expression. The importance of the interest in not being made to reveal one's political opinions may be an example. Meanwhile, obligations to reveal other kinds of personal mental states should sometimes be permissible, for example, in the case of the obligation of witnesses to disclose what they know about a criminal offence at trial, which is typically addressed as an issue of freedom of expression.⁵⁸ Presupposing that mental privacy is a core attribute of the *absolute* RFoT—as the UN Special Rapporteur seems to do in his latest report—raises foundational questions about how this form of privacy protection relates to more traditional approaches towards the *qualified* protection of privacy and personal data.

In considering these questions, we should distinguish between the law as it is and the law as it ought to be. Human rights should be considered as a living instrument, which must be interpreted in light of present-day conditions. Developments in emerging technologies, such as mind-reading techniques, may well give good reasons to strengthen the current level of privacy protection provided by qualified rights, for instance, by bringing some forms of privacy invasion within the scope of the absolute RFoT. But before we reinterpret the RFoT in such a way, it is essential to assess what exactly the right does and does not cover in our current understanding—also in relation to other rights, such as the right to freedom of expression and the right to privacy. Assigning absolute human rights protection to privacy interests requires a compelling justification. We should not just *assume* that the RFoT covers mental privacy without taking full account of how mental privacy is embedded in the broader framework of human rights law.⁵⁹

6. CASE STUDY: NON-CONSENSUAL PSYCHIATRY

In mental health care, talking therapies and medications are often used to alleviate psychological symptoms and to support patients on their road to recovery. A considerable part of the suffering of people with mental health difficulties experience relates to their thoughts. Treatment, therefore, may be directed, at least in part, towards changing thoughts. For instance, cognitive behavioural therapy for generalized anxiety disorder aims to diminish a person's catastrophic thoughts—thoughts about what can go terribly wrong—that may cause significant distress.⁶⁰ Antidepressant medication may aim to reduce a person's negative thoughts about herself, including suicidal thoughts. And antipsychotics may be used to diminish paranoid thoughts, such as the belief that someone is conspiring against the patient. Usually, such treatments will be given with the patient's consent. However, in some circumstances, treatment may be imposed without the patient's valid consent (or despite a patient's valid refusal). Such non-consensual treatment is, at least to our knowledge, a universal practice,⁶¹ and it is, in many jurisdictions, permitted by domestic laws and ethical principles guiding psychiatric treatment.⁶² Generally, in addition to the presence of a mental illness, an additional requirement for compulsory treatment

⁵⁶ Harris et al., *supra* n 24 at 595.

⁵⁷ Lighthart, *supra* n 1.

⁵⁸ *Wanner v Germany*, Application No 26892/12, Decision, 23 October 2018.

⁵⁹ Lighthart et al., *supra* n 3. See also: Lighthart, *Coercive Brain-Reading in Criminal Justice* (2022).

⁶⁰ Meynen, 'Generalized Anxiety Disorder and Online Intelligence' (2011) 6 *The Journal of Philosophy, Ethics, and Humanities in Medicine* 1.

⁶¹ Stenlund, 'Cognitive Liberty of the Person with a Psychotic Disorder' in Blitz and Bublitz, *supra* n 11.

⁶² *Ibid.* at 241; Bublitz, *supra* n 11 at 73.

is either the individual's *dangerousness* or *need for treatment*—where the danger must result from the mental disorder.⁶³

Though we think that compulsory treatment for mental health conditions can be ethically and legally justified in certain well-circumscribed situations,⁶⁴ it remains a topic of controversy. For example, anti-psychiatry patient movements have called for the abolition of compulsory psychiatric interventions for years. These movements sometimes invoke the absolute RFoT since psychiatric interventions often aim at changing thoughts, thought patterns or beliefs.⁶⁵ Against the backdrop of these considerations, it is unsurprising that the Special Rapporteur on Freedom of Religion or Belief recently considered mental health as a specific area of concern regarding the RFoT:

‘[P]sychotherapies, shock treatments, lobotomies and forced medication—some of which the medical community has denounced—reportedly have been used to coercively alter the thoughts of individuals, forcibly reveal thoughts (beyond legitimate therapeutic purposes), punish ‘inferred’ thoughts, or even physically modify brains, in separate or cumulative violations of the freedom’.⁶⁶

Here, the Rapporteur expresses concerns about abuse of tools for treating people with mental health conditions. At the same time, he acknowledges that treatment for mental health may sometimes be necessary for “restoring” a person’s freedom of thought, for example, if one experiences delusions.⁶⁷ In other words, treatment—and we understand this to include non-consensual treatment⁶⁸—may sometimes *enhance* a patient’s freedom of thought. Likewise, Stenlund highlights that some patients perceive psychosis to be like a ‘foreign entity that has seized them in its grip’. In those cases, psychosis can be considered an external force imposing itself to the *forum internum* dimension. On this approach, ‘involuntary treatment is not to restrict, but instead to return, the patient’s cognitive freedom’.⁶⁹ Moreover, the Human Rights Committee has observed that compulsory treatment may sometimes be ‘necessary and proportionate’ to protect a patient from serious harm or preventing injury to others—yet only as a measure of last resort, applied for the ‘shortest appropriate period of time’, accompanied by ‘adequate substantive and procedural (legal) safeguards’.⁷⁰

As already noted, present debates suggest a broad understanding of the RFoT, covering the non-consensual alteration of any kind of thought. At the same time, the RFoT is an absolute right—at least, that is the dominant view in influential handbooks on human rights and the current position of the Human Rights Committee and the UN Special Rapporteur on Freedom of Religion or Belief. Such an interpretation—a broad scope and absolute nature—could have far-reaching implications in the field of psychiatry. For example, in many countries, it is permitted to forcibly administer medication to an aggressive patient under mental health law to prevent harm to self or others.⁷¹ And, in the Netherlands, for instance, individuals can receive compulsory treatment at home, in particular, compulsory medication, for example, to avoid

⁶³ Olivia et al., ‘Compulsory Psychiatric Admissions in an Italian Urban Setting’ (2018) 9 *Frontiers in Psychiatry* 1; De Stefano and Ducci, ‘Involuntary Admission and Compulsory Treatment in Europe’ (2008) 37 *International Journal of Mental Health* 10.

⁶⁴ Cf. Harris et al., *supra* n 24 at 522–523.

⁶⁵ Bublitz, *supra* n 11 at 72–73; Cf. Alegre (2022), *supra* n 1 at 55.

⁶⁶ Shaheed, *supra* n 2 at 80.

⁶⁷ Shaheed, *supra* n 2 at 81.

⁶⁸ Also, because of the next sentence, which concerns compulsory treatment.

⁶⁹ Stenlund, *supra* n 61 at 250.

⁷⁰ General Comment No. 35 (2014) on liberty and security of person (CCPR/C/GC/35), at 19.

⁷¹ See also, CPT/Inf(2017)6; Craig, ‘Incarceration, Direct Brain Intervention, and the Right to Mental Integrity’ (2016) 9 *Neuroethics* 107 at 108.

less acute dangers such as a decline in social functioning.⁷² Similarly, in the United Kingdom, individuals can be subject to community treatment orders. Such compulsory interventions could well involve administering antipsychotic medication for months, to, for example, alter delusional beliefs, thereby allowing the person to participate more fully in society than would otherwise be possible. If a broad understanding of ‘thought’ is used and the RFoT is absolute, current practices of acute and less acute compulsory treatment will no longer be permissible—at least not according to the letter of the law. Even though it seems widely accepted that under certain circumstances involuntary psychiatric treatment would be allowed, or even necessary, it is not clear how this can be squared with a broad and absolute RFoT.

Finally, it is worth noting that the UN Special Rapporteur makes an explicit recommendation in this regard, prescribing that ‘[m]ental health professionals should firmly establish human rights as core values when prioritizing mental health interventions, including in relation to forced treatment’.⁷³ This is an important observation that, obviously, needs to be further developed, taking account of both the psychiatric and legal perspective. Together with the Rapporteur, we believe that the RFoT is consistent with some non-consensual treatments. From a legal perspective, this suggests that, if the RFoT is to be conceived as having a broad scope, the right must be formulated so as to include some exceptions.⁷⁴ Interestingly, regarding the absolute prohibition of ill-treatment pursuant to Article 3 ECHR, non-consensual treatment will in principle not amount to torture, inhuman or degrading treatment where it is of ‘medical necessity’.⁷⁵ Whether and how similar exceptions should be formulated and operationalised under the RFoT is an important issue for many people worldwide which needs urgent attention in light of Article 18 ICCPR and parallel provisions such as Article 9 ECHR.

Elsewhere, one of us has suggested a rule-and-exception model for the RFoT, which could serve as a starting point for further discussion. Assuming that certain paternalistic measures are generally acceptable, an intervention contravening the negative dimension of freedom of thought could be permissible if and only if:

- ‘(a) the person is not competent to make a decision about such interventions herself, (b) the intervention aims at improving the freedom to think by alleviating substantial deficits in thinking abilities or rational belief formation, (c) it is in the best (medical) interest of the person, (d) there are no less invasive means, and (e) the benefits of the intervention outweigh the setbacks, all things considered. (. . .) (f) interventions do not primarily pursue other governmental goals and (g) that they do not seek to imprint moral, political, or other values, they must strive to be content neutral’.⁷⁶

According to this exception, to be permissible, the primary and dominant characteristic of an infringement of the RFoT must be the promotion of freedom of thought itself. This would rule out the permissible use of, for example, neurotechnological memory detection or thought modulation in criminal justice, as such interventions shall often primarily aim at prosecuting and preventing crime rather than at fostering the person’s freedom of thought.⁷⁷ Meanwhile, other authors have suggested that, sometimes, the RFoT should enable these kinds of interferences too through a balancing of the competing interests at stake. For example, McCarthy-Jones argues

⁷² De Waard et al., ‘Compulsory Treatment in Patients’ Homes in the Netherlands’ (2020) 20 *BMC Psychiatry* 1.

⁷³ Shaheed, *supra* n 2, at 99.

⁷⁴ Bublitz, *supra* n 11; McCarthy-Jones, *supra* n 1.

⁷⁵ Harris, *supra* n 24 at 268.

⁷⁶ Bublitz, *supra* n 11 at 90–91.

⁷⁷ This exception is furthermore confined to people who lack decision-making capacity, which might have a considerable impact on current practices in psychiatry, where sometimes individuals are treated non-consensually even when they have decision-making capacity and are validly refusing treatment.

that it is possible to imagine a judge granting a warrant for a mental search with neurotechnology. If emerging technological applications like these are deemed to infringe the absolute RFoT, this could restrain further development of such technologies, which may not be desirable given the general benefits the technology could have. Hence, some exceptions should be possible. Part of a justification for permissible infringements, McCarthy-Jones argues, could be the claim that the RFoT needs to be balanced against other interests and rights.⁷⁸ In-depth elaboration on the question of whether and, if so, which limitations of the RFoT should be permissible, exceeds the scope of this paper. But it is clear that this issue is in need of further debate—especially, if we are to understand the RFoT as having a broad scope.

7. MORAL FOUNDATIONS

The drafters of Article 18 UDHR considered the suppression of free thinkers, scientists and dissidents as paradigmatic violations of the RFoT, relating the right to the foundations of a democratic society.⁷⁹ However, democratic justifications of the right are difficult to square with more recent interpretations of it, which see it as protecting against even interferences with trivial or politically irrelevant thoughts. These broader interpretations of the RFoT call for new moral foundations, and examining the candidate foundations may help us to more clearly and precisely specify the right's boundaries. One may look for different moral values that may justify a moral RFoT which in turn may ground the legal right. There are two important aspects to this task. One concerns the justification of the right, and the other concerns the justification of its absolute nature. We begin with the former, and we focus on the part of the right that protects against certain means of *altering* a person's thoughts, which we will refer to as the right against thought modulation. However, some of what we say may apply to other elements of the RFoT too.

One way of justifying the right against thought modulation appeals to its role in protecting our *autonomy of thought*,⁸⁰ which is understood as our ability to control the contents of our own mental lives. Certain ways of altering a person's thoughts typically compromise the person's control over her own thoughts.⁸¹ This is plausibly true, for example, of interventions that *irreversibly* alter aspects of a person's thought that would otherwise be within her voluntary control. Consider, for example, the use of subliminal imagery or brain-active drugs to produce powerful, addiction-like desires and thereby dictate a person's decisions. Insofar as these decisions would otherwise have been made voluntarily, these interventions will, at least in one respect, reduce a person's control over her own thought. Prohibiting such interventions by recognizing a right against thought modulation would protect against this loss of control, and it may be possible to justify the right on this basis.

Another candidate justification for the right against thought modulation appeals to its role in protecting *well-being*. Some forms of thought modulation reliably diminish the targeted individual's well-being. Consider, for instance, marketing practices that deceive people into parting with their money by exaggerating the benefits of a product, or falsely implying that the product has been recommended by other customers. Alternatively, consider online techniques, such as the use of bottomless news feeds and randomized rewards (so-called 'loot boxes'), to keep users online, or playing a game, for longer than they would, at other times, judge to be desirable. The right against thought modulation might be justified by the role that it plays in preventing such threats to well-being.

⁷⁸ McCarthy-Jones, *supra* n 1.

⁷⁹ Taylor, *supra* n 23, at 501; Bublitz, *supra* n 11; Shaheed, *supra* n 2, at 3.

⁸⁰ McCarthy-Jones, *supra* n 1.

⁸¹ Saghai, 'Salvaging the Concept of Nudge' (2013) 39 *Journal of Medical Ethics* 487 at 489; Faden, Beauchamp, and King, *A History and Theory of Informed Consent* (1986), 238; Kipper, 'Irresistible Nudges, Inevitable Nudges, and the Freedom to Choose' (2021) 8 *Moral Philosophy and Politics* 285 at 296.

Finally, a third way of justifying the right against thought modulation appeals to the concept of *respect* (and might alternatively be framed in terms of *dignity*). Some forms of thought modulation are disrespectful in the sense that they treat the targeted individual in a way that does not befit her status as a rational being.⁸² The most obvious examples here are forms of thought modulation that *undermine* or *diminish* a person's rational capacities, but it is plausible that thought modulation can also be disrespectful simply because it *fails to engage* a person's rational capacities—the modulator produces thoughts through non-rational mechanisms when she could and should instead have engaged the person's rational capacities.⁸³ Consider a case in which, rather than seeking to persuade customers to buy certain products, an advertiser uses subliminal images to make those products more salient in the customers' minds. One might seek to justify the right against thought modulation by appealing to the protection it provides against such disrespectful forms of influence.

These different moral justifications for the right against thought modulation would likely have different implications for what should be included within the content of that right. For example, if the right is justified by its role in protecting autonomy of thought, it is hard to see why the right should cover interventions that alter thoughts in ways that can easily be resisted (like mild 'nudges'). In such cases, the targeted individual retains control over her thoughts because she retains the option of resisting the alteration. By contrast, if the right is justified by reference to well-being or respect, it might cover even easily resistible alterations, for example, if these happen to reliably diminish well-being or undermine a person's rational capacities.

On the other hand, if the right against thought modulation is justified by reference to well-being, it is hard to see why it should cover cases in which a person's thought is modulated in order to protect the person from significant harm—what we might call *paternalistic* thought modulation. Consider, for example, the use of compulsory treatments under the 'danger to self' clauses of mental health legislation. However, if the right is justified by reference to respect, it could be argued that the right should be specified so as to cover many forms of paternalistic thought modulation. Thought modulation is arguably consistent with respect only where (a) the intervention engages and does not diminish, the patient's rational capacities (for example, because it consists in presenting a person with arguments or reasons, as in some 'talking therapies'), or (b) is employed on someone whose rational capacities have already been altogether destroyed (for example, by a severe psychiatric disease). Thus, questions regarding the moral foundations of the RFoT are relevant to the problems concerning psychiatric interventions laid out in the previous section.

Finally, there is also a need to consider the moral basis of the absolute level of protection provided by the RFoT. In moral philosophy, rights that can *never* be outweighed by other interests or the common good are hard to justify. The main strategies for defending rights in moral philosophy appeal to (a) the costs and benefits of recognising the right (consequentialist justifications), (b) whether the right would be chosen by the parties to some hypothetical social contract (contractualist justifications) and (c) whether the right can vindicate widely held intuitions about which actions are and which are not morally justified. But all of these strategies tend to favour the recognition of derogable non-absolute rights—rights that can be permissibly infringed at least in extreme cases (for example, in order to avoid a catastrophe)⁸⁴ and possibly in a much wider range of cases (for example, whenever respecting the right would have much worse consequences than infringing it). Some theories—most notably, that of Immanuel Kant—have been held to justify absolute rights against, for example, killing and

⁸² For this conception of respect, see, for example, Eidelson, 'Respect, Individualism, and Colorblindness' (2020) 129 *The Yale Law Journal* 1600.

⁸³ See Gorin, 'Towards a Theory of Interpersonal Manipulation' in Coons and Weber (eds), *Manipulation* (2014).

⁸⁴ Cf. Article 4 ICCPR, stipulating that the RFoT is non-derogable even in cases of public emergency.

lying. But these theories are highly controversial, in part, because they fail to accord with widely held moral intuitions; for example, most people intuitively judge that it is morally justified to lie in order to prevent someone from being killed. Therefore, the absolute human RFoT provides an intriguing challenge for moral philosophy. Conversely—and in line with the observations in the previous section—when interpreting the right, legal scholars should not simply assume that there is a clear and uncontroversial moral justification of its absolute nature.

8. CONCLUDING REMARKS

The RFoT is currently underdeveloped and the need for it to be more fully specified is growing—especially, as it appears that the right could be threatened in many ways. In this regard, lawyers, philosophers and psychiatrists as well as international actors, such as UNESCO and the Council of Europe, are drawing special attention to emerging challenges created by modern technologies. Likewise, the UN Special Rapporteur on Freedom of Religion or Belief, emphasizes that these technologies pose dilemmas about how to protect thoughts from impermissible manipulation, how to guarantee mental privacy and how to prevent these technologies from being used and abused to punish thoughts. When addressing these challenges, it is essential to consider the scope as well as the strength (for example, absoluteness) of the RFoT—and the possible need for exceptions, for instance, in mental health care. In addition, in order to prevent redundant legal protection and confusion regarding the strength of that protection, the RFoT must be interpreted and developed in close relation to other human rights over the mind, such as the rights to mental privacy and integrity, by taking account of how these rights play out in individual cases such as in non-consensual psychiatric treatment. Finally, in rethinking and further specifying the scope and strength of the RFoT, it may be possible to derive guidance from philosophical thinking regarding the moral foundations of the right.

ACKNOWLEDGEMENTS

CB is funded by the German Federal Ministry of Education and Research in the ERA-NET Neuron project Hybrid Mind (01GP2121A). TD is funded by the European Research Council (Consolidator Award 819757) and Uehiro Foundation on Ethics and Education. LF is funded by the British Academy (Postdoctoral Fellowship award pf170028) and Uehiro Foundation on Ethics and Education. GM and SL are funded by the Dutch Research Council (NWO Vici grant VI.C.201.067).