

## Persuasive Technologies and the Right to Mental Liberty

### *The ‘Smart’ Rehabilitation of Criminal Offenders*

*Sjors Ligthart, Gerben Meynen and Thomas Douglas*

#### 3.1 INTRODUCTION

Every day, millions of people use mobile phones, play video games and surf the Internet. It is thus important to determine how technologies such as these change what people think and how they behave. This is a central issue in the study of persuasive technologies. ‘Persuasive technologies’ (henceforth ‘PTs’) are digital technologies, such as mobile apps, video games and virtual reality systems, which are deployed for the explicit purpose of changing attitudes and/or behaviours, without using coercion, deception or extreme forms of psychological manipulation (such as hypnosis or indoctrination), and without exerting a direct (not psychologically mediated) physical effect on the brain.<sup>1</sup> Typically, PTs employ strategies such as prompting, information provision, and encouragement.

Since many important determinants of wellbeing, such as health, pollution, climate change and crime, have a large behavioural component, and since PTs are typically intended to *change behaviour*, some hold that PTs could potentially have large benefits, both for the persuaded individuals and for society at large.<sup>2</sup> For example, eco-feedback systems persuade car drivers to use less fuel by showing images of green leaves, and health apps might motivate users to exercise regularly, quit smoking, and stop drinking.<sup>3</sup> The mobile app Sobriety Counter helps users to stop

Thomas Douglas thanks the European Research Council (grant number 819757) for their funding.

<sup>1</sup> B. J. Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do* (Amsterdam/Boston: Morgan Kaufmann Publishers, 2003), pp. 1, 15–16 and 38–39; J. Smids, *Persuasive Technology, Allocation of Control, and Mobility: An Ethical Analysis* (Eindhoven: Eindhoven University of Technology, 2018), pp. 72, 93; N. Jacobs, ‘Two Ethical Concerns about the Use of Persuasive Technology for Vulnerable People’ (2019) *Bioethics* 1–3; J. M. Byrne and A. Pattavina, ‘Technological Innovation and Offender Reentry’ in: S. Leman-Langois (ed.) *Technocrime, Policing and Surveillance* (New York: Routledge, 2013), pp. 126–127; Y. Chow, W. Susilo, J. Phillips, J. Baek and E. Vlahu-Gjorgievska, ‘Video Games and Virtual Reality as Persuasive Technologies for Health Care: An Overview’ (2017) 8 *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* 18–35; L. Chittaro and N. Zangrando, ‘The Persuasive Power of Virtual Reality: Effects of Simulated Human Distress on Attitudes towards Fire Safety’ in: T. Ploug, P. Hasle and H. Oinas-Kukkonen (eds.) *Persuasive Technology: PERSUASIVE 2010. Lecture Notes in Computer Science*, vol. 6137 (Berlin, Heidelberg: Springer, 2010). We are not aware of others having explicitly excluded direct brain interventions or extreme forms of manipulation from the class of persuasive technologies. However, we think something like these exclusions are implicitly endorsed by most.

<sup>2</sup> Smids (note 1), pp. 3–4; R. Orji and K. Moffatt, ‘Persuasive Technology for Health and Wellness: State-of-the-Art and Emerging Trends’ (2018) 24(1) *Health Informatics Journal* 66–91.

<sup>3</sup> Smids (note 1), p. 4. For a review on PT for health and wellness, see Orji and Moffatt (note 2), pp. 66–91.

drinking alcohol by employing motivational techniques such as visualising the money a user saves by not drinking and providing health statistics about drinking.<sup>4</sup>

While PTs could, in various ways, contribute to solving different societal problems, they also raise legal and ethical questions, for instance regarding privacy and autonomy.<sup>5</sup> Since PTs typically aim to *change* what their users *think*, one legal-ethical question that arises is: do PTs infringe a legal or moral right to *mental liberty*, understood here as a right to freedom from (certain kinds of) interferences with one's mind?<sup>6</sup> In what follows we focus on the legal half of this question since, although neither a legal nor a moral right to mental liberty has been fully worked out, the possibility that we might enjoy a legal right to mental liberty has at least garnered significance support.<sup>7</sup>

In cases where a person validly consents to the use of a PT, it is doubtful that there would be any infringement of such a right. In such cases, we might think that the right has been partially waived by the act of consenting. It is in cases where there is some kind of external pressure on a person to use a PT, such that the validity of their consent is in doubt, that concerns regarding mental liberty come to the fore.<sup>8</sup> If such 'pressured use' of PTs will ever be deployed, it is likely to be in the field of criminal justice, where pressure or even full-blown coercion are commonly used to change an individual's mind and behaviour, for example, by commitment to a forensic mental institution, or by requiring participation in a sex offender treatment programme as part of a criminal sentence. Moreover, focusing on the use of PTs in criminal justice is also interesting because this is arguably a context in which infringements of fundamental legal rights might nevertheless be lawful, because they usually serve important objectives such as crime prevention. For example, coercively taking and examining a person's DNA infringes one's right to respect for bodily integrity and personal data, but can be lawful when used for the detection and prevention of crime.<sup>9</sup> Therefore, although infringed, this right is not violated. Similarly, although

<sup>4</sup> Jacobs (note 1), p. 2.

<sup>5</sup> J. Timmer, L. Kool and R. van Est, 'Ethical Challenges in Emerging Applications of Persuasive Technology' in: T. MacTavish and S. Basapur (eds.) *Persuasive Technology: PERSUASIVE 2015. Lecture Notes in Computer Science*, vol. 9072 (Cham: Springer, 2015); E. Mik, 'Persuasive Technologies: From Loss of Privacy to Loss of Autonomy' in: K. Barker, K. Fairweather and R. Grantham (eds.) *Private Law in the 21st Century* (Oxford: Hart Publishing, 2017); Smids (note 1).

<sup>6</sup> A similar question could be raised by technologies that do *not aim* to change the user's mental states but unintentionally (might) do so.

<sup>7</sup> J. C. Bublitz and R. Merkel, 'Crimes against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination' (2014) 8 *Criminal Law and Philosophy* 51–77; M. Ienca and R. Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotechnology' (2017) 13 *Life Sciences, Society and Policy*; A. Lavazza, 'Freedom of Thought and Mental Integrity: The Moral Requirements for Any Neural Prosthesis' (2018) 12 *Frontiers in Neuroscience*; J. C. Bublitz, 'The Nascent Right to Psychological Integrity and Mental Self-Determination' in A. von Arnould, K. von der Decken and M. Susi (eds.) *The Cambridge Handbook of New Human Rights: Recognition, Novelty, Rhetoric* (Cambridge: Cambridge University Press, 2020), pp. 387–403; S. Michalowski, 'Critical Reflections on the Need for a Right to Mental Self-Determination', in A. von Arnould, K. von der Decken and M. Susi (eds.), *ibid.*, pp. 404–412. By contrast, little attention has been paid to the moral right to mental liberty, though one of us has alluded to the possible existence of such a right: T. Douglas, 'Criminal Rehabilitation through Medical Intervention: Moral Liability and the Right to Bodily Integrity' (2014) 18 *The Journal of Ethics* 101–122, at 120. See also T. Douglas and L. Forsberg, 'Three Rationales for a Legal Right to Mental Integrity' in: D. A. G. Van Toor et al. (eds.) *NeuroLaw: Ways Forward for Neuroscience, Justice, and Security* (New York: Palgrave Macmillan, 2021) (forthcoming).

<sup>8</sup> Note that the possibility that PTs could be used coercively is consistent with the definitional claim made previously that PTs are themselves non-coercive; one can coerce someone to submit to an intervention that is itself non-coercive. For example, one can require, on pain of punishment, a person to attend an anger management course that itself employs non-coercive means to change behaviour.

<sup>9</sup> E.g. ECtHR 15 May 2018, appl.no. 41079/16, (*Caruana v. Malta*), § 28–42; ECtHR 4 June 2013, appl.nos. 7841/08, 57900/12, (*Peruzzo and Martens v. Germany*), § 44–49.

imprisonment seriously and constantly infringes prisoners' rights to freedom of movement, detaining criminal offenders is broadly accepted to be lawful in at least some cases.<sup>10</sup> Thus, criminal justice is both the context in which non-consensual PTs are perhaps most likely to be employed and the context in which their application is most likely to be lawful. For these reasons, we take criminal justice as our focus in this chapter. In particular, we focus on the use of PTs as aids to correctional rehabilitation, which we will call 'smart correctional rehabilitation' or simply 'smart rehabilitation'. It is worth noting, however, that much of what we have to say regarding PTs in forensic contexts may carry over to other contexts as well.

Smart rehabilitation is already being discussed. For example, some hold that PTs could be useful aids to rehabilitation in the context of parole, to monitor parolees and provide them with feedback in order to stimulate them to change mental states that support criminal lifestyles.<sup>11</sup> However, one issue that has not been addressed is whether smart rehabilitation might infringe a legal right to mental liberty.

By contrast, the issue of mental liberty has been discussed in relation to other non-consensual technological interventions that aim to alter the minds of criminal offenders. This discussion is focussed primarily on neurointerventions:<sup>12</sup> that is, interventions that seek to alter a person's brain and thus mind through direct (that is, not psychologically mediated) means, and that are performed without the valid consent of the person being targeted. Examples would include pharmaceutical interventions to reduce sex drive in sex offenders, and (non-)invasive brain stimulation, which also aims to change a person's mental states, and might be used, say, to enhance empathy in psychopaths. The focus on non-consensual *neurointerventions* is understandable; these are, perhaps, the interventions that most plausibly infringe a person's right to mental liberty. However, some psychologically mediated forms of influence, such as hypnosis and indoctrination, plausibly infringe the right to mental liberty too. And it can legitimately be asked whether the less extreme forms of psychological influence exerted by the PTs being considered today might infringe the right as well.

Smart rehabilitation raises challenges that have not been adequately addressed in the literature on neurointerventions. For example, it raises in a more challenging form the question of how we are to distinguish between mental liberty-infringing forms of influence and innocuous forms of persuasion such as the giving of dispassionate arguments. Neurointerventions are often distinguished from apparently innocuous forms of influence on the basis that they act directly – without psychological mediation – on the brain.<sup>13</sup> But this strategy is not available for PTs, whose influence on the brain is indirect. Smart rehabilitation also raises the question of whether infringement of mental liberty alone is sufficient for unlawfulness, or must be

<sup>10</sup> See also Douglas (note 7).

<sup>11</sup> B. Ticknor, 'Virtual Reality and Correctional Rehabilitation: A Game Changer' (2019) *Criminal Justice and Behavior* <https://doi.org/10.1177/0093854819842588>; R. S. Gable and R. K. Gable, 'Remaking the Electronic Tracking of Offenders into a 'Persuasive Technology' (2016) 34 *Journal of Technology Human Services* 13–31. Byrne and Pattavina (note 1), p. 125; A. Pattavina, 'The Use of Electronic Monitoring as Persuasive Technology: Reconsidering the Empirical Evidence on the Effectiveness of Electronic Monitoring' (2009) 4 *Victims and Offenders* 389; S. Elison-Davies, G. Davies, J. Ward, S. Dugdale and J. Weekes, 'The Role of Technology in Offender Rehabilitation' (2018) 5 *Advancing Corrections Journal* 107–119.

<sup>12</sup> E.g. D. Birks and T. Douglas (eds.), *Treatment for Crime* (New York: Oxford University Press, 2018); Bublitz (note 7); J. C. Bublitz, 'My Mind Is Mine!? Cognitive Liberty as a Legal Concept' in: E. Hildt and A. G. Franke (eds.) *Cognitive Enhancement* (Dordrecht: Springer, 2013) 233–264.

<sup>13</sup> Bublitz and Merkel (note 7).

accompanied, as it is typically with non-consensual neurointerventions, by an infringement of *bodily* integrity.<sup>14</sup>

In this chapter, we explore whether and how legal rights do, or ought to, protect mental liberty, with particular focus on protection against smart rehabilitation. We argue that current legal protection for mental liberty is unclear, and that, in seeking to clarify this protection for the purposes of assessing smart rehabilitation, three considerations are relevant: whether an intervention engages rational processes, whether it is resistible, and whether it operates via psychological channels. However we acknowledge that, although these considerations might provide helpful guidance in the examination of smart rehabilitation in light of mental liberty, they also raise difficult questions.

The outline of this chapter is as follows. In Section 3.2 we provide a fuller definition of PTs and present some possibilities that PTs offer for the smart correctional rehabilitation of criminal offenders. Next, in Section 3.3, we briefly discuss the right to mental liberty and the extent to which this right is guaranteed by existing European human rights. In Section 3.4, we discuss three factors that should be considered in specifying human rights protection against smart rehabilitation. Subsequently, in Section 3.5 we explore whether the use of PTs in the context of smart rehabilitation would infringe an appropriately specified legal right to mental liberty. We suggest that, in this context, it might be difficult to identify compelling distinctions between novel forms of smart rehabilitation and more traditional criminal legal interventions, such as the imposition of a prison sentence or a psychological treatment program.

### 3.2 PERSUASIVE TECHNOLOGIES: SOME POSSIBILITIES FOR SMART REHABILITATION

As briefly mentioned in the introduction, PTs are standardly defined as digital technologies that are explicitly designed to change attitudes and/or behaviours. They intend to change how their users think and/or act, not by coercion, deception, extreme manipulation, or direct intervention on the brain but through other means: typically information provision, prompting, and encouragement.<sup>15</sup>

An example of an existing PT is the system of Intelligent Speed Adaption (ISA).<sup>16</sup> This is an in-vehicle system that uses information on the position of a vehicle in relation to the speed limit in force at that particular location, which enables feedback to be provided to the driver. Whereas the *advisory* ISA only warns the driver that he exceeds the speed limit, the *supportive* ISA tries to limit the speed itself (for example via pressure on the accelerator pedal when the speed limit is exceeded), though it can be overridden by the driver at any time.<sup>17</sup> These technologies are intentionally designed to change the driver's choices about exceeding speed limits, but the driver has ultimate control over whether or not she continues speeding.

The example of ISA can be classified as a self-monitoring technology, which is one of the seven types of PTs that Fogg distinguishes.<sup>18</sup> According to Fogg, self-monitoring technologies

<sup>14</sup> E. Shaw, 'Against the Mandatory Use of Neurointerventions in Criminal Sentencing' in: D. Birks and T. Douglas (eds.) *Treatment for Crime* (New York: Oxford University Press, 2018), p. 334 argues that non-consensual neurointerventions are especially objectionable because they infringe *both* bodily and mental integrity.

<sup>15</sup> Fogg (note 1), pp. 15, 17, 32; W. IJsselstijn et al., 'Persuasive Technology for Human Well-Being: Setting the Scene' in: W. IJsselstijn, Y. de Kort, C. Midden, B. Eggen and E. van den Hoven (eds.) *Persuasive Technology* (Dordrecht: Springer, 2006), p. 1; Smids (note 1), p. 72.

<sup>16</sup> J. Smids, 'The Moral Case for Intelligent Speed Adaption' (2018) 35 *Journal of Applied Philosophy* 205–221

<sup>17</sup> *Ibid.*, p. 209.

<sup>18</sup> Fogg (note 1), p. 32. Other types are reduction, tunnelling, tailoring, suggestion, surveillance and conditioning technology.

allow ‘people to monitor themselves to modify their attitudes or behaviours to achieve a predetermined goal or outcome.’<sup>19</sup> Ideally, they work in real time, giving users ongoing data about their location, their progress on a task or about their physical or mental state, such as a (wearable) heart rate monitor that provides the user with feedback.<sup>20</sup> Another example of such a self-monitoring PT is the HAPIfork, a smart fork that monitors and tracks the user’s eating habits and aims to persuade users to eat more slowly.<sup>21</sup>

Whereas self-monitoring PTs enable individuals to learn about themselves, surveillance PTs enable an observer to learn about others. Fogg defines this latter type of PT as ‘any computing technology that allows one party to monitor the behaviour of another to modify behaviour [of the latter] in a specific way.’<sup>22</sup> The basic idea is that if you know that you are being observed, you behave differently. As an example of a persuasive surveillance technology, Fogg describes the surveillance system Hygiene Guard, which monitors whether employees wash their hands after visiting the rest room.

Even though PTs are by definition not themselves coercive, individuals may be coerced into using them. One area in which this is particularly likely to occur is in the monitoring of convicted offenders. For example, some advocate the integration of PTs into traditional monitoring tools to supervise probationers and parolees. While traditional tools such as GPS primarily focus on surveillance and control, PTs could add direct information and communication in relation to surveillance, with the aim of modifying mental states that support criminal lifestyles.<sup>23</sup> An example of a persuasive surveillance technology tool already used in criminal justice is SCRAM Continuous Alcohol Monitoring.<sup>24</sup> An ankle bracelet measures alcohol consumption through the user’s perspiration every thirty minutes.<sup>25</sup> The SCRAM CAM is connected to a mobile app (SCRAMnet Mobile), allowing officers to access real time data concerning a person’s alcohol consumption on probation or parole. We can easily imagine that criminal justice authorities might coerce an offender into wearing such an ankle bracelet, though the impact of the feedback from the bracelet on the individual’s alcohol consumption could remain non-coercive.

An example of persuasive *self*-monitoring that can potentially be used for the purpose of smart rehabilitation is biofeedback via wearable devices. A wearable device that measures biological features in real-time – such as physiological signals related to stress, engagement, and excitement<sup>26</sup> – communicates these features back to the wearer, for example, via a smartphone or smartwatch app.<sup>27</sup> Research has shown that wearables measuring cardiovascular and electrodermal parameters and motion were helpful in predicting aggression in youth with autism.<sup>28</sup>

<sup>19</sup> Ibid., p. 44. Of note, in the context of smart rehabilitation, the outcome could be predetermined by a third party, e.g., a judge.

<sup>20</sup> Ibid., pp. 44–45 and pp. 47–48.

<sup>21</sup> Jacobs (note 1), p. 2.

<sup>22</sup> Fogg (note 1), p. 46.

<sup>23</sup> Byrne and Pattavina (note 1), p. 125; Pattavina (note 11), p. 389; Gable and Gable (note 11).

<sup>24</sup> *State v. Lemler*, 2009 SD 86, 774 N.W.2d 272.

<sup>25</sup> [www.scramsystems.com/products/scram-continuous-alcohol-monitoring/](http://www.scramsystems.com/products/scram-continuous-alcohol-monitoring/)

<sup>26</sup> D. Kinnamon et al., ‘Portable Biosensor for Monitoring Cortisol in Low-Volume Perspired Human Sweat’ (2017) 7 *Scientific Reports*; A. Kaushik et al., ‘Recent Advances in Cortisol Sensing Technologies for Point-of-Care Application’ (2014) 53 *Biosensors and Bioelectronics* 499–512; M. Garbanino et al. ‘Empatica-E3 A Wearable Wireless Multi-sensor Device for Real-Time Computerized Biofeedback and Data Acquisition’ (2014) *International Conference on Wireless Mobile Communication and Healthcare* 39–42.

<sup>27</sup> <https://e4.empatica.com/e4-wristband>.

<sup>28</sup> M. S. Goodwin et al., ‘Predicting Aggression to Others in Youth with Autism Using a Wearable Biosensor’ (2019) 12 *Autism Res* 1286–1296.

Devices that measure such features and provide immediate feedback could potentially contribute to a convicted offender's recognition of increased risk, motivating him to change his mental state and behaviour in such situations, and ultimately prevent recidivism.<sup>29</sup>

There are also PTs that combine self-monitoring with surveillance, such as Phone-Coach Monitoring. In this context, probationers and parolees receive automated phone calls every day, asking questions tailored to their personal situations. Their answers are recorded and the phone-coach system tracks progress and provides immediate feedback to the participant and, if authorized, to the parole or probation officer.<sup>30</sup>

Besides persuasive surveillance and self-monitoring, it is possible to imagine other kinds of PTs that might be deployed in smart rehabilitation. Examples include virtual reality (VR) and augmented reality (AR).<sup>31</sup> According to Chittaro and Zangrondo, VR naturally supports the delivery of various forms of negative and positive feedback to the user,<sup>32</sup> based on which actions he chooses, and on how those actions have been categorized as right or wrong by the designers of the virtual environment.<sup>33</sup> According to Ticknor, VR has been useful in improving the effectiveness of traditional cognitive behavioural therapy, which is often used as treatment for offenders, by offering a safe, controlled and realistic environment where the individual can learn and practice new skills. Although research is still very limited, based on the ways in which VR has been used for other types of real-world problems, Ticknor argues that it can improve traditional cognitive behavioural therapy for criminal rehabilitation.<sup>34</sup>

If VR-based interventions will be used in criminal justice, it is likely that this will occur first in a forensic psychiatric or psychological setting, where such interventions might be used for treatment purposes, monitoring of progress, or risk assessment.<sup>35</sup> For example, VR allows safe monitoring of sexual offenders in a virtual supermarket where they encounter a child.<sup>36</sup> Suppose that an offender's risk, using traditional risk assessment tools, is considered too high for release but that a new, VR-PT application is available for risk assessment and, if desirable, for risk reduction. It will offer a personalised VR environment, closely monitoring and influencing the offender's physical and behavioural responses to the virtual stimuli. Such a technique may be non-coercive in the sense that the offender can always resist its effects, but it may still be imposed coercively; for example, the offender may be required to use the technology before release from a secure facility will be considered.<sup>37</sup>

<sup>29</sup> See L. Comet et al., *Neuroscientific Applications in the Juvenile Criminal Justice System: An Inventory of Measurement Instruments, Prevention and Intervention Methods* (The Hague: WODC, 2016), pp. 201–203.

<sup>30</sup> B. O. Burraston, S. J. Bahr and D. J. Cherrington, 'Reducing Juvenile Delinquency with Automated Cell Phone Calls' (2014) 58 *International Journal of Offender Therapy and Comparative Criminology* 525; N. R. VanDeMark et al., 'An Exploratory Study of Engagement in a Technology-Supported Substance Abuse Intervention' (2010) 5 *Substance Abuse Treatment, Prevention and Policy*; Byrne and Pattavina (note 1), pp. 126–127.

<sup>31</sup> Augmented reality means that only certain parts of the outside world are modified (virtual). For instance, a virtual spider can be seen on a real desk.

<sup>32</sup> For example, training impulse control for aggression regulation: S. Klein Tuente et al, 'Effect of Virtual Reality Aggression Prevention Training for Forensic Psychiatric Patients (VRAPT): Study Protocol of a Multi-Center RCT' (2018) 18 *BMC Psychiatry*.

<sup>33</sup> Chittaro and Zangrondo (note 1), p. 58.

<sup>34</sup> Ticknor (note 1); B. Ticknor, *Virtual Reality and the Criminal Justice System. Exploring the Possibilities for Correctional Rehabilitation* (Lanham, MD: Lexington Books, 2018).

<sup>35</sup> L. Comet and J. L. Van Gelder, 'Virtual Reality: A Use Case for Criminal Justice Practice' (2020) *Psychol Crime Law*, doi:10.1080/1068316X.2019.1708357.

<sup>36</sup> P. Fromberger, S. Meyer, K. Jordan et al., 'Virtual Reality Applications for Diagnosis, Risk Assessment and Therapy of Child Abusers' (2018) 36 *Behavioral Sciences and the Law* 235–244.

<sup>37</sup> Because of the setting, the person may be considered *vulnerable*, see P. Kellmeyer, N. Biller-Andorno and G. Meynen, 'Ethical Tensions of Virtual Reality Treatment in Vulnerable Patients' (2019) 25 *Nature Medicine* 1185–1188.

Altogether, a variety of PTs could potentially contribute to the smart rehabilitation of criminal offenders. However, as was briefly mentioned in the introduction, intentionally changing mental states of convicted offenders through smart rehabilitation also raises questions regarding the notion of mental liberty. In the next section, we discuss the notion of mental liberty and the extent to which it is protected by existing European human rights.

### 3.3 THE RIGHT TO MENTAL LIBERTY: A EUROPEAN HUMAN RIGHTS PERSPECTIVE

#### 3.3.1 Introduction

As yet, no European treaty or charter explicitly guarantees a human right to mental liberty. However, commentators are currently debating the case for introducing such a human right.<sup>38</sup> Within this discussion, different rights are proposed, such as rights to mental integrity, psychological continuity, cognitive liberty, and mental self-determination. Though there are significant differences around the edges, on many formulations these putative rights share the same core: they protect against certain kinds of mental interferences, where a mental interference consists in the non-consensual alteration of one person's mental states by the intentional action of another. We refer to a right against (certain kinds of) mental interference as the right to 'mental liberty'.

Existing discussion of mental liberty has mainly been driven by advancements in neurotechnology that enable – or may soon enable – us to intervene in a person's brain and thereby one's mind. The most widely discussed examples include transcranial direct current stimulation (tDCS) and transcranial magnetic stimulation (TMS) – both of which allow the stimulation of particular brain regions via non-invasive means – as well as deep brain stimulation, in which brain regions are stimulated via implanted electrodes.

To some extent, a right to mental liberty is already guaranteed by existing European human rights law. The European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), as well as the Charter of Fundamental Rights of the European Union (ECFR), guarantee the right to freedom of thought. In addition, both the ECHR and the ECFR recognise a right to mental, psychological and moral integrity. The meaning and scope of these rights are briefly discussed next.

#### 3.3.2 *The Right to Freedom of Thought*

Article 9(1) ECHR prescribes that 'Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief and freedom, either alone or in community with others and in public or private, to manifest his religion or belief, in worship, teaching, practice and observance.' The same right is guaranteed by Article 10(1) ECFR.<sup>39</sup>

<sup>38</sup> Ienca and Andorno (note 7); Bublitz (note 7); Bublitz 2013 (note 12); Lavazza (note 7); L. Kirchmair 'Objections to Coercive Neurocorrectives for Criminal Offenders: Why Offenders' Human Rights Should Fundamentally Come First' (2019) 38 *Criminal Justice Ethics* 19–40; see J. M. Muñoz, 'Chile: Right to Free Will Needs Definition' (Correspondence) (2019) *Nature* 634.

<sup>39</sup> According to Article 52(3) ECFR, the meaning and scope of Articles 10 ECFR and 9 ECHR are the same. Therefore, we mainly focus on (the more elaborated) literature and case-law on Article 9 ECHR. See also Article 18 International Covenant on Civil and Political Rights and Article 18 Universal Declaration of Human Rights.

This right encompasses an internal and external dimension (the *forum internum* and *forum externum*). The right to freedom of thought, conscience and religion, which is exercised inside the individual's mind,<sup>40</sup> comprises the internal dimension. The right to manifest one's religion and belief comprises the external dimension.<sup>41</sup> The *forum internum* is absolute. This part of the right may not be infringed: any infringement will constitute a violation of the right. Infringement of the *forum externum* can, however, be justified under certain circumstances (Article 9(2) ECHR).

According to the *Human Rights Handbooks* of the Council of Europe, the internal dimension of Article 9 ECHR seeks at its most basic level to prevent state indoctrination of individuals: every person is free to have, develop, refine and change one's thought, conscience and religion.<sup>42</sup> Other handbooks take a similar approach, under which the *forum internum* aims to enable individuals to continue their belief, and guarantee that states may never interfere with the most inner sphere of its citizens, for instance by coercing a person to change his beliefs, dictating what one has to believe (for example, through brainwashing), or by using inquisitorial methods to discover a person's thoughts and convictions.<sup>43</sup>

Much case-law and scholarship exist on the right to freedom of conscience and religion.<sup>44</sup> However, how far the right extends to other aspects of our mental lives via the notion of 'thoughts' remains unclear.<sup>45</sup> De Jong distinguishes three general interpretations of *thought*, according to which the *forum internum* covers

- (1) religion: i.e., all and only thoughts and convictions that are related to religion,
- (2) conscience: i.e., all and only thoughts and convictions (including those of a religious nature) that have a major impact on one's way of living,
- (3) all thoughts.<sup>46</sup>

Owing to a lack of case-law, however, it remains unclear which approach would apply in the European context. Some argue for a broad interpretation on which the right protects even trivial thoughts such as decisions about which socks to wear.<sup>47</sup> Others argue that the legal history and case-law of the European Court on Human Rights (ECtHR) rather indicate a narrow approach

<sup>40</sup> D. Gomien, *Short Guide to the European Convention on Human Rights* (The Hague: Council of Europe Publishing, 2005), p. 95.

<sup>41</sup> D. J. Harris et al., *Harris, O'Boyle, and Warbrick: Law of the European Convention on Human Rights* (New York: Oxford University Press, 2018), pp. 573–574.

<sup>42</sup> J. Murdoch, *Protecting the Right to Freedom of Thought, Conscience and Religion under the European Convention on Human Rights* (Strasbourg: Council of Europe, 2012), p. 18.

<sup>43</sup> Harris et al. (note 41), p. 573; B. Vermeulen and M. Roosmalen, 'Freedom of Thought, Conscience and Religion' in: P. Van Dijk et al. (eds.), *Theory and Practice of the European Convention on Human Rights* (Cambridge: Intersentia, 2018), pp. 738–739.

<sup>44</sup> Council of Europe, *Guide on Article 9 of the European Convention on Human Rights* (Strasbourg: Council of Europe/European Court of Human Rights, 2019); Murdoch (note 42), pp. 16–21; P. M. Taylor, *Freedom of Religion: UN and European Human Rights Law and Practice* (Cambridge Core, 2005); C. Evans, *Freedom of Religion Under the European Convention on Human Rights* (New York: Oxford University Press, 2001); M. D. Evans, *Religious Liberty and International Law in Europe* (Cambridge: Cambridge University Press, 1997).

<sup>45</sup> S. Lighthart et al., 'Forensic Brain-Reading and Mental Privacy in European Human Rights Law: Foundations and Challenges' (2021) *Neuroethics* 14 191–203; J. C. Bublitz, 'The Soul is in the Prison of the Body: Mandatory Moral Enhancement, Punishment & Rights against Neuro-rehabilitation' in: D. Birks and T. Douglas (eds.) *Treatment for Crime* (New York: Oxford University Press, 2018), pp. 300–301; Evans (note 44), p. 52; Taylor (note 44), p. 119.

<sup>46</sup> D. de Jong, *The Freedom of Thought, Conscience and Religion or belief in the United Nations (1946–1992)*, (Antwerp/Groningen/Oxford: Intersentia/Hart, 2000), p. 23.

<sup>47</sup> S. Alegre, 'Rethinking Freedom of Thought for the 21st Century' (2001) *European Human Rights Law Review* 3 221–233. See J. C. Bublitz 'Freedom of Thought in the Age of Neuroscience' (2014) 100 *Archiv Für Rechts- Und*



to the *forum internum*.<sup>48</sup> For example, according to the Grand Chamber of the ECtHR, not all thoughts, convictions and religions are protected by Article 9 ECHR. Instead, ‘the right to freedom of thought, conscience and religion denotes only those views that attain a certain level of cogency, seriousness, cohesion and importance’.<sup>49</sup> This seems to indicate some sympathy for the second (‘conscience’) approach described by De Jong. For instance, to be protected under Article 9 ECHR, convictions should amount to more than mere opinions or ideas, and must concern a ‘weighty and substantial aspect of human life and behaviour’.<sup>50</sup> Applying this reasoning to ‘thoughts’ more generally, one could argue that these need to involve more than mere knowledge, opinions or ideas but should concern a weighty and substantial aspect of the individual’s life.<sup>51</sup> An intention to vote for a specific political party, for example, may be a thought that is sufficiently weighty and substantial to be protected by the *forum internum*,<sup>52</sup> as may some scientific, philosophical and political beliefs.<sup>53</sup>

Yet, one might wonder why the ECtHR would follow such a narrow approach, instead of guaranteeing freedom of all thoughts, regardless of their (relative) importance. One argument for a narrow approach, mentioned by Bublitz,<sup>54</sup> relates to the absolute nature of the *forum internum*: because any infringement is prohibited, applying a broad right to freedom of thought could have extreme and far-reaching implications, for instance in the context of nudging or, as Evans argues, advertising, which both aim to change what people think.<sup>55</sup> Since it may be impractical or undesirable to absolutely prohibit a very broad range of mental influences, freedom of thought must be construed narrowly, such that most influences are not conceived as infringements.<sup>56</sup> However, as yet case-law has not provided a verdict on this argument. All we can say with confidence is that the scope of the right to freedom of thought remains to be elaborated.

Since the precise scope of the right to freedom of thought is still unclear, it is hard to predict whether the ECtHR would consider smart rehabilitation to interfere with ‘thoughts’ in the meaning of Article 9 ECHR. But even if we assume that the scope of this right is broad,<sup>57</sup> and that (at least some) types of smart rehabilitation can therefore contribute to changing thoughts protected under Article 9 ECHR, it is not clear that the right to freedom of thought would prohibit the use of such smart rehabilitation tools in a criminal justice context, since PTs do not completely extinguish the user’s control over what he thinks, and thus do not clearly diminish freedom of thought.<sup>58</sup> This aspect of PTs – leaving (some level of) ultimate control to the user – would probably be particularly relevant in the determination of the legal implications of smart

*Sozialphilosophie* 1–25; S. McCarthy-Jones, ‘The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century’ (2019) 2 *Frontiers in Artificial Intelligence* 1–17.

<sup>48</sup> Ligthart et al. (note 45); Harris et al. (note 41), p. 573; Murdoch (note 42), p. 16.

<sup>49</sup> ECtHR (GC) 1 July 2014, appl.no. 43835/11 (*S.A.S. v. France*), § 55; ECtHR (GC) 26 April 2016, appl.no. 62649/10 (*İzzettin Doğan and others v. Turkey*), § 68.

<sup>50</sup> ECtHR *Campbell and Cosans v. UK*, appl.nos. 7511/76, 7743/76, § 36 (22 February 1982).

<sup>51</sup> Ligthart et al. (note 45).

<sup>52</sup> ECtHR 8 July 2008, appl.no. 9103/04 (*Georgian Labour Party v. Georgia*), § 120; ECtHR 11 January 2007, appl. nos. 55066/00, 55638/00 (*Russian Conservative Party of Entrepreneurs and Others v. Russia*), § 76. However, probably less ‘weighty’ but also protected under the freedom of thought, is a person’s wish to give one’s child a particular name: ECHR 2 July 1997, appl.no. 27868/95 (*Salonen v. Finland*).

<sup>53</sup> W.A. Schabas, *The Universal Declaration of Human Rights: The Travaux Préparatoires* (New York: Cambridge University Press, 2013), pp. 1766, 2489, 2500 (Soviet Union), 1766–1767 (France) and 2499 (Uruguay).

<sup>54</sup> Bublitz (note 7), p. 1317. See also Taylor (note 44), p. 118.

<sup>55</sup> Evans (note 44), p. 295.

<sup>56</sup> Bublitz (note 7), p. 1317; Evans (note 44), p. 295.

<sup>57</sup> As is argued, for example, by Alegre (note 47).

<sup>58</sup> Smids (note 1), pp. 3, 79–85.

rehabilitation in light of the freedom of thought pursuant to Article 9 ECHR. For example, as Harris et al. argue, ‘as long as individuals are able to continue in their beliefs, Article 9(1) will not be violated’.<sup>59</sup> In Section 3.4, we explore some further considerations that might be important for specifying the scope of human rights protections against persuasive technologies, both from a legal and ethical point of view.

### 3.3.3 *The Right to Respect for Mental, Psychological and Moral Integrity*

According to Article 3 of the EU Charter of Fundamental Rights, ‘Everyone has the right to respect for his or her physical and mental integrity.’ Although not explicitly mentioned by the European Convention on Human Rights, a similar right – to psychological and moral integrity – is recognised by the Grand Chamber of the ECtHR, covered by the right to respect for private life as set out in Article 8 ECHR,

The concept of “private life” is a broad term which is not susceptible to exhaustive definition. It covers the physical and psychological integrity of a person, and can therefore embrace multiple aspects of the person’s identity such as, for example, gender identification, sexual orientation name and elements relating to a person’s right to his or her image.<sup>60</sup>

The Court has found that health, together with physical and moral integrity, falls within the realm of private life.<sup>61</sup>

Similar to the right to freedom of thought, the precise meaning and scope of the European rights to mental, moral and psychological integrity are subject to disagreement.<sup>62</sup> Whereas some authors suggest that they could imply a right to mental self-determination – that is, to control over the content of one’s mental life – others presume that they are meant as a right to mental health.<sup>63</sup> While case-law of the ECtHR does not clearly indicate that the Court recognises a right to mental self-determination, interpreting the right to mental, moral and psychological integrity as nothing more than a right to mental health, may be in one way too restrictive. The Court associates moral and mental integrity with issues of mental health;<sup>64</sup> however, it also considers one’s reputation and honour to be part of the psychological integrity of a person, though these do not bear any close connection to mental health.<sup>65</sup> Moreover, the Court has acknowledged that one’s psychological integrity embraces various other aspects of a person’s (social) identity, such as one’s name, gender identification and sexual interest.<sup>66</sup> It seems

<sup>59</sup> Harris et al. (note 41), p. 573. See also Evans (note 44), p. 295.

<sup>60</sup> ECtHR (GC) 29 March 2016, appl.no. 56925/08 (*Bédát v. Switzerland*), § 72. Also: ECtHR (GC) 25 September 2018, appl.no. 76639/11 (*Denisov v. Ukraine*), § 95.

<sup>61</sup> ECHR (GC), 12 September 2012, 10593/08 (*Nada v. Switzerland*), § 151. Cf. EHRM 24 July 2014, appl. no. 7446/12 (*Remetin v. Croatia* (nr. 2)), § 70: ‘Under Article 8 States have a duty to protect the physical and moral integrity of an individual from other persons.’

<sup>62</sup> Bublitz (note 45), p. 303.

<sup>63</sup> J. C. Bublitz, ‘My Mind Is Mine!? Cognitive Liberty as a Legal Concept’ in: E. Hildt and A. G. Franke (eds.) *Cognitive Enhancement: An Interdisciplinary Perspective. Trends in Augmentation of Human Performance 1* (Dordrecht: Springer, 2013), p. 248; Ienca and Andorno (note 7), p. 18; Bublitz (note 7); Michalowski (note 7).

<sup>64</sup> ECtHR 6 February 2001, appl.no. 44599/98 (*Bensaid v. UK*), § 47; ECtHR 15 November 2016, appl.nos. 28859/11 and 28473/12 (*Dubská and Krejzová v. Czech Republic*), § 163; ECtHR (GC) 11 July 2006, appl.no. 54810/00 (*Jalloh v. Germany*), § 79, 82.

<sup>65</sup> ECtHR 28 October 2014, appl.no. 20531/06 (*Ion Carstea v. Romania*) § 38; ECtHR 21 November 2013, appl. no. 16882/03 (*Putistin v. Ukraine*), § 32.

<sup>66</sup> ECtHR (GC) 25 September 2018, 76639/11 (*Denisov v. Ukraine*), § 95; ECtHR (GC) 29 March 2016, appl.no. 56925/08 (*Bédát v. Switzerland*), § 72.

plausible, then, that the right could protect against some mental interferences<sup>67</sup> that do not threaten mental health. However, the precise meaning and scope of the rights to mental, moral and psychological integrity remain unclear. It is thus unclear what protection they provide against smart rehabilitation, especially where they do not threaten mental health. Of note, unlike infringements of the right to freedom of thought, an infringement of the right to moral and psychological integrity as covered by Article 8(1) ECHR, can, according to Article 8(2) ECHR, be justified if it (1) is based on a foreseeable and accessible legal ground, (2) serves a legitimate interest and (3) is necessary for the realisation of the legitimate interest pursued.

### 3.3.4 *Is a New Right Needed?*

The basic idea of a right to mental liberty – i.e., protecting against non-consensual alteration of one person’s mental states by the intentional action of another – is, to some extent, captured by existing European human rights: the right to freedom of thought and the right to respect for mental, psychological and moral integrity will protect against some interferences with the mind: for example, those that seriously threaten mental health or modulate religious beliefs. Yet, the precise meaning and scope of these rights remain unclear. As a consequence, these rights provide little guidance on which interventions on a person’s mental states are permitted or prohibited. Clearer legal regulation could be developed by (1) introducing a novel human right to mental liberty,<sup>68</sup> or (2) further elaborating existing human rights.<sup>69</sup> Either way, it might make sense to speak of a right to mental liberty – a right against (certain kinds of) mental interference – though on the former approach this would be a stand-alone human right, whereas on the latter, it would derive from others.

In our view, the latter approach would be sufficient. Current human rights could, in principle, protect all conceivable (mental) interests that are worth protecting,<sup>70</sup> such as particular thoughts and mental states. However, in what follows we will remain neutral between options (1) and (2). We will continue to speak of a ‘right to mental liberty’ but without thereby implying that the right is a separate and independent fundamental right rather than an implication of existing rights.

In the next section, we will suggest that attempts to precisely specify the scope of the right to mental liberty in the domain of smart rehabilitation face serious challenges.

## 3.4 CHALLENGES IN SPECIFYING THE SCOPE OF THE RIGHT TO MENTAL LIBERTY IN THE DOMAIN OF SMART REHABILITATION

Regardless of whether a novel human right to mental liberty is developed, or the approach is instead to specify the implications of existing rights, we can ask whether and, if so, when smart rehabilitation will infringe an appropriately specified human right to mental liberty. In this section, we raise some issues that will need to be addressed in specifying the scope of the right in this domain. We believe that three considerations should be relevant here: (1) the *significance* of the mental alterations caused by the smart rehabilitation, (2) the *intentions* of the intervener

<sup>67</sup> That is, the non-consensual alteration of one person’s mental states by the intentional action of another.

<sup>68</sup> Ienca and Andorno (note 7); Bublitz (note 7).

<sup>69</sup> Michalowski (note 7).

<sup>70</sup> See *ibid.*; Ligthart et al. (note 45); J.C. Bublitz ‘Novel Neurorights: From Nonsense to Substance’ *Neuroethics* 2022, 5, Article 7.

and (3) the *means* via which those alterations are produced. Each consideration raises difficult questions.

On *significance*: In law, the principle of proportionality plays an important role in justifying infringements of different human rights.<sup>71</sup> For example, more severe interferences with someone's *bodily* integrity (protected by Article 8 ECHR) require stronger reasons to be justified.<sup>72</sup> By analogy, the more severe the interference with a person's mental liberty would be, the stronger reasons we would need to justify it. In the case of bodily integrity, the severity of an interference might be determined by the physical invasiveness of an intervention. In the case of mental liberty, we suggest that it would be determined by the *significance* of the alteration to a person's mental states. This is consistent with the view, discussed earlier in the chapter, that highly significant mental states, such as conscience and religious beliefs, will be covered by Article 9 ECHR, which prohibits any infringement.

However, the question arises of how the significance of an alteration should be determined: by the pervasiveness of the alteration (roughly, what proportion of a person's mental states is affected), the magnitude of the alterations (how much the mental states are changed), the importance of those mental states (for example, whether they are central to a person's self-conception or personality) or some combination of these? If the latter, how do these factors combine to determine significance?

On the *intentions* of the intervener: It is plausible to think that, to qualify as an instance of bodily interference, the alteration of a person's bodily states must be the result of an intentional action on the part of another. If a gust of wind blows A into B, we would not think that A has physically interfered with B: at least, not in a way that might infringe B's rights. Similarly, we suggest that mental interferences must result from intentional actions, or at least, that they must do so if they are to infringe a right to mental liberty.

However, what is not clear is whether the intentions of the intervener play any role beyond this. Should the particular aim of the intentional PT matter? For example, suppose one correctional PT aims to make society safer (see next section) and the person healthier, while another similar PT only serves the financial self-interests of the intervener. Should human rights protections cover the latter but not the former? Put differently: should the precise *objective* of changing a mental state be essential in determining whether the right to mental liberty has been *infringed*? And how should we think about cases where an intervener intends to produce some mental alterations in another but her action also produces other, unintended, mental effects? In such a case, should unintended though foreseeable side-effects of an intervention contribute to determining whether the intervention infringes a right to mental liberty?

We believe, however, that the most important questions in relation to smart rehabilitation will concern the *means* via which an alteration is produced. Even when the intended effects of smart rehabilitation are highly significant, it is not clear that it should qualify as a human rights infringement. Imagine a PT that operates solely by providing information that the individual then considers in her reflective deliberation, such as the Sobriety Counter app mentioned in

<sup>71</sup> L. Lavrysen 'System of Restrictions' in: P. Vermeulen et al. (eds.) *Theory and Practice of the European Convention on Human Rights* (Cambridge: Intersensia, 2018); B. Rainey, E. Wicks and C. Ovey, *The European Convention on Human Rights* (New York: Oxford University Press, 2017), pp. 340–368.

<sup>72</sup> For example, whereas the ECtHR considers compulsory taking of blood and saliva to involve a minor, 'relatively slight' infringement of a person's right to bodily integrity, it qualified non-consensual genealogical examination of detainees as an 'intrusive and serious interference', and therefore harder to justify. ECtHR 7 December 2006, appl.no. 29514/05, (*Van der Velden v. The Netherlands*), § 9; ECtHR 15 May 2018, appl.no. 41079/16, (*Caruana v. Malta*), § 28–42; ECtHR 13 May 2008, appl.no. 52515/09 (*Juhnke v. Turkey*).

Section 3.1, persuading addicts to stop drinking, by merely providing information about general alcohol-related health issues and the money one saves by not drinking. Using such a PT might eventually have highly significant mental effects, regardless of how significance is determined, and might be intended to have highly significant effects. Nevertheless, it seems doubtful that this PT infringes human rights. On the other hand, there are certain types of (non-consensual) mental interferences that plausibly should be regarded as infringing (though not necessarily violating) human rights, even when their effects (both intended and unintended) are relatively minor. This point is best illustrated by moving beyond the realm of PTs. Though not a PT, the compulsory administration of aggression-lowering drugs to prisoners might be an example of an intervention that should be taken to infringe a right to mental liberty even if its effects are mild (and regardless of the fact that the infringement might be justified). Similar thoughts apply to the use of extreme forms of psychological manipulation, such as indoctrination and hypnosis.

An important issue, therefore, concerns the means via which the mental alteration is produced. Some means of producing an alteration seem to warrant stronger human rights protection than others. Think, for example, of direct neurointerventions such as deep brain stimulation and chemical castration, compared with indirect interventions such as cognitive behavioural therapy.<sup>73</sup> But how exactly are we to characterise the relevant differences in means? Three possibilities suggest themselves, and they have rather different implications for smart rehabilitation.

First, perhaps what should matter is the extent to which a technology exerts its influence via *rational* means: means that engage a person's rational capacities, such as their capacity for conscious deliberation. Typically, at least, the provision of information does indeed engage such capacities. On the other hand, the administration of a drug does not: it operates via purely chemical means.<sup>74</sup> Between these extremes, however, there is a lot of 'grey area'. Consider, for example, PTs that do not just provide information but provide it in a way designed to maximise its effect on a particular user, for example, by providing at the time of day at which the user is known to be most susceptible to persuasion.<sup>75</sup> Or consider a virtual environment that tweaks motivations of users in various ways, including by confronting them with fear-inducing images of prison cells, inmates, courtrooms or police agents. It is, we think, unclear whether (and to what degree) these PTs engage rationality or bypass it. Part of the unclarity here derives from uncertainty about how we should understand rationality and which processes exactly qualify as rational; for example, do the 'quick and dirty' heuristics often exploited by advertisers qualify as rational processes? Part of it derives from uncertainty regarding what it is, exactly, to *engage* rationality. And part of it derives from uncertainty about the precise mechanisms via which these technologies work; which psychological processes do they influence and how? There would thus be challenges in operationalising this distinction. Still, it might seem that the distinction is of moral importance. In general, it might be argued that rationality-engaging technologies are more respectful of the person than rationality-bypassing interventions, since they engage with

<sup>73</sup> See J. C. Bublitz 'Why Means Matter: Legally Relevant Differences Between Direct and Indirect Interventions into Other Minds' in: N. A. Vincent, T. Nadelhoffer and A. McCay (eds.) *Neurointerventions and the Law: Regulating Human Mental Capacity* (New York: Oxford University Press, 2020); F. Focquaert and M. Schermer 'Moral Enhancement: Do Means Matter Morally?' (2015) 8 *Neuroethics* 139–151.

<sup>74</sup> Bublitz (note 73), pp. 58–60.

<sup>75</sup> For discussion of similar cases, see N. Levy, 'Nudges in a Post-Truth World' (2017) 43 *Journal of Medical Ethics* 495–500; G. Keeling, 'Autonomy, Nudging and Post-Truth Politics' (2018) 44 *Journal of Medical Ethics* 721–722; N. Levy, 'Nudges to Reason: Not Guilty' (2018) 44 *Journal of Medical Ethics* 723–723; N. Levy, 'Nudge, Nudge, Wink, Wink: Nudging Is Giving Reasons' (2019) *Ergo, an Open Access Journal of Philosophy*.

precisely those capacities that are distinctive of persons.<sup>76</sup> Similar worries may extend to at least some PTs.

Second, it may be that what should matter is whether a technology exerts its influence via *resistible* means, since, other things being equal, irresistible influences are more threatening to autonomy than resistible ones.<sup>77</sup> The administration of brain-active drugs will typically have at least some irresistible mental effects: those effects that are produced entirely through neurochemical means, without any psychological processing. But the resistibility of most PTs is also open to question. For example, even the mere provision of information might have some irresistible effects, since it is disputed whether we have control over which beliefs we form when we are presented with evidence.<sup>78</sup> In addition, if PTs employ several strategies – such as prompting, information provision, and encouragement – simultaneously and during longer periods of time, it may become increasingly difficult to resist them.

Third, it may be that what should matter is whether an intervention *engages some mental process*, regardless of whether that process is a rational or resistible one, rather than inducing mental changes through a purely physico-chemical mechanism (such as through pharmaceuticals or brain stimulation). On this view, it would be enough for a technology to fall outside of the scope of the right to mental liberty that the target has to see, hear, smell, touch or taste something for the influence to occur.<sup>79</sup> On this view, while the non-consensual administration of mind-altering drugs infringes the right to mental liberty, PTs do not, since they always operate via perceptual means. This view has an advantage in operationalisability over the two views considered earlier in this section, since it is generally relatively clear whether an influence operates via some sort of mental process or whether it causes a change of mental states through physico-chemical means. However, this view arguably lacks any theoretical basis; it is not clear why the mere fact that an influence operates via some sort of mental process should be relevant to autonomy, respect or any other moral value, so it is not clear why human rights protections should be sensitive to this distinction. It also implausibly suggests that extreme forms of psychological manipulation, such as hypnosis and indoctrination, do not threaten mental liberty since they operate via perceptual and other mental processes.

Each of the three factors that we have discussed – significance, intention, and means – raises difficult questions. But it should be noted that the situation is further complicated by possible interactions between them. Though we have discussed these three considerations separately, they will ultimately need to be taken into account together. For instance, a very limited circumvention of a person's rationality that alters a person's core beliefs may be more problematic than a less rationality-engaging PT that affects a less significant mental state.

<sup>76</sup> See Bublitz and Merkel (note 7), pp. 70–73; Bublitz (note 7), pp. 1322–1325.

<sup>77</sup> For a defence of the moral relevance of resistibility in relation to 'nudges' see Y. Saghai, 'Salvaging the Concept of Nudge' (2013) 39 *Journal of Medical Ethics* 487–493.

<sup>78</sup> B. Williams, 'Deciding to Believe' in: H. E. Kiefer and M. K. Munitz (eds.) *Language, Belief, and Metaphysics* (Albany: SUNY Press, 1970); E. M. Curley, 'Descartes, Spinoza, and the Ethics of Belief' in: M. Mandelbaum and E. Freeman (eds.) *Spinoza: Essays in Interpretation* (LaSalle, IL: Open Court Publishing, 1975); W. Alston 'The Deontological Conception of Epistemic Justification' in: *Epistemic Justification: Essays in the Theory of Knowledge* (Ithaca, NY: Cornell University Press, 1989); C. Ginet, 'Deciding to Believe' in: M. Steup (ed.) *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue* (New York: Oxford University Press, 2001); N. Shah, 'Clearing Space for Doxastic Voluntarism' (2002) 805 *The Monist* 436–445; S. Ryan, 'Doxastic Compatibilism and the Ethics of Belief' (2003) 114 *Philosophical Studies* 47–79.

<sup>79</sup> See Bublitz (note 73).

### 3.5 SMART REHABILITATION AND MENTAL LIBERTY: CONFRONTING THE CHALLENGES

Although an exhaustive examination of the relation between smart rehabilitation and the right to mental liberty exceeds the scope of the present chapter, we here briefly discuss some views that could inform further research.

What is important, at least in our view, is that changing people's ways of thinking, and their behaviour, is not an exclusive feature of emerging technologies that can be used for smart rehabilitation. In fact, in criminal justice we often try to change how convicted offenders think and behave all the time, for example through coerced participation in treatment programs for sexual offenders or addicts, but also by imprisonment, which is often intended in part to stimulate self-reflection and deter future offending by the imprisoned individual.<sup>80</sup> These interventions induce mental changes. However, traditional criminal justice interventions such as these are not normally considered to raise any issues under existing legal rights such as freedom of thought and mental liberty.<sup>81</sup> Perhaps this is because those who commit serious crimes are deemed to have forfeited aspects of the right to mental liberty or because the objectives of criminal justice are thought sufficiently important to justify infringement of the right. Or perhaps there is something about the means or significance of the alterations produced by these interventions that takes them beyond the scope of the right. Either way, if one were to argue that smart rehabilitation infringes, or even violates, a person's right to mental liberty, one should either identify some compelling distinctions between smart rehabilitation and existing criminal interventions that also (aim to) change how criminal offenders think and behave, or one should concede that many of the latter infringe (or violate) the right as well.

Probably, compelling distinctions would not be found in the significance of the mental alterations that are (indirectly) caused by PTs nor in the *intentions* that motivate them. For example, imagine that a convicted offender of sexual assault is allowed to leave prison on parole, under the condition that he participates in a sex offender treatment programme that aims to change his thoughts about sexual relationships, intending to ultimately change his sexual behaviour. However, suppose that there is now also an app-based PT, which aims to alter the exact same mental states, intending to change the exact same sexual behaviour, and which is equally successful in achieving these aims. Could one argue that requiring the use of this PT as a condition of parole would infringe the right to mental liberty, while requiring participation in the treatment programme would not? In both cases, the *significance* of the mental alterations as well as the *intention* with which they are produced are the same. What differs is the *means* that produces the change of mind: participating in the programme's courses versus a communicative mobile app.

Whether this difference should be relevant in the context of mental liberty can be debated. On the one hand, one could argue that the PT-app is more intrusive, since it is not only effective during particular times at particular places, as the courses of a treatment programme are, but it

<sup>80</sup> Which might even influence certain brain functions: R. Dillon, 'Banning Solitary for Prisoners with Mental Illness: The Blurred Line between Physical and Psychological Harm' (2019) 14 *Northwestern Journal of Law & Social Policy* 265; R. Umbach, A. Raine and N. R. Leonard, 'Cognitive Decline as a Result of Incarceration and the Effects of a CBT/MT Intervention: A Cluster-Randomized Controlled Trial' (2018) 45 *Criminal Justice and Behavior* 31–55; J. Meijers et al. 'Reduced Self-Control after 3 Months of Imprisonment; A Pilot Study' (2018) 9 *Frontiers in Psychology*.

<sup>81</sup> See Bublitz (note 78), pp. 77–81. See cases in which the Court considered that (compulsory) courses of sex education did not amount to indoctrination: ECtHR 13 September 2011, appl.no. 319/08 (*Dojan and others v. Germany*); ECtHR 7 December 1976, appl.nos. 5095/71, 5920/72, 5926/72 (*Kjeldsen and others v. Denmark*).

actually tries to influence the user's mental states at any time, at any place: even in the bedroom and bathroom, where a tutor would normally not be allowed.<sup>82</sup> In addition, whereas in human–human persuasion each party can stop the interaction, recognise emotions, ask for clarification and show in various ways that one feels uncomfortable with the influential process, PTs are limited in their capacities for this kind of two-directional communicative interaction.<sup>83</sup> Instead, PTs are 'proactively persistent': they do not get tired, embarrassed, feeling uneasy or guilty – persuasive messages can pop up continuously.<sup>84</sup> As Smids writes, 'PTs can continue their persuasive attempt until the user capitulates in either a moment of weakness or a moment of unawareness.'<sup>85</sup> As a consequence, one could argue that PTs might be *less resistible* than, for example, a sexual treatment programme.<sup>86</sup>

On the other hand, however, one could argue that since both PTs and treatment programmes typically engage rational processes to at least an extent, both are unproblematic with respect to mental liberty. Both considerations – the more limited resistibility of PTs and the fact that PTs somehow engage rational processes – could be relevant to whether the right to mental liberty is infringed. However, how we should balance these considerations is yet unclear.

### 3.6 CONCLUSION

In this chapter we discussed some possibilities that PTs offer for the smart correctional rehabilitation of criminal offenders. We explored whether such smart rehabilitation might infringe a right to mental liberty. Since the scope of existing human rights to freedom of thought and to respect for one's mental liberty are as yet ambiguous, it is unclear whether they protect against smart rehabilitation. We argued that, in order to specify the scope of human rights protections against smart rehabilitation, at least three considerations are relevant: the significance of the mental alterations that are caused, the intentions for which they are produced and the means via which they are produced. Whether smart rehabilitation would, in light of these considerations, infringe the right to mental liberty remains unclear and deserves closer examination. In this context, it should be relevant to what extent compelling distinctions can be identified between smart rehabilitation and current interventions that intend to alter the minds of criminal offenders.

<sup>82</sup> Jacobs (note 1), p. 3; Fogg (note 1), p. 10–11.

<sup>83</sup> Smids (note 1), p. 24; Fogg (note 1), pp. 216–218.

<sup>84</sup> Fogg (note 1), p. 216; Smids (note 1), p. 24.

<sup>85</sup> Smids (note 1), p. 24.

<sup>86</sup> Note that other arguments are conceivable as well. For example, one could contend that unlike participation in a mandatory treatment programme, messages on a smartphone can easily be ignored, e.g., by downloading the mandatory PT app to an old smartphone that is not normally used anymore.