



Exploring the Essence of the Freedom of Thought – A Normative Framework for Identifying Undue Mind Interventions

Timo Istance 

Received: 3 July 2024 / Accepted: 13 January 2025 / Published online: 23 January 2025
© The Author(s), under exclusive licence to Springer Nature B.V. 2025

Abstract The freedom of thought (FoT) has recently gained attention in human rights scholarship, emerging as a key component in the human rights protection of the human mind. However, this newfound interest has exposed significant gaps in the protection offered by the FoT. While the underdevelopment of the FoT is mainly examined in relation to the mind's vulnerability to emerging neurotechnologies, there are numerous other ways to interfere with the privacy, freedom, and integrity of the mind. Conversations, education, online marketing, and abusive indoctrination programs can all interfere with individual's minds. These different forms of mind interventions (MIs) appear to pose varying degrees of threat to the FoT, particularly to the right not to have one's thoughts impermissibly altered. Despite this, the theoretical exploration of the impact of different MIs on this aspect of FoT remains largely unexplored. This article explores the core of the right not to have one's thoughts impermissibly altered. First, the current shortcomings in the conceptualisation of FoT will be outlined. Second, the article defines mind interventions to delineate the scope of the analysis. Third, it introduces a normative framework to identify Undue Mind Interventions (UMIs). To this end, a

general maxim will be presented: *a MI qualifies as an UMI when the agent intentionally undermines a subject's mental autonomy by bypassing, distorting, or disarming the mental abilities that enable it*. The core elements of this maxim will be detailed. Finally, the impact of this normative assessment on the meaning of FoT will be discussed.

Keywords Freedom of thought · Mind intervention · Mental autonomy

Introduction

The freedom of thought (FoT) has recently gained attention in human rights scholarship, emerging as a key component in the human rights protection of the human mind. However, this newfound interest has exposed significant gaps in the protection this freedom offers. While the underdevelopment of the FoT is mainly examined in relation to the mind's vulnerability to emerging neurotechnologies, there are numerous other ways to interfere with the privacy, freedom, and integrity of the mind. Conversations, education, online marketing, and abusive indoctrination programs can all interfere with individual's minds. These different forms of mind interventions (MIs) appear to pose varying degrees of threat to the FoT, particularly to the right not to have one's thoughts impermissibly altered. Despite this, the theoretical exploration of the impact of different MIs on this aspect of FoT remains largely unexplored.

T. Istance (✉)
Research Group Personal Rights and Property Rights
of the University of Antwerp, FWO Fellow Fundamental
Research, Antwerp, Belgium
e-mail: Timo.Istance@Uantwerpen.be

This article explores the core of the right not to have one's thoughts impermissibly altered. First, the current shortcomings in the conceptualisation of FoT will be outlined. Second, the article defines mind interventions to delineate the scope of the analysis. Third, it introduces a normative framework to identify Undue Mind Interventions (UMIs). To this end, a general maxim will be presented: *a MI qualifies as an UMI when the agent intentionally undermines a subject's mental autonomy by bypassing, distorting, or disarming the mental abilities that enable it*. The core elements of this maxim will be detailed. Finally, the impact of this normative assessment on the meaning of FoT will be discussed.

Freedom of Thought and the Normative Qualification of MIs

In recent years, legal scholars have sought to clarify the freedom of thought (FoT) to protect the privacy, autonomy, and integrity of the *forum internum*—the inner sanctum where mental faculties are developed and exercised [1].¹ Despite being fundamental to many other rights and freedoms, the FoT has received sparse attention in legal doctrine and case law, resulting in its unclear scope and meaning. This dilutes its practical utility as a legal safeguard to protect the *forum internum* against undue interferences.² In 2021, the UN Special Rapporteur on freedom of religion or belief (SR) issued a first substantive report on FoT, outlining its core elements and identifying practices that violate this freedom [1]. The FoT includes: i) the freedom not to reveal one's thoughts, ii) the freedom from punishment for thoughts, iii) the protection from impermissible alteration of thought, and iv) a state's obligation to ensure an enabling environment for free thought. This analysis focuses on the protection from impermissible alteration of thought. While this is an established human rights protection, the definition of "impermissible alteration" remains unclear. Not every form of persuasion violates the FoT. However,

the point where acceptable social persuasion becomes impermissible alteration remains undefined. The SR identifies coercion, modification, and manipulation as methods that are past that point and interfere with the FoT. Fleshing out the meaning of these concepts is crucial in the process of clarifying the FoT, and the creation of a wider mind-protection framework.

The emergence of NTs and the growing concerns regarding their ability to interfere with people's mental processes, has highlighted the underdevelopment of the FoT, and incentivised legal experts to elaborate the core principles of this freedom [4]. Nevertheless, the FoT does not only face threats of violations by misuses of NT. On the contrary, many ways exist to intervene into the human mind. While many of these mind interventions (MIs) may be less spectacular, the effects they have on people's mental sphere and the ethical concerns they raise may be highly similar, and even more pressing than those generated by state-of-the-art NTs. Therefore, the ongoing debate on the development of the FoT as the central pillar of a human rights approach towards NT, should be taken as an opportunity to develop the FoT as a coherent and effective safeguard to protect the human mind against Undue Mind Interventions (UMIs) of all kinds. Moreover, efforts to clarify the FoT should be part of broader initiatives to establish a mind-protection framework within the human rights framework – akin to the collection of well-developed human rights principles protecting the integrity of the human body. While the FoT takes central stage in such a framework, it should be interpreted in relation to other rights, most notably the right to privacy and the right to (mental) integrity. These rights should be further developed collectively as to form a coherent set of safeguards protecting the privacy, autonomy, and integrity of the *forum internum*.

Mind Interventions: Definition and Case Examples

Definition

As the aim of this article is to indicate which MIs would amount to UMI and therefore interfere with the FoT, it is necessary to define 'mind interventions'. First, a definition of 'mind' is required. While the rich philosophical debate on the definition of the mind and its relation to the body or brain is extensive, it falls

¹ While acknowledging the dual structure of the FoT, protecting both an internal space for thought (*forum internum*) and an external space for its expression (*forum externum*), this analysis will solely focus on the former.

² While crucial for shaping the FoT, this article does not focus on the question of which mental states fall within the FoT's scope. For such an analysis, see [2, 3].

outside this article's scope. For the purpose of this analysis, the mind encompasses 'all conscious and non-conscious mental states, events and processes, i.e. thoughts and beliefs, emotions and moods, as well as the underlying psychological mechanisms operating in the complex mental machinery' [5], p. 387. Furthermore, without endorsing a specific theory in the mind-body debate, it is widely accepted that mental phenomena are (somehow) related to brain processes. This premise will be assumed in this article, without endorsing any further claims on the philosophical and scientific mystery as to how the brain and mind relate.

Against this background, MIs are defined as *deliberate actions undertaken by an agent to influence the mind of another individual, commonly referred to as the subject*. Several key elements in this definition deserve further elaboration. First, an agent can be a natural person, but also a collective entity – legal or otherwise – such as a company or government through which (a group of) natural persons operate. Second, 'altering' broadly refers to any influence to the mind of the subject that would not have occurred without the deliberate action of the agent. Third, MIs refer to deliberate actions. Alterations of a subject's mind emanating from environmental factors such as air quality or exposure to sunlight – and thus not stemming from deliberate human action – do not qualify as MI for the purpose of this analysis. Moreover, certain situational factors – general or specific – may have profound impacts an individual's mind. For instance, poverty affects both the shape of the brain and the mind of individuals [6], or the socio-cultural environment wherein individuals live, characterised by a dominant value system, will impact the course of an individual's beliefs, thoughts, and affective states. Yet, while these factors influence an individual's mind, they are not reducible to a deliberate action (or coherent set of actions) to change minds attributable to an identifiable actor. These rather constitute latent, situational elements which results from an interplay between a myriad of factors, and therefore do not qualify as MIs.³ Fourth, this definition

only encompasses interpersonal MIs taking place between an agent and a subject. This again excludes environmental and situational factors that influence the mind, but also self-induced alteration of an individual's mind – such as changes in moods or cognitive capacities due to engaging in physical exercise, mindfulness practices, medication use or food consumption. While these MIs lie at the root of an interesting debate on the potential moral limits of cognitive enhancement and its interplay with fundamental moral concepts such as autonomy and human dignity, they fall out of the scope of this analysis. Fifth, for an action to be classified as a MI, it must result in actual changes in the subject's mind. The mere intention to create mental effects, while necessary, is not sufficient.

Case Examples

Some case examples of MI will be outlined to illustrate this concept. While currently largely hypothetical, neurotechnological MIs, particularly through neuromodulation (or neurostimulation) technologies, are the category of MI that gave rise to concerns about the freedom of individuals' minds. These technologies, like Deep Brain Stimulation (DBS) and Transcranial Brain Stimulation (TBS), target specific brain regions with electrical currents [7, 8]. Originally developed for treating neurological motor disorders like Parkinson's Disease and epilepsy, clinical studies have revealed unintended effects on patients' mental states, as reflected by reports of increased impulsivity, aggression, gambling behaviour or symptoms of depression and mania [9]. Moreover, neuro-modulation techniques are now employed with the specific aim to directly influence mental states, treating conditions like Obsessive Compulsive Disorder, depression, addiction, and anorexia nervosa [10, 11]. Some commercially available applications claim to enhance cognitive abilities such as memory, attention, and problem-solving [12, 13]. Notably, research is underway to explore memory manipulation through brain-modulation techniques [14]. This suggests that NTs have the potential to alter a subject's mind.

³ While these contextual factors will not qualify as MIs on their own, they may be exploited or capitalised on by an agent in their effort to alter a subject's mind, and will in that capacity be of importance in the normative assessment of MIs. For instance, when poverty makes an individual more vulnerable to abusive exploitation methods due to their weaker socio-

Footnote 3 (continued)
economic status, this will be a relevant element when assessing the moral acceptability of a MI (*infra*).

While the current unpredictability makes this method unattractive for agents seeking targeted interventions into another person's mind, it remains a conceivable future avenue for such interventions. Moreover, it has been demonstrated that an agent could potentially hack a brain-stimulating device to manipulate their brain processes, simultaneously influencing their mental states [15].

Nevertheless, while NTs have sparked a vivid debate on the moral acceptability of MIs, they are not the sole, nor the most effective means for altering minds [16]. Numerous other interventions achieve this goal more efficiently. Tying in with the medical context, neuropharmaceuticals⁴ are a pertinent example. The use of neuropharmaceuticals aimed at altering individual mental states constitutes a MI when administered by a third party. While typically consensual in research trials or medical treatments (or even self-induced when individuals themselves administer the drugs without a physician's involvement), emblematic yet morally contentious examples include forcibly administering antipsychotic drugs to psychiatric patients or detained individuals to modify behaviour or ensure trial competency. For instance, prevalent in many jurisdictions is the administration of antipsychotic drugs to persons with a mental illness prone to violent behaviour.⁵ Alongside their intended effects such as mood stabilisation and aggression reduction, neuropharmaceuticals can yield unpredictable side effects affecting mental states. For instance, fluphenazine (Prolixin) stabilises – and thus alters – moods, thoughts and behaviour in schizophrenia treatment but may induce depression, confusion, unusual dreams, or intensify psychosis [18]. The

mental changes resulting from the administration of neuropharmaceuticals reflect deliberate actions by identifiable agents intending to alter mental states.

Furthermore, marketing practices, ranging from traditional advertising (e.g., billboards and print ads) to online marketing, present an increasingly scrutinised area of MIs. Certain persuasion strategies, such as subliminal advertising, are explicitly prohibited in many jurisdictions due to their manipulative nature. In recent years, growing concern has emerged over the widespread use of AI-driven persuasion strategies. These represent a significant advancement in persuasive technology, blurring the line between acceptable and problematic online persuasion [19]. For instance, AI algorithms enable marketing tactics such as microtargeting. Microtargeting, a pivotal tool in online advertising, creates individual user profiles based on extensive data to tailor marketing efforts.⁷ The use of AI and Big Data approaches further enhances profiling capabilities for commercial, electoral, or governance purposes. Companies like Amazon, Netflix, and Facebook employ AI algorithms to analyse customer data, discern preferences, and steer users toward specific products and services [21]. Similarly, in the political realm, microtargeting may play a significant role, as exemplified by its use in the 2016 US presidential elections, where political action committees (PACs) utilised Cambridge Analytica datasets to influence voter behaviour, raising concerns about democratic integrity [22–24]. This AI-mediated, informationally empowered advertising effectively influences decision-making processes and mental states, including emotions, intentions, and beliefs, drawing moral scrutiny due to exploitation of psychological vulnerabilities [25]. The deliberate creation and operationalisation of such online algorithms by companies or political entities aim to influence users' minds and behaviour, thus qualifying as MIs.

A final example of MI is cult indoctrination. The notion “indoctrination” is ambiguous, evading a strict definition, as reflected in its varied conceptualisations across scientific disciplines and common language. This ambiguity stems from the diverse contexts,

⁴ Defined as a drug used to treat neuropsychiatric, neuropsychological, or nervous-system disorders; Merriam Webster Dictionary.

⁵ For instance, the US Supreme Court established that involuntary medication of prisoners by the state is acceptable if due substantial and procedural safeguards are in place, e.g., an established mental illness, the treatment must be in the interest of the imprisoned person, and a system of administrative review is in place; US Supreme Court *Washington v. Harper*, 494 U.S. 210 (1990).

⁶ Nevertheless, this acceptance represents a present, possibly temporary balance among significant medical and legal concerns, including informed consent, competency, treatment refusal, due process, and privacy rights. Weller suggests that, in the future, people may look at this practice in the same way they look at the practice of lobotomising patients [17].

⁷ While research cautiously suggests that microtargeting may wield a greater influence potential than traditional advertising, it is important to underscore that it does not have the spectacular manipulative capabilities often attributed to it [20].

purposes (including political, religious, educational, or commercial), and methods associated with indoctrination. Identifying the common elements connecting indoctrination practices—ranging from prisoners of war indoctrinated by governments to cult indoctrination, religious indoctrination in schools, or indoctrination in Multi-Level-Marketing Schemes—is not straightforward. Therefore, cult indoctrination serves as a focused example. Cult indoctrination involves an identifiable agent (cult leader/leadership) systematically manipulating subjects' mental states through a combination of specific actions. Hassan's model for identifying cults defines a destructive cult as 'a group that violates its members' rights and damages them through the abusive techniques of unethical mind control' [26]. This mind control encompasses the manipulation of behaviour, information, thought, and emotion. Tactics employed to gain and maintain control include relocation, deceit, fostering confusion, social isolation, dissent restriction, group pressure, guilt induction, conditional affection, threats, and fear [27, 28]. For the analysis of cult indoctrination as a MI, it is useful to present an epistemological account of indoctrination, which conceptualises indoctrination as a form of manipulation shaping subjects' reasoning patterns so as to generate predetermined mental states [29]. This deliberate reconfiguration of mental processing by subjecting individuals to harmful practices constitutes a MI, profoundly interfering with minds. Hand aptly suggests that 'to indoctrinate people, is to interfere with their minds in the most serious way possible' [30], p.352.

These examples highlight the vulnerability of individuals' minds to external interventions, across a wide variety of contexts. They will serve as crucial reference points throughout this analysis. However, it is essential to recognise that commonplace practices such as political debates, education, and simple social interactions can also qualify as MIs. While these daily interventions shape minds, their moral acceptability is often overlooked as compared to NTs or indoctrination schemes which raise substantial ethical concerns. Nonetheless, the normative framework set up below aims to contribute to distinguishing between acceptable and undue mind interventions in a way that applies not only to spectacular and uncommon MIs – such as NTs or cult indoctrination – but also to everyday mental influences.

Mental Autonomy as Normative Criterion to Identify UMIs

The Quest for a Normative Criterion

To establish such a framework, it is essential to formulate a criterion for identifying UMIs.⁸ For instance, consoling a nervous friend before a job interview by offering reassurance and citing past successes is benign and ethically unproblematic. Similarly, engaging in persuasive conversations and using rhetorical tools and body language to advocate political beliefs is deemed acceptable within the realm of social influence. Human social influencing is integral to public interaction, with persuasive dialogues serving as linchpins of democratic societies, warranting no moral reproach or infringement of human rights.⁹ Individuals experience continuous influence across various domains, including interactions with businesses, political organisations, parents, friends, and romantic partners [31]. Mental persuasion in itself is not per definition morally problematic. It is therefore important not to define UMIs in an overly inclusive fashion. Conversely, moral intuition indicates that altering the thoughts and emotions of that same friend who is awaiting their job interview through means such as wireless neurostimulation (and without their consent) crosses ethical boundaries. Despite potentially similar cognitive and affective outcomes, engaging in a MI through the unsolicited alteration of brain processes appears as morally problematic.¹⁰ The critical task involves exploring arguments that may support this intuition by identifying characteristics of MIs, where reassuring a friend through conversation constitutes Acceptable Mind Intervention (AMI) whereas influencing their mental states

⁸ The term Undue Mind Intervention is selected over other terms commonly found in philosophical and legal literature, such as manipulation or 'mind control', which can have varying and sometimes ambiguous definitions. By opting for UMI, I aim to avoid any unintended associations with these pre-existing notions.

⁹ Even when acknowledging that a considerable part of this communication will consist of body language and unexplicit rhetorical style methods that unconsciously influence the other person.

¹⁰ However, considering today's state of the art in NT, conversational tools will be more successful and produce stronger effects than MI via NT.

through unsolicited brain stimulation amounts to an UMI. These characteristics serve as criteria for establishing ethical boundaries for MIs on a case-by-case basis. This, subsequently, will inform the formulation of appropriate human rights principles – within the scope of the FoT – concerning the autonomy and integrity of the human mind that can simultaneously apply to, for instance, indoctrination at school, government nudging, and online manipulation.

Levy's parity principle offers a suitable starting point for establishing a normative criterion to support the moral distinction between different MIs. He suggests that MIs with identical or highly similar effects on the mind should be treated equally on a normative level, regardless of the means used to realise the MI [32].¹¹ However, the moral irrelevance of the *modus operandi* for MIs is challenged by Bublitz and Merkel, who argue that 'if two interventions have the same or sufficiently similar effects, they may still differ in the way the effect is produced, i.e. the causal routes into the mind, and this difference is crucial' [33], p.71. The source and method of mind alteration therefore appear to be of moral importance. This leads to the important tasks to identify inherent characteristics that differentiate various MIs in a morally relevant way.

In neuroethical literature, this issue is often addressed by highlighting the distinction between direct and indirect MIs. Indirect MIs operate through perceptual and cognitive pathways, altering mental states via the processing of external stimuli by the senses (e.g., vision, hearing) [34]. In contrast, direct MIs bypass perceptual pathways and cognitive processes and influence the mind through the alteration of neuronal function or brain structures [35].¹² While

moral intuition typically labels direct interventions as UMI and indirect interventions as AMI, this binary classification proves overly simplistic from a normative standpoint [35, 36]. Practices like cult indoctrination, despite utilising perceptual and cognitive pathways, are widely condemned as morally unacceptable. Conversely, direct interventions such as emergency administration of neuropharmaceuticals to psychologically distressed individuals, while potentially controversial, are deemed morally acceptable by some as they ultimately positively affect the mental processes of the individual [37]. Moreover, the direct–indirect dichotomy fails to address more complex scenarios. For instance, subliminal stimuli operate through perceptual pathways yet evade (conscious) cognitive processing, challenging easy categorisation. Thus, this binary distinction does not seamlessly align with ethical judgments.

Therefore, other criteria have been proposed to assess the moral acceptability of MIs. Bublitz and Merkel argue that MIs are morally unacceptable when they violate the right to mental self-determination by undermining an individual's control over cognitive mechanisms, in this way bypassing their ability to rationally process information [33]. In the same vein, Birks and Buyx label the alteration of a subject's thoughts unacceptable when it is realised by intentionally evading their autonomous thought [38]. Shaw and Douglas similarly emphasise the importance of whether MIs engage with a person's rationality, suggesting it as a relevant moral criterion [39–41]. Additionally, Focquaert and Schermer contend that MIs amount to UMI when they disrupt personal (narrative) identity due to the obstruction of critical endorsement of mental changes [36].

Identifying UMI

In line with the proposals outlined here, this analysis suggests the following maxim for identifying UMIs:

A MI qualifies as an UMI when the agent intentionally undermines a subject's mental autonomy by bypassing, distorting, or disarming the mental abilities that enable it.

To elaborate on this overarching principle, this article will: i) clarify the concept of mental autonomy (MA), ii) delineate what it means to bypass, distort, or disarm the mental abilities underlying MA, iii)

¹¹ Levy's argument does not suggest that we should dismiss concerns about modern mind-altering methods. He rather argues that we should not yet be overly concerned about modern mind-altering technologies. He emphasises that direct interventions, such as neurotechnologies, are not currently as effective in influencing minds as indirect interventions such as persuasive rhetoric or education. Today, these indirect methods have a stronger impact on people's mental lives than neurotechnology.

¹² Bublitz and Merkel pertinently notice that both categories of interventions change the mind, and "both affect (or are realised by) brain activity" [33, p.69]. Without explicitly engaging in the body-mind debate, it is clear that a strict dualistic perspective is unconvincing, as all mental changes necessitate physical (brain) changes.

discuss the significance of consent, and iv) examine the role of intentionality.

At the outset, it is, however, important to acknowledge the limitations of this framework. First, this framework aims to identify mind interventions that are morally problematic due to their *modus operandi*, i.e., the way in which they alter mental states. Some MIs may be morally problematic either because of the content of the mental states being altered (e.g., instilling racist beliefs in a subject) or due to the consequences of those alterations (e.g., persuading a subject to pursue financially irresponsible investments that will inevitably result in grave financial losses). Yet, these fall outside the scope of this framework as their impermissibility does not pertain to the alteration itself, but to the content of the result or accompanying externalities. Safeguards under the FoT, which are in essence designed to protect free thinking and unimpeded thought processes, should remain neutral regarding the content of thoughts and potentially undesirable consequences arising from them.¹³ Second, it is crucial to emphasise that the normative evaluation of MIs requires a case-by-case assessment, considering the specific circumstances of each instance. Nevertheless, the proposed framework provides essential normative and conceptual foundations, within which such assessment could be conducted, and more detailed, operational strategies for such assessments can be developed.

Mental Autonomy

a) Definition

At its core, MA can be understood as the ability to self-determine at the mental level [42]. This definition should be framed within a broader concept of

personal autonomy as developed by Gerald Dworkin. Personal autonomy, though complex and defined in various ways, generally explores the motivational forces guiding individuals' behaviour. Mental elements such as beliefs, desires, emotions, and judgments are pivotal among these motivational factors [43]. MA is to be situated within this realm of mental elements that drive behaviour. Dworkin builds upon Frankfurt's second-order account of personal autonomy, which posits that personal autonomy involves the endorsement of first-order desires through second-order attitudes [44]. These first-order desires, encompassing thoughts, beliefs, and desires – mental states – form the basis of individual motivation that prompts their behaviour (including speech). The ability to critically reflect on and adopt attitudes toward these first-order mental states, in a second-order account of autonomy, is a necessary condition for an individual to be autonomous. These reflective attitudes represent evaluative assessments of first-order mental states and empower individuals to modify them. Dworkin expands upon Frankfurt's theory, acknowledging its limitations in addressing various forms of manipulation in the formation of these higher-order desires. He argues that personal autonomy also necessitates the capacity to critically form, evaluate, and potentially change mental states. This process of internal self-reflection must uphold procedural independence, ensuring it remains untainted by manipulation, deceit, or similar practices. Mental self-determination requires guarantees of non-interference with the processes necessary for critical reflection on one's mental states [45]. This reflection on the formation of second-order desires need not be conscious, as individuals naturally engage in this process without active awareness, according to Dworkin.

The requirement of procedural independence embodies the concept of MA. Individuals should be given the opportunity to critically engage with mental influences in the face of MIs they are subjected to. To allow for self-determination on a mental level, changes in mental states should be explainable through referring to the subject's capacity for the formation of second-order mental states through which they can critically endorse these changes. MIs that actively undermine procedural independence, and obstruct the critical self-reflection on mental changes, violate a subject's MA. The formation of second-order mental states towards one's first-order mental

¹³ Hence, while these MIs may indeed raise moral concerns and warrant attention under human rights law, these should arguably not be framed within the protection of the *forum internum* by the FoT – and specifically the right to protection against impermissible alterations of thought. Instead, such issues might be more appropriately addressed under the freedom of opinion and expression, which currently accounts for the content and effects of certain expressions that underpin these MIs. Moreover, the right to mental integrity could potentially offer protection against mental distress and discomfort caused by such MIs.

states, and the ability to change the latter accordingly, should be a process under the individual's control. Such control can only be realised when MIs engage the subject's mental abilities for critical self-reflection on alterations of their (first-order) mental states. These mental abilities should facilitate: 1) identifying external influences, 2) evaluating induced mental alterations, and 3) deciding whether to endorse or reject these alterations. Thus, when an agent changes a subject's mind by intentionally undermining these mental abilities, preventing the subject from identifying, evaluating, and endorsing or rejecting changes to their mental states, this MI amounts to UMI.

The crucial question here is to articulate which mental abilities facilitate critical reflection needed to deliberately endorse mental changes induced by a MI. The mental abilities required would include attention, concept formation, episodic memory, planning, and rational deliberation [46]. Ideally, however, as Metzinger points out, these mental abilities are further refined to the degree that they consist of empirically observable psychological processes whose operability can be assessed. The integrity and functionality of these abilities are ultimately determinant for the classifications of MIs as UMI or AMI in the respective contexts these MIs take place.¹⁴

It is important to emphasise that the normative criterion discussed here pertains to a procedural understanding of MA, focusing on the capacity for MA rather than its effective utilisation. A MI is not morally unacceptable simply because the subject fails to utilise their abilities to identify, evaluate, and endorse the inhibited mental effects. MA concerns functional autonomy, specifically referring to the presence and proper functionality of mental abilities necessary for reflecting on changes in mental states. It does not pertain to substantial autonomy which would entail the actual use of these abilities, and ultimately the control over one's mental states [47]. Whether the subject exercises their capacity to critically endorse mental changes induced by a MI is inconsequential for normative assessment. For example, television advertisements typically aim to influence consumers'

decision-making processes by targeting first-order mental states and minimising engagement with second-order mental states, leveraging insights from behavioural sciences [48]. However, these advertisements do not (necessarily) bypass, distort, or disarm an individual's mental abilities such as attention, memory, or rational deliberation. Therefore, this form of influencing does not impede the independent formation of second-order mental states, preserving the capacity for MA. It leaves intact the possibility for critical endorsement, leaving it up to the individual to seize that opportunity [49]. This procedural nature eliminates substantive considerations, making questions about the alignment of induced mental changes with the subject's beliefs or values, the agent's interests [50, 51], or the instillation of false beliefs irrelevant. These elements, often suggested to distinguish manipulation from persuasion, are not considered in the assessment of MA.

b) Rationality and MA: important caveats

Incorporating the requirement of procedural independence into the conceptualisation of MA naturally involves integrating a rationality standard. MA requires that mental changes can be endorsed after rational reflection that ultimately searches for reasons supporting such endorsement. This perspective aligns with a Kantian approach to autonomy, which integrates rationality into the concept—a departure from Frankfurt's second-order account of personal autonomy [52]. However, this inclusion introduces significant caveats for MA as a normative criterion.

First, it is crucial to steer clear of portraying absolute sovereignty over one's mind as the default state of human existence. The notion of a purely rational agent who only accepts changes in their mental states and processes if substantive reasons for these changes can be identified, forms a flawed premise for analysing MA. A significant portion of mental processes and events originates from 'the psyche's mysterious, irrational, inexplicable, and oftentimes unconscious forces' [34], p. 366. Many mental processes underlying conscious experience operate beyond conscious control, while much cognitive processing unfolds automatically. As Metzinger contends, 'conscious thought is a subpersonal process'. Therefore, attributing high levels of MA to individuals is misguided [33, 46].

¹⁴ Metzinger suggests that attentional agency and cognitive agency are among the important psychological functions that underlie the capacity for MA, and which need to be defined and refined in a way that they can be empirically assessed.

Second, extending beyond the scope of what is defined here as MI, individuals are constantly exposed to an array of external and internal stimuli that may not necessarily engage their rational faculties, challenging the notion of a completely rational actor unaffected by subconscious forces. Environmental factors exert a continuous influence on individuals as embedded and embodied social beings. It is crucial to acknowledge the diverse range of stimuli—both external (such as air quality, social dynamics, and ambient noise) and internal (like hormonal fluctuations and fatigue)—that impact mental states, including moods, intentions, and decision-making processes. Additionally, even self-initiated actions which may not be perceived as directly influencing the mind (such as eating certain foods or exercising), can affect an individual's mind. Therefore, this analysis refrains from endorsing the concept of entirely autonomous, rational actors immune to mental influences as a basis for the normative evaluation of MIs.

Third, MA should be seen as a capacity that has to be developed and can be lost. Not all individuals possess the same level of mental abilities necessary for critical reflection on induced mental changes – ultimately rendering some (groups of individuals) more susceptible to UMIs. Children, for instance, lack fully developed mental capacities required to identify, evaluate, and endorse mental alterations induced by external agents. This is evident in discussions on indoctrination in children, particularly in its contrast to moral education. Ideas, beliefs, and thoughts are imparted to children by parents, teachers, and other caregivers in a manner that may not fully engage rational capacities, as these are still in the process of development. This creates a delicate tension: while it clearly demonstrates that children are more vulnerable to UMIs, the developmental nature of their mental abilities does not make every MI involving children inherently morally problematic [53]. The criterion of MA thus appears more challenging to apply in contexts involving young children. Similarly, individuals suffering from dementia or brain lesions gradually lose the cognitive abilities necessary for forming new values, beliefs, or long-term memories [54]. While these individuals retain essential human qualities such as emotions, social sensitivity, affection, self-expression, and humour [55] their abilities required for identifying and critically reflecting on interventions into their mental sphere deteriorate.

Despite these considerations, it remains reasonable to acknowledge individuals' partial control over their *forum internum* [56, 57]. Therefore, it is both justified and crucial to examine situations where an individual's capacity for MA, albeit inherently limited (to a varying degree in different individuals), is intentionally bypassed, distorted, or disarmed by a malicious agent attempting to alter their mind. Moreover, the fact that certain groups possess underdeveloped mental abilities should be considered as relevant contextual factor in the case-by-case moral assessment of MIs.

c) Behavioural autonomy versus mental autonomy

The definition of MA focuses solely on the abilities necessary to exert control over one's mind and does not extend to behavioural autonomy. The close interconnection between behavioural and mental autonomy, however, complicates the strict dichotomy between the two. Many instances of undue influence on personal behaviour, often termed behavioural manipulation, involve distorting and disarming individuals' mental abilities and exploiting their cognitive or emotional vulnerabilities, thus impairing MA [58]. However, there are actions that impact behavioural autonomy without necessarily affecting MA. For example, cases of persuasion through peer pressure, where an individual feels compelled to stay for another drink due to a strong group dynamic, may affect their behavioural autonomy. They may feel constrained in expressing their thoughts or independently determining their behaviour based on their inner motivations. Nevertheless, their ability to critically reflect on the influence exerted by their friends' persuasive encouragement on beliefs and intentions remains intact.

Within the broader model of personal autonomy, MA precedes behavioural autonomy. It concerns the capacity for self-determination over mental states, processes, and events which ultimately drive and shape external behaviour. Embodied, embedded, and enactive accounts of the mind would challenge a strict dichotomy, as they do not conceive of the relationship between the mental and behavioural levels as causal and unidirectional, or even just dual. This article does not take a stance in this debate. It considers MA as part of the broader concept of personal autonomy, as outlined by Dworkin. It does, however, not posit

mental states as the ultimate causal sources for behaviour, acknowledging that a one-way determining relationship between the internal and external realms is misleading. Without delving into a detailed philosophical exploration, it recognises the bidirectional influence between speech and behaviour on the one hand and mental processes on the other. Nonetheless, when evaluating whether MI respect the integrity of the mental abilities underlying MA, this bidirectionality is not of decisive importance.

d) Rationale

Given that the mind is the most intimate aspect of an individual, embodying all human experience including perceptions, emotions, and actions, it is evident that safeguarding it from unwarranted intrusions is paramount. Historically, unsolicited meddling with others' beliefs and behaviour has been deemed morally problematic throughout ethical discourse [59]. But why is MA a suitable anchor point for this moral judgement? The concept of MA, as developed on the basis of Dworkins' account of personal autonomy, highlights the idea that autonomy requires the mental (and/or cerebral) space free from external disturbance where individuals are free to reflect (and ultimately evaluate and endorse mental states). This form of mental self-determination forms the fundament of human dignity as in Kant's metaethics [45]. Within his moral framework, there is a fundamental presumption that thoughts, beliefs, and actions are, to some extent, shaped by individual choice and rationality [60]. An individual's mind deserves inherent respect owing to their indispensable role in forging a meaningful life as an autonomous human being. Consequently, the moral imperative to uphold human dignity necessitates the respect for MA of each individual involved in interpersonal interaction. Interfering with someone's mind should be done in a manner that acknowledges their capacity to evaluate their mental states and shape them in rapport to their environment and within the broader scheme of their lived experience. This approach facilitates the flourishing of the individual and the cultivation of a meaningful life. When MIs sideline the individual's capacity for mental self-determination, the individual becomes instrumentalised in their interaction with the agent, generally for the benefit of the latter. This would involve a fundamental violation of the subject's human dignity.

It is crucial to recognise that while MA is deeply rooted in human dignity, it does not necessarily override other moral values that also stem from human dignity. Values such as integrity and well-being, for instance, can sometimes clash with MA. In certain scenarios, such as in psychiatric contexts, the imposition of neuropharmaceuticals may compromise MA, yet simultaneously bolster the integrity and well-being of the individual affected. Similarly, consider governmental initiatives nudging individuals toward healthier or safer lifestyle choices. While this might encroach upon MA, it can significantly enhance the health and overall quality of life for many—a commendable aim that is also rooted in respect for human dignity. This is an important consideration that will profoundly affect the question as to how MA should be protected within human rights law (*infra*).

Bypassing, Distorting, and Disarming the Mental Abilities

UMIs are MIs that occur when an agent intentionally undermines a subject's MA by bypassing, distorting, or disarming the mental abilities underlying MA. These three types of interferences with one's mental abilities will be outlined separately, although their distinctions are often not clear-cut. Moreover, UMIs typically involve tactics that span across multiple categories. For instance, in cult indoctrination, agents will often aim to both distort and disarm the subject's abilities underlying MA.

The undermining of MA essentially represents an abuse of a power disparity. Various contextual factors elevate an agent into a position of power over the subject, enabling them to intentionally bypass, distort, or disarm the subject's relevant mental abilities in order to alter their mind as desired. Identifying these factors is crucial. These factors can be of various natures, including technical (i.e., the power position is inherent to availability of the means used for a MI, e.g. the use of neuropharmaceuticals, methods for channelling subliminal stimuli, or AI algorithms), social (i.e., the power position originates in hierarchical structures like employer-worker, teacher-student, government-citizen), and psychological (i.e., a power position exists in the light of the subject's socio-psychological vulnerabilities). It is paramount to recognise that the assessment of whether a MI ultimately bypasses, distorts, or disarms a subject's capacity to critically appreciate mental alterations must be conducted on

a case-by-case basis. The criterion of bypassing, distorting, or disarming MA cannot decisively categorise general types of MI as UMI or AMI, but requires contextual evaluation.

a) Bypassing MA

MI bypass MA when they alter mental states without engaging in any way with the subject's mental functions, and thus actively circumvent the subject's capacity to identify, evaluate, and endorse the mental changes that are being induced. Direct interventions that alter a subject's mental states by chemical or electrical modifications of their brain activity, fall into this category.¹⁵ These MIs completely side-line perception and cognition, and thus the necessary features for a subject to identify and evaluate reasons for the alteration of their (first-order) mental states.

Apart from direct interventions, however, there also is a grey zone. Subliminal advertising, for instance, could be considered to bypass one's capacity for MA. Subliminal persuasion relies on the idea of perception without awareness. It entails the alteration of mental processes of a subject through exposure to stimuli that are not consciously perceived [61]. While the effectiveness of subliminal persuasion is subject of debate within psychological literature, a vast amount of scientific research acknowledges the potential of subliminal stimuli as a means of interfering with individual's minds [62]. Due to the reliance on auditory or visual stimuli, subliminal advertising in principle proceeds through perceptual pathways. Nonetheless, since it escapes conscious perceptual and, importantly, cognitive processing, subliminal persuasion gives the agent the power to influence the mental states of the subject without them recognising being influenced, and to bypass any opportunity for the subject to critically evaluate reasons for the newly incited emotions, intentions, or beliefs.

b) Distorting MA

In addition, there are ways of intervening in the mind that, while engaging perceptual and cognitive pathways and therefore not fully bypassing a

subject's abilities for critical deliberation, alter the mind through distorting these abilities. This category of UMI is closest to what is generally understood as manipulation or brain-washing in contexts like (cult) indoctrination. These MIs typically rely on exploiting the subject's psycho-social vulnerabilities responsible for an innate susceptibility of having their MA distorted.¹⁶ For instance, mental characteristics and predispositions stemming from a subject's personal history (e.g. childhood trauma such as abuse or neglect that affects adults psychiatric status or financial and educational functioning [63], the lack of social support system, or personality disorders [64]) may inherently weaken the robustness of the mental capacities underpinning MA. Similarly, the developmental status of a subject will be of determining importance. Children's underdeveloped capacities for rational deliberation and critical assessment, combined with the position of dependency they generally find themselves in towards others, make them particularly susceptible to interferences that distort their psychological processes [65]. For instance, young children are particularly vulnerable to the persuasive effects of advertising [66]. Furthermore, various contextual factors can transform education for children, which *prima facie* appears morally unproblematic, into an UMI because children's developing capacities for MA are distorted. In short, when agents find themselves in a position of power towards a subject due to societal, institutional, or structural factors, or through exercising force, techniques are at their disposal to intervene in and control the mind of the subject by taking advantage of existing vulnerabilities to distort their abilities to critically engage with mental changes.¹⁷

A stark example of MIs distorting MA is that of 'brainwashing', 'reeducation', or 'psychological coercion' of prisoners of war [67]. These practices employ violent techniques such as isolation, physical punishment (and torture), and sleep deprivation in order to

¹⁵ This category of direct MIs is referred to by the SR as tools for 'modification' of thoughts.

¹⁶ This does not suggest that certain individuals are immune to distortions of their mental abilities to the degree that they are incapable of rationally reflecting on mental changes brought about by a MI.

¹⁷ To substantiate and coherently delineate individuals' susceptibility and vulnerability to distortions of MA, further exploration of behavioural sciences — particularly social and developmental psychology — and sociology is essential. Additionally, as indicated above, a crucial aspect of this effort lies in identifying the core mental abilities that underpin MA.

destabilise and distort a subject's cognitive and affective processes. This allows for the alteration of their mind in a manner that, although passing through cognition and perception, proceeds in such a way that the subject's mental abilities are unable to critically identify, evaluate, and endorse newly imposed thoughts, beliefs, intentions, or emotions. Similar tactics are used in cult indoctrination. Disconnection from familiar environments, deceit, fostering of chaos, isolation from society, suppression of dissent, instilling guilt, providing affection contingent on obedience, issuing threats of harm, and instilling fear are among the tactics cult leaders deploy in order to distort individuals' psychological capacities, thereby undermining their ability for mental self-determination [27].

c) Disarming MA

A third way MIs inappropriately interact with a subject's MA is by disarming the mental abilities that facilitate it, rather than affecting their integrity. The subject's mental faculties for critically evaluating mental changes are engaged in a manner that renders them ineffective and obstructs meaningful self-determination at a mental level. The agent manipulates external factors to ensure that the mental abilities underpinning MA cannot be utilised to critically evaluate reasons that may support mental changes. This does not necessarily circumvent the capacity for reflection entirely; rather, it manipulates the environment to predetermine the progression of the subject's mental states.

A prominent, yet complex example is coercion. Coercion, in a philosophical [68] as well as a legal [1] sense, generally refers to influencing a person to act against their will through the use or threat of force.¹⁸ It can be argued that coercion does not necessarily constitute a MI. It represents an overt form of influence where an agent exerts coercive pressure at the level of the *forum externum*, imposing constraints on behaviour and speech, without necessarily affecting

the deeper level of the *forum internum* [71].¹⁹ For instance, an employee is coerced by their employer to be loyal to the ruling monarch under threat of losing their job. If the threat is sincere, the employee has two options. They can either pretend to share the employer's view on the monarchy by exhibiting this view in their behaviour or speech (by, for instance, participating in the greeting of the royal portraits and the singing of the national anthem every morning). They also remain free to quit their job. In both cases, the agent's actions did not necessarily result in actual change in the mind of the employee regarding their appreciation of the monarchy. The employee's opportunity to feel, deliberate, and reflect on their feelings, beliefs, and thoughts regarding the monarchy, without aligning them with those envisaged by the agent, remains intact.

Hence, one could argue that coercion does not constitute a MI since it may not directly alter the *forum internum*. However, it could also be argued that coercion leads to a state of mind the subject would not have reached without the coercive influence—for example, the decision to quit their job. Coercion highlights the complexities of how external pressures can shape individuals' beliefs and decision-making processes, even when they retain (a level of) MA.²⁰ Brain research has shown that coercive pressure does affect the brain processes underlying decision-making, and generates a cognitive mechanism that reduces the experience of agency and increases compliance [75]. There is therefore some form of influence on the mental processes of the individual. Coercive influence is not be a MI that as such bypasses or distorts the subjects' mental abilities for MA, but it may in many cases – which vary due to contextual factors such as the socio-psychological situation of the subject – prevent them from meaningfully engaging with

¹⁸ In a human rights context, the Human Rights Committee clarified that this includes threats of physical violence or penal consequences, along with limitations of access to healthcare, employment, education, or public life as a whole, as well as the promise of preferential treatment [69, 70].

¹⁹ These limits are the result of an elimination of acceptable alternatives for action and speech besides those predetermined by the agent [72, 73].

²⁰ Another consideration that complicates the moral analysis of the mental impact of coercion is the idea that the person subjected to coercion, who changes their beliefs or intentions due to the coercive actions, is able to rationalise these newly imposed beliefs or intentions. In that sense, they could be seen as critically endorsing the mental changes as they can identify, evaluate, and endorse these in the light of the understanding they have of the constraints on their freedom to choose the agent imposed on them [74].

the mental interference in a way that allows for mental self-determination.

In addition to coercion, information control constitutes another method of disarming an individual's MA. It functions as a means to neutralise MA, impeding the effective functioning of mental abilities for the identification and critical endorsement of mental influences. This typically requires a form of hierarchy. As Van Dijk points out in his analysis of manipulation through discourse, 'parents can manipulate their children because of their position of power and authority in the family, professors can manipulate their students because of their institutional position or profession and because of their knowledge, and the same is true for politicians manipulating voters, journalists manipulating the recipients of media discourse or religious leaders manipulating their followers' [76], p. 362. For instance, teachers have the power to impose mandatory participation in religious practices in schools for students. Combined with religious education focusing on one particular religion and exposure to religious symbols, this is in some instances labelled "indoctrination" that unduly influences or imposes ideas and beliefs onto children. These MIs rely on the control and directing specific input that engage these mental capacities. In certain situations, such as the special vulnerability of children, this may be considered as illegitimately neutralising their (still developing) capacity for MA, by intentionally obstructing the access to tools necessary for the critical appreciation of mental influences. Moreover, undue online influence includes practices like microtargeting, personalised search engines, and AI-driven social platforms [59]. These tools can be seen as a form of power abuse where agents push specific stimuli or information onto a subject in a way that does not necessarily bypass their cognition but overwhelms their abilities for mental self-determination with information designed to produce a predetermined mental outcome, while simultaneously restricting access to stimuli that would stimulate for critical reflection. The challenge in these cases lies in determining which forms of input manipulation actually disarm MA, as not every unannounced suggestion by a personalised search engine or advertisement necessarily disarms MA.

It is worth highlighting that coercion and information control are tools also used –among a wide array of methods– to distort one's capacity for MA. As

previously indicated, these methods are integral to indoctrination schemes aimed at altering the minds of cult members or re-educating prisoners of war. Therefore, even if not considered MIs on their own, they undoubtedly contribute to the distortion of the mental abilities underlying MA. Moreover, in contrast to the distinction made in this analysis, Bublitz advances an argument that would classify coercion as a MI that distorts the mental capacity for critical deliberation. In this view, the creation of a psychologically burdening situation allows for the exploitation of psychological weaknesses. The agent anticipates the subject to yield to coercive pressure to dissolve the untenable psychological stress and satisfy their psychological needs [77].

Concealment & Consent Concealment is often pinpointed as a pivotal aspect of morally objectionable methods of interfering with an individual's mental states, setting it apart from acceptable forms of persuasion [78, 79]. Morally objectionable MIs typically operate beyond the awareness of the subject as the agent aims to keep them in the dark about how they are impacting their minds. Nevertheless, while the subtle, covert, or concealed nature is often crucial for MIs to be effective, it is not a constitutive element of an UMI [80].

When UMIs occur through the electrical or chemical alteration of the subject's brain processes, the concealed character is not crucial for the effectiveness of the MI. Whether or not a subject knows they are subjected to brain stimulation, for instance, will not impact the fact that their mental processes are interfered with through electrical pulses in their brain. However, when the subject is aware, they have the opportunity to consent to the MI. Recognising individuals as autonomous beings with inherent dignity does not only entail the duty to abstain from MIs that bypass, distort, or disarm their MA, but also requires respect for their voluntary choice to be subjected to MIs that undermine MA in these ways. Within the limits of fundamental rights and freedoms of others, the autonomy of an individual to consent to MIs should prevail.²¹ People indeed occasionally

²¹ This consideration ties into a lively debate in ethical literature about human enhancement. The use of drugs and technology for human enhancement will become increasingly important in light of technological, scientific, and societal

forfeit their MA and consent to MIs that undermine MA. This is the case, for instance, when they agree to be subjected to neurotechnological and neuropharmacuetical treatments, or forms of psychotherapy. Individuals should be able to make positive choices for these MIs, which would only be morally unacceptable if no consent would be obtained. This is especially true considering that such MIs may, ultimately, have a positive effect on the individual's capacity for MA and overall well-being [81]. For instance, neuropharmaceuticals can have a beneficial effect on MA, by alleviating symptoms of clinical depression that reduce an individual's mental abilities necessary to critically reflect on the mental influences they are subjected to in their daily lives.

Consequently, individuals should be free to consent to receive, for instance, treatment that chemically alters their brain activity when they anticipate that these interventions will have a positive effect on their well-being – on a mental and physical level. While these interventions affect their mind in a way that may escape their control, the choice to opt for these treatment options should be open, given that the subject is accurately informed and willing to accept the risks and side-effects. Consent to such MIs should be freely given, informed, and specific. This is especially relevant and challenging when it concerns vulnerable persons. 'Free' means that consent is given voluntary and not obtained under coercive circumstances. For instance, in situations where employers encourage the use of TBS to enhance focus, the hierarchical dynamic may pressure workers into using TBS even without explicit obligation. This will compromise their ability to provide free consent. 'Informed' implies that the subject should be fully aware of the potential mental changes that may result from the MI. In cases where the mental effects of certain actions, such as neurotechnological interventions [39], may be unpredictable, the subject should be informed about these potential effects or informed that they are not foreseeable. 'Specific' indicates that the MI cannot extend beyond the intervention as described to the

subject. For instance, when an individual consents to the use of DBS, but the device is hacked and maliciously controlled by a third party, any resulting mental changes would not be covered by the initial consent.²² Another example is individuals who freely choose to participate in an educational program. They consent to being exposed to certain thoughts, ideas, and beliefs presented through specific educational methods aimed at engaging with their mental states. However, this consent does not extend to, for instance, re-education programs, as these may employ methods – such as physical violence – go beyond most traditional educational practices.²³

In cases whereby MA is distorted or disarmed, concealment will typically affect the success of a MI. For example, in online microtargeting, individuals are generally unaware of both the fact that, and the exact manner in which their online data is collected, processed, and aggregated to allow advertisers to more effectively influence them. If users would be made aware – through public education for instance – of being subjected to such online influences, they could enhance their ability to recognise and critically evaluate their impact on their mental sphere. Moreover, if website users would be required to explicitly consent to the use of their data for improving the microtargeting strategies to influence their online behaviour (after being appropriately informed), a similar effect may occur. An agent who discloses its intention to influence the subject's mental states, boosts the latter's capacity to effectively engage their mental abilities, enabling them to self-determine their thoughts, emotions, beliefs, and intentions.

An interesting question in the normative assessment of everyday MIs is whether there exists a prevailing, basic level of consent to certain established practices that constitute MIs. Revisiting the example of the profiling of website users: do people who browse online in 2025 implicitly accept that their online experiences are manipulated by digital algorithms? Similarly, when individuals visit a physical store, do they implicitly consent to being exposed to visual, auditory, and even olfactory stimuli designed

Footnote 21 (continued)

developments. A key ethical question in this regard is whether there should be limits to human enhancement, possibly based on principles such as human dignity, or whether individual autonomy should prevail. However, these important questions will not be addressed in this paper.

²² It has been shown that brain stimulation devices and Brain-Computer interfaces are vulnerable to hacking [82].

²³ It should be noted that it might in some cases be difficult to draw the line between acceptable and unacceptable education practices.

to influence their decision-making and nudge their behaviour? In other words: does engaging in established societal practices imply implicit consent to certain MIs? Or should there be a formal consent procedure that explicitly informs users not only about the collection of their online data but also about its use for AI-driven influencing strategies? This is an interesting line of inquiry when assessing established societal practices against the normative framework presented here.

Intent Intent is a fundamental aspect of every MI. However, for an MI to qualify as an UMI, a specific type of intent is required. To qualify as an UMI, the agent must intend to bypass, distort, or disarm the subject's capacity to identify and reflect on mental changes. Without intent, an UMI cannot occur [71]. This means that the agent must be aware that their method of influencing the subject's mind undermines their capacity for MA. For instance, forcible administering neuropharmaceuticals inherently bypasses a subject's MA. Choosing this method to alter a subject's mind indicates the intent to engage in an UMI (if no consent was obtained). However, if there has been a mix-up in an hospital treatment plan and a neuropharmaceutical were mistakenly given to a patient, the interference with that person's mental states would not constitute an UMI – and arguably not even a MI if the originally intended medicine had no psychoactive effects.

Nevertheless, it is useful to broaden the concept of intentionality to include 'non-accidental' actions [83]. This implies that an agent can intend to bypass, deteriorate, or disarm MA without fully understanding how the MI will specifically affect the subject's psychological processes. Agents do not necessarily need to actively believe that their actions will amount to an UMI, but upon reflection, they should recognise it as a foreseeable consequence of their actions within the broader context. For instance, an agent who designs and implements an online algorithm with the goal of steering users towards certain beliefs and feelings about veganism, without engaging their critical deliberation process, may not be fully aware of how the algorithm achieves this. Even when third parties purchase the same algorithm for political influencing, their use of the algorithm cannot simply be considered innocent ignorance [84]. While direct intent may

be more difficult to prove in such cases, the mere fact that the algorithm is designed to influence the beliefs, intentions, emotions, and decision-making processes of users in a way that exploits vulnerabilities and evades cognitive engagement indicates its potential to unduly influence people's mental states. The inherent functionality of the means for mental influencing renders the unacceptable character of MI non-accidental. An agent can 'be unconscious of the depth of their intention' [85], p.30, yet still act with that intention.

UMI and the Freedom of Thought

At present, the right not to have one's thoughts impermissibly altered remains highly opaque. There is a lack of established principles to determine whether a MI violates the FoT. However, when this right has been assessed by a court, some noteworthy conclusions were reached. For example, the courts consider power disparities when assessing whether there has been undue influencing of thought. In *Larissis and Others v. Greece*, the European Court of Human Rights (ECtHR) discussed the distinction between appropriate and inappropriate means of religious influencing.²⁴ In this decision, the power imbalance between a high-ranking officer attempting to convert hierarchical subordinate soldiers to Jehovah's Witness was an important factor in establishing improper proselytism [86]. Similarly, in *Kokkinakis v. Greece*, the ECtHR indicated that, when delineating between appropriate and inappropriate influencing, the vulnerable status of the subject and the use of violence and 'brainwashing' are significant considerations in determining whether there is a violation of the FoT [87]. Yet, while these rulings rightly highlight the importance of contextual factors, they do not indicate what 'brainwashing' means. They do not provide a detailed, universally applicable analysis, defining the decisive criteria for identifying impermissible influence. Therefore, further development of the FoT requires identifying the core conditions for interference with the right not to have one's thoughts altered, and determining which factors would indicate that a

²⁴ This case entailed an alleged violation of the right to religion (article 9 ECHR), but was decided on relevant elements for the entire freedom of the *forum internum*.

MI meets these conditions and exceed the limits of acceptable persuasion.

The normative framework for identifying UMIs as set out above can serve as a basis and guide ongoing international academic and policy efforts to interpret the FoT. Cases involving manipulation, coercion, or modification of thought should be conceptualised as abuses of power, where agents alter an individual's mind by using various tools and exploiting psychosocial vulnerabilities to bypass, distort, or disarm an individual's mental abilities essential for MA. These MIs interfere with the FoT. This approach aims to develop FoT with a central focus on protecting and promoting MA [48]. It is crucial to refine this normative framework by integrating insights from psychological research on the mental abilities that underlie MA and the methods of mental influence that interfere with them.

It is crucial to underscore that significant questions remain. What exactly constitutes 'thought'? Does it cover all mental states, or only specific thoughts, opinions, or beliefs that carry significant weight and coherence [2, 3]? Furthermore, how does the extended mind thesis impact the scope of the FoT [88]? This remains a topic of debate. Additionally, and closely linked to the analysis above, is the question whether the protection offered by the FoT against UMIs should be absolute. Labelling all UMIs as violations of the FoT would have profound implications, potentially categorising established societal practices such as (forced) psychiatric treatments or certain types of marketing strategies as human rights violations. Therefore, balancing individuals' interest in avoiding UMIs with broader societal interests poses considerable challenges in shaping the FoT. Should the law adopt a deontological approach, where principled reasons mandate the absolute protection of MA regardless of specific circumstances? This approach would advocate for preserving the absolute character of the FoT, thereby suggesting that UMIs are never justifiable. In that case, practices like forced psychiatric interventions might need to be completely banned, with significant consequences for healthcare and public safety frameworks. Alternatively, should the law take a more consequentialist stance, weighing the harm caused by violating an individual's MA against other potential harms and interests, including bodily integrity, personal well-being, and public safety? This approach would suggest a relative protection of MA, where UMIs may not always be covered by the absolute

protection of the FoT, but could instead fall under the relative protection provided by the rights to privacy and mental integrity [89, 90]. UMIs may be considered justifiable in cases where less intrusive measures – such as AMIs – are inadequate to achieve a legitimate goal. For example, persuading individuals to get vaccinated for public health reasons by nudging them towards this option, or medically intervening in a patient with an acute psychiatric disorder without their consent to guarantee public safety, could be seen as justifiable.

Probably, a combination of absolute and relative protections constitutes the most comprehensive and effective human rights approach. UMIs that profoundly impact an individual's fundamental interests should fall under the absolute protection of the FoT. This absolute protection should be supplemented by relative protection against UMIs that have a lesser personal impact and may be justifiable, when proportionate, in the light of a legitimate societal goal. For instance, a MI that disarms a person's capacity for MA may be a proportionate measure to pursue a legitimate goal (e.g., government nudging towards healthy lifestyle choices), but may violate one's right to mental integrity when pursued for other goals (e.g., nudging consumers solely for economic gain). Similarly, using neurostimulation devices to interfere with an imprisoned person's brain and affect their mental states could be considered disproportionate (due to violation of both mental and bodily autonomy), despite the potentially legitimate aim of ensuring public and personal safety, while pursuing the same goal through talk-therapy may be deemed proportionate. This relative protection against UMIs leaves room for a nuanced, case-by-case balancing of interests, and should be complemented by the absolute protection of the FoT when an individual's fundamental interests are violated in a way that would never be acceptable (e.g., impermissible interference with political opinions or sexual orientation). It is important to emphasise that this approach requires careful consideration of how the respective scopes of absolute and relative protection are interact.

Conclusion

The analysis suggested a guiding conceptual and normative framework for the moral evaluation of MIs. When individuals exert mental influence by

intentionally bypassing, distorting, or disarming another's MA, this form of mental influencing is morally problematic and constitutes an UMI. This framework needs further integration of insights from behavioural sciences (including social and developmental psychology) to identify and clarify the mental abilities underlying MA, and enabling individuals to identify, evaluate, and endorse changes to their mental states. This necessary refinement of the framework would allow for a more fine-grained outline of how UMIs operate in real-life cases, and which (groups of) individuals may be particularly susceptible. This further elaboration is needed for an effective operationalisation of the moral criterion of MA to apply in specific, real-life situations.

This normative framework for identifying UMIs should inform the interpretation of the FoT, which holds protection against impermissible alterations of thought. A case-by-case contextual assessment should enable the determination as to whether individuals abuse their position of power vis-à-vis another person in order to bypass, distort, or disarm their MA, and thereby violates their FoT. Yet, the question remains as to whether the FoT should categorically prohibit any UMI. From a deontological point of view, such a prohibition appears justified, as the inherent value of MA dictates that individuals should never be treated in a way that denies their autonomy and objectifies them in social interaction. Alternatively, a consequentialist approach may be deemed more pragmatic, as it would allow for UMIs under certain circumstances where a careful balancing of interests indicates that other fundamental interests rooted in human dignity –such as personal well-being and public safety– would justify infringing upon MA.

Ultimately, the conclusion of this analysis raises challenging legal and public policy questions about which forms of interference with the mind should be deemed unacceptable. Established practices such as forced psychiatric interventions may need to be completely banned in light of this analysis, with far-reaching consequences for current practices in healthcare and public safety law. Moreover, the operationalisation of MA may lead to the realisation that certain marketing strategies, currently deemed acceptable, actually constitute UMIs. As McCarthy-Jones notes, powerful financial interests inherent in the neoliberal market system may resist legal action based on these conclusions [91]. Are we prepared to reconsider

entrenched (capitalist) practices in the name of FoT, even if it means significant disruption to the existing economic framework? Similarly, political interests within social democratic systems may incite reluctance of stringently protecting MA, as interferences may contribute to the advancement of a social good. These considerations are crucial when further elaborating on the scope and meaning of key elements of the FoT. As aptly stated: 'If the first serious international attempt to define the right to freedom of thought does not alarm you, you're not paying attention' [86], p. 27.

Funding Fonds Wetenschappelijk Onderzoek, Grant no. 45218.

Declaration

Conflicts of Interest N.A.

References

1. UN Special Rapporteur on freedom of religion or belief. 2021. *Interim report on the freedom of thought*. UN Doc. A/76/380. Accessed on 27 May 2024 via <https://undocs.org/Home/Mobile?FinalSymbol=A%2F76%2F380&Language=E&DeviceType=Desktop&LangRequested=False>.
2. Ligthart, Sjors. 2020. Freedom of thought in Europe: Do advances in 'brain-reading' technology call for revision? *JLB* 7(1). <https://doi.org/10.1093/jlb/lsaa048>.
3. Istace, Timo. 2023. Protecting the mental realm: What does human rights law bring to the table? *NQHR* 41(4). <https://doi.org/10.1177/09240519231211823>.
4. Ligthart, Sjors, Bublitz, Jan-Christoph, Douglas, Thomas, Forsberg, Lisa, Meynen, Gerben. 2021. Rethinking the right to freedom of thought: A multidisciplinary analysis. *Human Rights Law Review* 22(4). <https://doi.org/10.1093/hrlr/ngac028>.
5. Bublitz, Jan-Christoph. 2020. The Nascent Right to Psychological Integrity and Mental Self-Determination. In *The Cambridge Handbook of New Human Rights*, eds. von Arnault Andreas, von der Decken Kerstin, and Susi Mart, 387–403. Cambridge: CUP. <https://doi.org/10.1017/9781108676106.031>.
6. Farah, Martha J. 2018. Socioeconomic status and the brain: Prospects for neuroscience-informed policy. *Nature Reviews Neuroscience* 19 (7): 428–438. <https://doi.org/10.1038/s41583-018-0023-2>.
7. Klein, Eran. 2020. Ethics and the emergence of brain-computer interface medicine. *Handbook of Clinical Neurology* 168: 329–339. <https://doi.org/10.1016/B978-0-444-63934-9.00024-X>.
8. Higgins, Edmund S., and Mark S. George. 2020. *Brain Stimulation Therapies for Clinicians*, 2nd ed. Washington DC: American Psychiatric Publishing.

9. Hildt, Elisabeth. 2006. Electrodes in the brain: Some anthropological and ethical aspects of deep brain stimulation. *IRIE* 5:33–39. <https://doi.org/10.29173/irrie193>.
10. Tan, Shawn Zheng Kai, Fung, Man-Lung, Koh, Junhao, Chan, Ying-Shing, and Lim, Lee Wie. 2020. The paradoxical effect of deep brain stimulation on memory. *Aging and Disease* 11(1):179–190. <https://doi.org/10.14336/AD.2019.0511>.
11. Elias, Gavin J.B., Alexandre Boutet, Suresh E. Joel, Jürgen. Germann, Dave Gwun, Clemens Neudorfer, et al. 2021. Probabilistic mapping of deep brain stimulation: Insights from 15 years of therapy. *Annals of Neurology* 89: 426–443. <https://doi.org/10.1002/ana.25975>.
12. Cinel, Caterina, Valeriani, Davide, and Poli, Riccardo. 2019. Neurotechnologies for human cognitive augmentation: current state of the art and future prospects. *Frontiers in Human Neuroscience* 13. <https://doi.org/10.3389/fnhum.2019.00013>.
13. Hossain, Soaad Qahhar, and Syed Ishtiaque Ahmed. 2020. Ethical analysis on the application of neurotechnology for human augmentation in physicians and surgeons. *Proceedings of the Future Technologies Conference (FTC)* 3: 78–99. https://doi.org/10.1007/978-3-030-63092-8_6.
14. Zawadski, Przemyslaw, and Agnieszka K. Adamczyk. 2021. Personality and authenticity in light of the memory-modifying potential of optogenetics. *AJOB Neuroscience* 12 (1): 3–21. <https://doi.org/10.1080/21507740.2020.1866097>.
15. Pugh, Jonathan, Laurie Pycroft, Anders Sandberg, Tipu Aziz, and J. Savulescu. 2018. Brainjacking in deep brain stimulation and autonomy. *Ethics and Information Technology* 20 (3): 219–232. <https://doi.org/10.1007/s10676-018-9466-4>.
16. Bublitz, Jan-Christoph. 2023. What an international declaration on neurotechnologies and human rights could look like: Ideas, suggestions, desiderata. *AJOB Neuroscience* 15 (2): 96–112. <https://doi.org/10.1080/21507740.2023.2270512>.
17. Weller, Chuck. 2011. Forced administration of antipsychotic drugs to civilly committed mental patients in Nevada: A remedy without clear statutory authorisation. *Nevada Law Journal* 11: 759–769.
18. <https://www.rxlist.com/prolixin-drug.htm> (Accessed 1 July 2024).
19. Floridi, Luciano. 2024. Hypersuasion – On AI's persuasive power and how to deal with it. *Philosophy & Technology* 37(64). <https://doi.org/10.1007/s13347-024-00756-6>.
20. Zarouali, Brahim, Tom Dobber, Guy de Pauw, and Claes de Vreese. 2020. Using a personality-profiling algorithm: Assessing the persuasion effects of personality-tailored ads on social media. *Communication Research* 48 (8): 1066–1091. <https://doi.org/10.1177/0093650220961965>.
21. Coglianese, Cary. 2018. Optimizing regulation for an optimizing economy. *University of Pennsylvania Journal of Law and Public Affairs* 4(1). <https://ssrn.com/abstract=3300304>.
22. Channel 4 News Investigations Team. 2020. Revealed: Trump campaign strategy to deter millions of black Americans from voting in 2016. <https://www.channel4.com/news/revealed-trump-campaignstrategy-to-deter-millions-of-black-americans-from-voting-in-2016>. Accessed 2 July 2024.
23. Zuiderveen, Frederik J., Judith Mölier, S. Kruikemeier, Ronan Fathaigh, Kristina Irion, T. Dobber, et al. 2018. Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review* 14 (1): 82–96.
24. Krotzek, Lennart J. 2019. Inside the voter's mind: The effect of psychometric microtargeting on feelings toward and propensity to vote for a candidate. *International Journal of Communication* 13: 3609–3629.
25. Büchi, Moritz, Eduard Fosch-Villaronga, Christoph Lutz, Tamò-Larrieux. Aurelia, Shruthi Velidi, and Salome Viljoen. 2020. The chilling effects of algorithmic profiling: Mapping the issues. *Computer Law & Security Review* 36. <https://doi.org/10.1016/j.clsr.2019.105367>.
26. Hassan, Steven A. 2020. The BITE model of authoritarian control: Undue influence, thought reform, brainwashing, mind control, trafficking and the law. *Doctoral dissertation, Fielding Graduate University*. <https://freedomofminds.com/wp-content/uploads/2021/01/Dr-Hassan-Dissertation-Published.pdf>. Accessed 2 July 2024.
27. Anderson, Susan, and Philip Zimbardo. 1980. Resisting mind control. *USA Today* 109 (2426): 44–47.
28. Ayrapetova, Alyona G. 2020. Indoctrination as a mechanism of psychological manipulation in the process of involvement in destructive religious organisations. *European Journal of Research and Reflection in Educational Sciences* 8 (11): 17–23.
29. Yaffe, Gideon. 2003. Indoctrination, coercion, and freedom of will. *Philosophy and Phenomenological Research* 67 (2): 335–356.
30. Hand, Michael. 2004. The problem with faith schools: A reply to my critics. *Theory and Research in Education* 2 (3): 343–353. <https://doi.org/10.1177/1477878504046529>.
31. Coons, Christian, and Weber, Michael (eds). 2014. *Manipulation: theory and practice*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199338207.001.0001>.
32. Levy, Neil. 2007. *Neuroethics. Challenges for the 21st century*. New York: Cambridge University Press.
33. Bublitz, Jan-Christoph., and Reinhard Merkel. 2014. Crimes against minds: on mental manipulations, harms and human right to mental self-determination. *Criminal Law and Philosophy* 8: 51–77. <https://doi.org/10.1007/s11572-012-9172-y>.
34. Bublitz, Jan-Christoph., and Reinhart Merkel. 2009. Autonomy and authenticity of enhanced personality traits. *Bioethics* 23 (6): 360–374. <https://doi.org/10.1111/j.1467-8519.2009.01725.x>.
35. Bublitz, Jan-Christoph. 2020. Why Means Matter. Legally relevant differences between direct and indirect interventions into other minds. In *Neurointerventions and the law*, eds. Vincent, Nicole A., Nadelhoffer, Thomas, and McCay, Allan, 49–88. New York: Oxford University Press. <https://doi.org/10.1093/oso/9780190651145.003.0003>.
36. Focquaert, Farah, and Maartje Schermer. 2015. Moral enhancement: Do means matter morally? *Neuroethics* 8: 139–151. <https://doi.org/10.1007/s12152-015-9230-y>.
37. Shaw, Elisabeth. 2014. Direct brain interventions and responsibility enhancement. *Criminal Law*

- and *Philosophy* 8: 1–20. <https://doi.org/10.1007/s11572-012-9152-2>.
38. Birks, David, and Alena Buyx. 2018. Punishing intentions and neurointerventions. *AJOB Neuroscience* 9: 133–143. <https://doi.org/10.1080/21507740.2018.1496162>.
 39. Shaw, Elisabeth. 2015. The use of brain interventions in offender rehabilitation programs: should it be mandatory, voluntary, or prohibited? In *Handbook of Neuroethics*, eds. Clausen, Jens, and Levy, Neil, 1381–1398. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-4707-4_169.
 40. Douglas, Thomas. 2022. If nudges treat their targets as rational agents, nonconsensual neurointerventions can too. *ETMP* 25 (2): 369–384. <https://doi.org/10.1007/s10677-022-10285-w>.
 41. Douglas, Thomas. 2018. Neural and environmental modulation of motivation: What's the moral difference? in *Modulation of motivation. Treatment for crime: Philosophical essays on neurointerventions in criminal justice*, eds. David Birks, and Douglas, Thomas, 208–223. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198758617.003.0012>.
 42. Metzinger, Thomas K. 2015. M-Autonomy. *Journal of Consciousness Studies* 22 (11–12): 270–302.
 43. Dworkin, Gerald. 1976. Autonomy and behaviour control. *Hasting Centre Report* 6 (1): 23–28. <https://doi.org/10.2307/3560358>.
 44. Frankfurt, Harry G. 1971. Freedom of the will and the concept of the person. *Agency and Responsibility* 68 (1): 5–20. <https://doi.org/10.2307/2024717>.
 45. Lavazza, Andrea, Giorgi, Rodolfo. 2023. Philosophical foundation of the right to mental integrity in the age of neurotechnologies. *Neuroethics* 16(10). <https://doi.org/10.1007/s12152-023-09517-2>.
 46. Metzinger, Thomas K. 2013. The myth of cognitive agency: Subpersonal thinking as a cyclically recurring loss of mental autonomy. *Frontiers in Psychology* 4. <https://doi.org/10.3389/fpsyg.2013.00931>.
 47. Calleja, De., and Mirja Pérez. 2019. Autonomy and Indoctrination: Why we need an emotional condition for autonomous reasoning and reflective endorsement. *Social Philosophy and Policy* 36 (1): 192–210. <https://doi.org/10.1017/S026505251900027X>.
 48. McCarthy-Jones, Simon. 2019. The autonomous mind: The right to freedom of thought in the twenty-first century. *Frontiers in Artificial Intelligence* 2. <https://doi.org/10.3389/frai.2019.00019>.
 49. Carter, Adam. 2021. Varieties of (Extended) thought manipulation. In *The Law and Ethics of Freedom of Thought (vol. 1)*, eds. Blitz, Marc Jonathan, and Bublitz, Jan-Christoph, 291–305. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-84494-3_10.
 50. Goodin, Robert E. 1980. *Manipulatory politics*. New Haven: Yale University Press.
 51. Scanlon, Thomas M. 1998. *What we owe to each other*. Cambridge: Harvard University/Belknap Press. <https://doi.org/10.2307/j.ctv134vmrn>.
 52. Owen, Gareth S., Fabian Freyenhagen, Genevra Richardson, and Matthew Hotopf. 2009. Mental capacity and decisional autonomy: An interdisciplinary challenge. *Inquiry* 52 (1): 79–107. <https://doi.org/10.1080/00201740802661502>.
 53. Merry, Michael S. 2005. Indoctrination, moral instruction and nonrational beliefs: A place for autonomy? *Educational Theory* 55 (4): 399–420. <https://doi.org/10.1111/j.1741-5446.2005.00002.x-i.1>.
 54. Huang, Yuanyuan, Hui Liu, and Cong Yali. 2022. Is deception defensible in dementia care? A care ethics perspective. *Nursing Ethics* 29 (7–8): 1589–1599. <https://doi.org/10.1177/096973302210923>.
 55. Kitwood, Tom, and Kathleen Bredin. 1992. Towards a theory of dementia care: Personhood and well-being. *Ageing & Society* 12 (3): 269–287. <https://doi.org/10.1017/s0144686x0000502x>.
 56. Swaine, Lucas. 2020. Freedom of thought in political history. In *The Law and Ethics of Freedom of Thought (vol. 1)*, eds. Blitz, Marc Jonathan, and Bublitz, Jan-Christoph, 1–26. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-84494-3_1.
 57. Swaine, Lucas. 2020. *Ethical autonomy; the rise of self-rule*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780190087647.001.0001>.
 58. Filho, Flores, Edgar Gastón Jacobs, and Marina Castro Firmo. 2023. Human dignity and neurorights in the digital age. *Brazilian Journal of Law, Technology and Innovation* 1 (2): 87–107. <https://doi.org/10.59224/bjlti.v1i2.87-107>.
 59. Ienca, Marcello. 2023. On artificial intelligence and manipulation. *Topoi* 42 (3): 833–842. <https://doi.org/10.1007/s11245-023-09940-3>.
 60. Nuffield Council on Bioethics. 2002. *Genetics and human behaviour: The ethical context*. chrome-extension://efaidnbnmnibpcjpcglclefindmkaj/<https://www.nuffieldbioethics.org/assets/pdfs/Genetics-and-human-behaviour-short-guide.pdf> (Accessed on 1 July 2024).
 61. Moore, Timothy E. 1988. The case against subliminal manipulation. *Psychology & Marketing* 5 (4): 297–316. <https://doi.org/10.1002/mar.4220050403>.
 62. Dijksterhuis, Ap, Aerts, Henk, and Smith, Pamela K. 2006. The power of the subliminal: On subliminal persuasion and other potential applications. In *The new unconscious*, eds. Hassin, Ran R., Uleman, James S., and Bargh, John A., 77–106. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195307696.003.0005>.
 63. Copeland, William, Shanahan, Lilly, Hinesley, Jennifer, et al. 2018. Association of childhood trauma exposure with adult psychiatric disorders and functional outcomes. *JAMA Network Open* 1(7). <https://doi.org/10.1001/jamanetworkopen.2018.4493>.
 64. Reid, Joan, Rachael Haskell, Christina Dillahunt-Aspillaga, and Jennifer Thor. 2013. Contemporary review of empirical and clinical studies of trauma bonding in violent or exploitative relationships. *International Journal of Psychiatry Research* 8 (1): 37–73.
 65. Cuypers, Stefaan E. 2010. Autonomy in R.S. Peters' educational theory. *Philosophy of education* 43 (1): 189–207. <https://doi.org/10.1111/j.1467-9752.2009.00733.x>.
 66. Radensky, Jenny, Chassiakos, Yolanda L.R., Ameenuddin, Nusheen, and Navsaria, Dipesh. 2020. Digital Advertising to children. *Pediatrics* 146(1). <https://doi.org/10.1542/peds.2020-1681>.

67. Hopper, Elizabeth, and Jose Hidalgo. 2006. Invisible chains: Psychological coercion of human trafficking victims. *IHRLR* 18: 185–210.
68. Lunenburg, Fred C. 2012. Power and leadership: An influence process. *International Journal of Management, Business, and Administration* 15 (1): 1–9.
69. Human Rights Committee. 16 July 2003. *Communication no. 878/199*, UN Doc. CCPR/C/78/D/878/1999, para. 7.2. <https://digitallibrary.un.org/record/503369>. Accessed 2 July 2024.
70. Human Rights Committee. 1993. *General Comment No. 22*. UN Doc. HRI/GEN/1/Rev.9, p. 205, para. 5. <https://www.ccprcentre.org/ccpr-general-comments>. Accessed 2 July 2024.
71. Baron, Marcia. 2003. Manipulativeness. *Proceedings and Addresses of the American Philosophy Association* 77 (2): 37–54. <https://doi.org/10.2307/3219740>.
72. Barnhill, Anne. 2014. What is manipulation? In *Manipulation: Theory and Practice*, eds. Coons Christian, and Weber Michael, 51–72. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199338207.003.0003>.
73. Wood, Allen W. 2014. Coercion, manipulation, exploitation. In *Manipulation: Theory and Practice*, eds. Coons, Christian, and Weber Michael, 17–50. <https://doi.org/10.1093/acprof:oso/9780199338207.003.0002>.
74. Pugh, Jonathan. 2020. *Autonomy, rationality, and contemporary bioethics*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198858584.001.0001>.
75. Caspar, Emilie A., Julia F. Christensen, Axel Cleeremans, and Patrick Haggard. 2016. Coercion changes the sense of agency in the human brain. *Current biology* 26 (5): 585–592. <https://doi.org/10.1016/j.cub.2015.12.067>.
76. Dijk, Van, and A. Teun. 2006. Discourse and manipulation. *D&S* 17 (3): 359–383. <https://doi.org/10.1177/0957926506060250>.
77. Bublitz, Jan-Christoph. 2020. Freedom of Thought as an international human right: Elements of a theory of a living right. In *The law and ethics of freedom of thought (vol. I)*, eds. Blitz, Marc Jonathan, and Bublitz, Jan-Christoph, 49–101. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-84494-3_1.
78. Lau, Stephan. 2022. The good, the bad, and the trade-craft: HUMINT and the ethics of psychological manipulation. *Intelligence and National Security* 38 (4): 592–610. <https://doi.org/10.1080/02684527.2022.2129159>.
79. Süsner, Daniel, Beate Roessler, and Helen Nissenbaum. 2019. Online manipulation: Hidden influences in a digital world. *Georgetown Law Technology Review* 4 (1): 1–45.
80. Klenk, Michael. 2022. (Online) Manipulation: Sometimes hidden, always careless. *Review of Social Economy* 80 (1): 85–105. <https://doi.org/10.1080/00346764.2021.1894350>.
81. Mackenzie, Catriona, and Walker, Mary. 2014. Neurotechnologies, personal identity, and the ethics of authenticity. In *handbook of neuroethics*, eds. Clausen, Jens, and Levy, Neil, 373–392. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-4707-4_10.
82. Ienca, Marcello, and Haselager Pim. 2016. Hacking the brain: Brain-computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology* 18 (2): 117–129. <https://doi.org/10.1007/s10676-016-9398-9>.
83. Klenk, Michael. 2024. Ethics of generative AI and manipulation: A design-oriented research agenda. *Ethics and Information Technology* 26: 1–15. <https://doi.org/10.1007/s10676-024-09745-x>.
84. Jongepier, Fleur, and Wieland, Jan Willem. 2022. Micro-targeting people as a mere means. In *the philosophy of online manipulation*, eds. Klenk, Michael and Jongepier, Fleur, 156–175. New York: Routledge. <https://doi.org/10.4324/9781003205425-10>.
85. Dunstan, James. 2023. *Manipulation and influence: A trickery account of manipulation applied to three scopes*. A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, The University of Sheffield. <https://etheses.whiterose.ac.uk/33436/> (Accessed on 1 July 2024).
86. ECtHR. 24 February 1998. *Larissis and others v. Greece*. App. No. 140/1996/759/958–960.
87. ECtHR. 25 May 1993. *Kokkinakis v. Greece*. App. No. 14307/8.
88. Blitz, Marc J. 2010. Freedom of thought for the extended mind: cognitive enhancement and the constitution. *Wisconsin Law Review* 4: 1049–1118.
89. Ligthart, Sjors, Bublitz, Jan-Christoph, Douglas, Thomas, Forsberg, Lisa, and Meynen, Gerben. 2022. Rethinking the right to freedom of thought: A multidisciplinary analysis. *Human Rights Law Review* 22(4). <https://doi.org/10.1093/hrlr/ngac028>.
90. Istace, Timo. 2024. Human rights law: An incomplete but flexible framework to protect the human mind against neurotechnological intrusions. *LIT* 16 (1): 309–340. <https://doi.org/10.1080/17579961.2024.2313796>.
91. McCarthy-Jones, Simon. 2023. *Freethinking*. London: Oneworld.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Timo Istace is a PhD researcher in the field of Human Rights Law, Health Law, and Legal Philosophy. He specialises in exploring the implications of emerging technologies, with a particular focus on neurotechnology, within these domains.