Neural Implants and the TRICK to Autonomy

MAXIMILIAN KIENER & THOMAS DOUGLAS[1]

Forthcoming in *Ethics in Practice* (6th Edition), ed. Hugh LaFollette (Wiley-Blackwell)

Biomedical technologies do not merely influence our bodies. They are starting to have a profound impact on our minds too. One area where this is especially salient is the burgeoning field of brain implants (Hughes 2014; Cho et al. 2021; Kawala-Sterniuk et al. 2021). These devices can bridge the gap between the biological and the digital and have the potential to treat neurodegenerative diseases, enhance cognitive capabilities, and even integrate AI-enhanced functionalities directly into our neural makeup. Modern neural implants could even go beyond therapeutic applications and create a future of augmented cognition in which our brains interface seamlessly with digital information. However, as Gilbert and colleagues note, the 'concern about (…) effects on patients' sense of self, autonomy and identity is growing' (Gilbert et al. 2019, 84). See also (Klein 2016; Glannon 2016; Clausen et al. 2017).

In this article, we concentrate on the concern that neural implants might, even when initially employed with the patient's free and informed consent, still impede an individual's autonomy. We propose a novel framework called TRICK for assessing this concern. The acronym stands for Transparency, Rationality, Irresistibility, and Consent Kinetics (a term of art that we will explain later). TRICK builds on insights from the neighbouring debate on nudging, where considerations relating to transparency, rationality, and (ir-)resistibility have been prominently explored, but also highlights the limitations of, qualifies, and supplements these principles. We first motivate TRICK by explaining why the literature on nudging is relevant to the appraisal of neural implants, and then explain each component principle in turn.

[1] We thank Jan Christoph Bublitz, Marcello Ienca, David Storrs-Fox, and the WIP discussion group at the Institute for Ethics in Technology at Hamburg University of Technology for feedback on an earlier version of this paper.

The Link between Nudging and Neural Implants

Nudges are techniques that influence our behaviour by prompting the use of rules of thumb like, ‘stick with what you know’, ‘don’t fix what ain’t broken’ or ‘listen to people you recognise’. Examples include placing salads in prominent locations in cafeterias to promote healthier food choices, making pension contributions the default option for new employees, or using heart-warming images to increase donations to a charity.

Nudges and neural implants certainly differ. Nudges work through arranging the environment in which people make decisions, while neural implants work through direct physical interaction with people’s brains. Yet, they both influence people’s minds and decision-making. Thus, it may not be too far-fetched to consult the former debate to enrich our understanding of the latter.

Nudges have been criticised for impeding the ‘nudgee’s’ autonomy. But many have argued that at least some nudges preserve, or even promote, autonomy. To determine the impact of nudges on autonomy, a number of considerations have been proposed (cf. Kiener 2021b), converging towards three main principles.

The first principle is transparency: the nudge must be in some sense observable. Hansen and Jespersen claim that ‘the distinction between transparent and non-transparent nudges (…) serves as a basis for distinguishing the manipulative use of nudges from other types of uses’ (Hansen and Jespersen 2013, 6); and Bovens even argues that ‘every Nudge should be such that it is in principle possible for everyone who is watchful to unmask the manipulation’ (Bovens 2009, 217) (where ‘manipulation’ need not carry a pejorative meaning but only signify a type of influence.)

The second principle is rationality: a nudge must not undermine or subvert a person’s rationality.[2] It should still be the case that a person’s

[2] See, for example, Wilkinson (2013, 349); Cohen (2013, 5): 5; Engelen (2019, 206); Schmidt (2019, 515); Bovens (2009). Some hold further that nudges must not harness

‘decision-making procedure should yield accurate results somewhat reliably and not just accidentally’ (Schmidt 2019, 512). After all, there is a ‘legitimate worry that nudges decrease people’s process-rationality’ (Engelen 2019, 220) or, more generally, that they make 'people less rational than they are, can be or should be' (Engelen 2019, 204).

Finally, the third principle is resistibility: nudgees should be able to resist the nudge. The pioneers of nudging, Thaler and Sunstein, demanded from the outset that, to preserve autonomy (and indeed even to count as a ‘nudge’ in the first place), influences ‘must be easy and cheap to avoid’ (Thaler and Sunstein 2009, 6). People should be able to avoid exposure to or block the effect of a nudge and it should not require any significant psychological effort, special resilience, or insensitivity to do so (Saghai 2013; Wilkinson 2013).

We propose that analogues of these principles could be used to assess the threat to autonomy posed by neural implants.[3] To see how this might work, let us focus on the treatment of severe epilepsy where neural implants have played an important role (Nune, DeGiorgio, and Heck 2015; Kremen et al. 2018; Romanelli et al. 2018; Gilbert, Ienca, and Cook 2023). Let us imagine a possible further implant, which we call ‘NeuroGuard’, for patients with this condition. The implant predicts and prevents seizures by monitoring neural activity and delivering corrective electrical pulses when neural changes suggesting an imminent seizure are detected. An associated app provides the patient with real-time information about the implant’s activity so that they can decide whether or not to allow the implant to intervene on a given occasion. AI plays an important role too; it allows the device to learn over time, analysing vast amounts of neural data in real-time to become better and better at identifying early neural signatures that indicate potential seizure onset in the particular individual, thus enhancing the timeliness and precision of interventions.

Consider the following scenario:

irrational or bypass rational processes within the nudgee. See, for example, Bovens (2009, 209), Conly (2013, 30), Schmidt (2019, 511), MacKay and Robinson (2016, 3-4).

[3] See also the growing literature that explores how nudges and novel digital technologies start to intersect, e.g. Ienca and Vayena (2021)

> ***Epilepsy.*** Maria, a 35-year-old journalist, has suffered from severe epilepsy since childhood. Traditional medications have been ineffective and her unpredictable seizures make independent living dangerous. Maria has also been very anxious about her condition and her fear often distorts a realistic assessment of her situation. Hearing about NeuroGuard, Maria undergoes the implantation procedure. Once activated, the device works wonders for her. Having previously experienced weekly seizures, Maria now suffers around one or two per year, and her anxiety is greatly attenuated.

This scenario shows how brain implants could promote a patient's autonomy—their ability to control the course of their own life. Maria now has greater control over whether she will be afflicted by a seizure, and thus over what activities she can safely pursue. What is more, it seems that the three abovementioned principles, which were developed in the debate on 'nudges', can help to explain why Maria's autonomy was promoted.

In Maria's case, her app informs her about each possible intervention of the neural implant and thus gives her full transparency regarding the operation of the implant; and it seems that such transparency is an important factor for Maria retaining and even increasing control over her life. Moreover, regarding rationality, Maria's anxiety previously distorted her judgment, but with the implant in place, her anxiety recedes. This plausibly improves her ability to appreciate her situations more appropriately, making Maria more rational overall. Finally, consider resistibility. It seems important for her autonomy that being in the loop gives Maria the opportunity to act against any intervention and thus, let us assume, makes it easy and cheap for her to veto any activity of the implant. Maybe Maria could even identify with the following statement from a patient with a similar neural implant: 'I felt more in control when I used the device. I could push on [or off] and do what I wanted to do' (Reported in Gilbert 2015, 7); or another patient

who said ‘I felt like I could do anything [now]’ (Reported in Gilbert, Ienca, and Cook 2023, 785).

Thus, it seems that the debate on nudging has relevance for the ethical assessment of neural implants: in particular, the former provides three specific principles that could guide the assessment of the effect that neural implants have on user’s autonomy. In what follows, we shall deploy the three principles to draw out further insights concerning the effect of neural implants on autonomy. Yet, we also show that the three principles have significant limitations, require further qualification, and need to be supplemented with an additional desideratum concerning consent.

Unless otherwise stated, we understand autonomy as an exercise concept, which refers to the exercise of certain capacities of thought and agency, most notably the capacity to make and execute decisions that are in line with one’s own values, beliefs, and preferences. So understood, autonomy involves not only formulating plans but also actively carrying them out, and this could be done either independently or with support from others (cf. Lillehammer 2012, 197). Accordingly, we investigate how, and to what extent, neural implants impede or promote the exercise of these capacities.

Transparency

Unlike in Maria’s initial case, a lack of transparency could sometimes be desirable. Consider the following case:

> ***Stealth Mode***. Maya, a scientist dealing with disruptive and unsafe epileptic seizures, utilises NeuroGuard which, under normal circumstances, supplies predictive analytics on potential seizure incidents. However, Maya experiences anticipatory anxiety knowing a seizure is forthcoming—nearly as incapacitating as the seizures themselves. There is the option of switching the device into a “Stealth Mode”, in which it would predict and counter seizures without informing Maya of impending episodes, thus obviating her anxiety. Following

> comprehensive discussions with the medical team, Maya, who understands the trade-offs involved, elects to activate Stealth Mode.

This example shows that full transparency does not always promote autonomy: for Maya, full transparency would cause severe anxiety thereby impede her autonomy. Moreover, in addition to causing anxiety, full transparency might also cause informational overload, overburdening Maja with complex information rather than specifically assisting her decision-making.

Additional challenges could arise when patients have the control to permit or veto any single intervention of a device, or even change its settings. Admittedly, such control may promote *some* people's autonomy, as we suggested it did in Maria's case *Epilepsy*. But for others, it may not. As Klein and colleagues note: 'if a patient is given control over device settings, the temptation to increase stimulation settings to feel better and better may be difficult to resist, and patients may fear the introduction of a new kind of addiction' (Klein et al. 2016, 2).[4] Here is an example:

> ***New Calm.*** Daniel can control NeuroGuard's activities and finds a setting that brings him profound tranquillity. Over time and captivated by this newfound calm, Daniel began choosing more frequent interventions and higher intensities. This pursuit of continuous tranquillity begins dominating his life.

Although NeuroGuard's transparency was meant to empower patients, in Daniel's case, it inadvertently facilitated an obsessive quest for non-therapeutic emotional states. Thus, full transparency could carve the way to further influences that impede rather than promote autonomy.

We propose that it is possible to capture the significance of transparency while accommodating cases like *Stealth Mode* and *New*

[4] See Palacios-González (2015) and Klein (2015). This is noteworthy as it is normally thought that transparency increases rather than decreases resistibility. For discussion of the relationship between transparency and resistibility in relation to nudges, see De Marco and Douglas (*Forthcoming*).

Calm by means of two complementary strategies. First, one could preclude *type* transparency and only pursue *token* transparency. *Type* transparency is transparency about the general use of the neural implant, for example, that it is in a person's brain and operates to prevent seizures. *Token* transparency, by contrast, is transparency regarding individual instances in which the device operates. We suggest that token transparency will sometimes be detrimental to autonomy overall—as it was for Maya and Daniel—because it causes anxiety or informational overload or facilitates compulsion. By contrast, *type* transparency does not have these risks, or not to the same degree, and remains important for a person to retain the option of autonomously refusing or terminating an intervention. Second, we introduce delayed disclosure as an additional strategy. This approach involves withholding specific operational details of the neural implant temporarily (i.e. withholding full type transparency) to manage the potential negative impacts on the user's autonomy—such as anxiety, informational overload, or compulsion—while still committing to eventual type transparency at a later point. This delayed approach can be particularly beneficial as it allows individuals to process information at a more manageable pace.[5] Overall, we suggest that, in assessing the effects of neural devices on autonomy, it will be important, first, to prioritise *type* transparency and then, second, implement *token* transparency only if, or *when*, doing so enhances the individual patient's autonomy.

Rationality

In *Epilepsy*, our starting example, Maria's rationality was plausibly enhanced by NeuroGuard. Her anxiety reduced and she became better able to appreciate her situation. But in other cases, neural implants can affect people's decision-making in a problematic way. This becomes especially apparent in the use of neural implants for the treatment of bipolar disorder (Holtzheimer et al. 2012; Gippert et al. 2017; Mutz 2023). To see how, let us consider another possible future neural implant,

[5] We thank David Storrs-Fox for raising this issue of delayed disclosure.

call it the ‘EmoTune Neural Chip’, which is designed to regulate and balance emotional responses in patients with severe mood disorders. It works by detecting specific neural patterns associated with extreme emotional states and sending small electrical pulses to adjust brain activity, aiming for emotional stabilisation. In EmoTune, AI could dynamically regulate emotional states by deciphering neural patterns associated with various emotions through deep learning algorithms. This may include pre-emptively recognising emotional states that could benefit from modulation and subsequently triggering neural interventions. Here is an example:

> ***New Job.*** Robert, a 40-year-old with a long history of bipolar disorder, has the EmoTune chip implanted after traditional treatments prove ineffective. The chip successfully stabilises his emotions and prevents extreme emotional swings. A few months later, Robert has to make a significant decision: whether to accept a high-pressure job offer in a city far away from his social support system. Rationally, moving would mean leaving behind his family, friends, and therapist. The job also involves longer hours and more stress, factors that previously exacerbated his condition. However, every time Robert begins to feel anxious while considering the offer, emotions that would typically inform his rational decision-making process, the EmoTune chip activates and stabilises his emotions. As a result, the natural emotional cues that would help Robert weigh the potential consequences of his decision are dampened and Robert ends up being indifferent about the possible challenges and risks. Robert eventually decides to take the job, thinking it is a logical step forward in his career. However, once relocated, the cumulative stress and isolation begin to affect his mental health and lead to unforeseen complications.

The EmoTune Neural Chip is effective in managing Robert’s bipolar symptoms. But it also negatively affects his decision-making and mutes emotional cues integral to informed judgments, thereby plausibly decreasing Robert’s overall rationality, i.e. his ability to respond to reasons and appraise his situation accurately. Robert’s case showcases

the role of emotions in rationality and meaningful decision-making. As Charland says:

> Individuals cannot be said to appreciate fully the choices they face unless the choices mean something to them personally [and] emotions are one important source of value. They define and shape many of one's most basic preferences, which in turn help to define and shape many of one's goals and values. (Charland 1998, 368)[6]

But although the curtailment of emotions may reduce rationality and, by extension, impede autonomy, as it seems to do in Robert's case, the relation between emotions, rationality, and autonomy is complicated, as the following example illustrates:

> ***Healing.*** Taylor, grappling with emotional flashbacks and anxiety post-trauma, elects to utilise EmoTune, under professional advice, to momentarily suppress emotions that he experiences as negative. Taylor's decision-making is curtailed like Robert's in *New Job.* Yet, this suppression paves the way to productive therapeutic involvement as well as daily activities, and it also gives Taylor some emotional breathing space, necessary for his psychiatric recovery.

At first glance, the *Healing* and *New Job* cases are close parallels. Both protagonists, Robert and Taylor, experience reduced emotions and are arguably in one way less rational: they are more emotionally disconnected from reality and thus (to build on Charland's view) less able to appraise what certain options or circumstances mean to them personally. Yet, for Taylor, unlike for Robert, the temporary loss of emotions and emotional connectedness is a strategic measure to promote long-term autonomy. Just as a climber must on some occasions momentarily move away from their goal, stepping sideways or downwards, to circumvent an obstacle and discover a new path to the

[6] The role of emotions for our rationality has been widely acknowledged. (Scarantino and de Sousa 2018; Hursthouse 1991; Railton 2014; Damasio 2006; Greenspan 2004; Nussbaum 2001)

destination, so might Taylor have to opt for a temporary step away from (certain forms of) rationality to ensure long-term progress.

Thus, just as we could justify a reduction in transparency for the sake of promoting autonomy, we can also sometimes justify a reduction in rationality for the same purpose. But there is an interesting difference: the reduction of transparency was directly intended in Maya's case and a means to preclude anxiety, informational overload, or compulsion, whereas the reduction of rationality was not a means to something else, but only a foreseen side-effect of the required curtailment of certain emotions.

Irresistibility

Discussing Maria's case, *Epilepsy*, we presented resistibility as a consideration that would be relevant in assessing the effects of the intervention on Maria's autonomy. In the initial scenario, we assumed that Maria could easily resist and act against the influence of her implant, simply by blocking it using an app on her phone, and that this suggested that the intervention was conducive to her autonomy. Moreover, we added that if Maria, or others using a similar device, experienced pressure through the constant notifications, making the influence of the device hard to resist, this could be a problem for the protection of autonomy.

However, we have also already encountered cases where the lack of an ability or opportunity to resist the influence of an implant was not problematic. Recall *Stealth Mode*, where Maja opted for a 'set it and forget it' option; the NeuroGuard device would influence her mind without her being in the loop anymore. We suggested that this could promote her autonomy.

We can further build on such cases and consider scenarios in which the patient remains in the loop but is likely to find it difficult to resist the technology's influence. Consider the following case:

> ***Persuasive Partner.*** Alex, a college student with epilepsy, relies on NeuroGuard. Whenever NeuroGuard detects potential seizure triggers in Alex's brain, such as during prolonged study sessions or exposure to loud noises, it electrically stimulates neural circuits so as to create an inclination for Alex to take certain seizure-preventing actions. For example, if Alex needs to take a break from studying to prevent a seizure, the device stimulates regions of his brain associated with relaxation and fatigue, making him feel a strong urge to rest. Similarly, if the noise level is too high, the device affects Alex's auditory processing in such a way that he instinctively seeks a quieter environment. This direct brain intervention makes it hard for Alex to resist the device's recommendations. However, NeuroGuard succeeds in significantly reducing Alex's seizure risk.

Suppose that, in many of the cases in which NeuroGuard intervenes, the actions caused are ones that promote Alex's autonomy, by freeing him from the constraint of frequent seizures. Moreover, suppose that in some of these cases, Alex would not have taken the autonomy-promoting actions had NeuroGuard employed easier-to-resist influences. In other words, NeuroGuard serves as a guardrail in situations where Alex himself would be incapable of staying on track on his own. Under these assumptions, the employment of hard-to-resist influences will tend to promote, rather than impede, Alex's overall autonomy.

Moreover, such use of neural implants is not vulnerable to some of the criticisms that have been levelled against nudges or manipulative tactics more generally, where scholars most strongly advocated a requirement of resistibility. For instance, Noggle argues that certain influences become manipulative and undermine autonomy when they skew our perception of what is truly important, making certain things appear more prominent to us than they really are (Noggle 2017). However, in Alex's situation, this does not hold true. The device's influence is carefully calibrated to align with his personal goal of taking care of his health. Furthermore, Bovens criticises manipulative tactics for causing a disconnect between what motivates our actions and what

should rationally justify them (Bovens 2009, 210). While the factors influencing Alex's behaviour are still not equivalent to rational justifications, they do correlate with and reinforce his rational reasons, boosting Alex's ability to track and respond to the reasons that apply to him. Consequently, Boven's general critique loses at least much of its bite against the device's impact on Alex. Thus, despite its somewhat irresistible nature, NeuroGuard and the app can promote Alex's autonomy and is less vulnerable to some key criticisms that they were raised against nudges and manipulative influences, although it may, of course, still require scrutiny with regard to others. (One aspect that may threaten the promotion of Alex's autonomy is that the control over his actions is partially externalised to an AI-powered device.)

Overall, we should understand low resistibility not as perfect indicator that a technology will diminish a person's autonomy, but as an imperfect proxy. Whenever there is a strong influence that people cannot easily resist, this is a first indication that there might be an impediment to their autonomy. But as the cases of Maja *(Stealth Mode)* and Alex *(Persuasive Partner)* show, the lack of easy resistibility could sometimes also promote certain aspects of autonomy, that is the exercise of the capacity to make and execute decisions that are in line with one's own values, beliefs, and preferences.

The Three Principles Re-Considered

So far, our analysis has produced mixed results. In our initial case, *Epilepsy*, principles of transparency, rationality, and resistibility suggested that Maria's autonomy was enhanced, and this, we think, is the right verdict on this case. But in subsequent examples, we have also found that these principles do not accurately predict a technology's effect on a patient's autonomy. In this section, we make some of the limitations of these principles more explicit.

Consider first transparency. Transparency promotes autonomy insofar as it facilitates opportunities for meaningful control, but it can also impede autonomy, for example when complete information about the

neural implant's functions might induce anxiety, hamper the therapeutic efficacy of the device, or overwhelm the user with technical details that are not directly pertinent to their experience or use of the technology. We outlined the concepts of token and type transparency, alongside a strategy of delayed disclosure, to illustrate how transparency can serve as an effective tool for enhancing, rather than impeding, autonomy. However, the extent to which transparency enhances autonomy will depend on the specific circumstances of each individual case.

Consider next rationality. Promoting rationality also promotes autonomy insofar as it helps people to reach decisions that better reflect their genuine values and preferences, but how technology affects a person's rationality can be a complicated matter. One reason for this is that rationality has an emotional component; emotions can be an impediment to rationality, as we saw in *Healing,* but they can also support it, as we saw in *New Job*. Moreover, as with transparency, there might be instances where temporarily diminishing rationality may facilitate therapeutic outcomes and enhance long-term autonomy.

Finally, resistibility is a good, but imperfect proxy for the promotion of autonomy. Influences that are hard to resist may impede a person's autonomy. But as our examples showed, this is not invariably true and we need to assess the overall context of an influence, and not just its psychological strength. Just as the metaphor of a climber who steps downwards or sidewards to avoid an obstacle could help explain why temporarily reducing rationality could enhance autonomy, the metaphor of a guardrail could help explain why the lack of resistibility can sometimes help people to stay on (their self-chosen) track.

Consent Kinetics

We have been assuming throughout that neural implants will be used with the patient's *initial* consent.[7] However, consent is not a single

[7] See also Eran Klein's further work on consent in the context of BCIs. Klein outlines several areas of disclosure of risks, some of which also transcend normal disclosure requirements, such as risks of security and cyber-attacks (Klein 2016). On extended medical disclosure and AI, see also (Kiener 2021a).

discrete event, but an ongoing process. Even when people give their initial valid consent, such consent can lapse, either because the person's decision-making capacity has deteriorated or because the situation has changed in a way that requires renewed consent.[8] Second, the right to authorise a new medical procedure includes not only the right to *give* initial consent, but also 'the right to stop a [medical] procedure', that is, *withdraw* consent (Ciarlariello v Schacter [1993], 121) (Grubb et al. 2010, 449-452). To see how these features pose challenges in the context of neural implants, consider the following case:

> ***Lost Passion.*** Victor, a skilled pianist, has long found solace and expression in his music, particularly as a means of navigating life with bipolar disorder. On the advice of his psychiatrist, he agreed to EmoTune. Though EmoTune was effective in managing his mental health, it unexpectedly dampened his once fervent passion for music.
>
> ***Subsequent Scenario 1: Continued Consent.*** At a later point, the treatment with EmoTune develops side-effects, one of which significantly reduces Victor's dexterity. There are treatment alternatives which could preserve Victor's dexterity and musical abilities, but since Victor has lost interest in music, he is not interested in them and continues to consent to EmoTune, despite the side effects.
>
> ***Subsequent Scenario 2: Withdrawn Consent.*** EmoTune continues to be effective without side-effects, but Victor withdraws consent to a different treatment—a drug that he had been taking to reduce a hand tremor—despite the fact that stopping this treatment will prevent him from playing the piano well. As he is now apathetic regarding his musical career, he does not find the tremor to be a problem.

In both subsequent scenarios, we could assume that Victor passes the current capacity tests. He may also be well-informed about the consequences of his decisions, and be free from external interferences on

[8] Klein and Ohemann also emphasised the importance of consent as either an ongoing or iterative process, rather than a discrete event (Klein and Ojemann 2016). Further discussion in: (Lidz, Appelbaum, and Meisel 1988; Wallace et al. 2016; Maclean 2009).

his decision-making. Yet, the cases remain problematic, despite Victor's initial consent to EmoTune, because it is not clear that Victor's newfound indifference to music is itself an expression of his autonomy. For this reason, in *Subsequent Scenario 1*, it would be reckless of the medical team to simply accept Victor's continued consent without further support. The requirements for Victor's continued consent ought to be especially high. The lesson to be drawn then is that continued consent cannot just be a matter of maintaining capacity but requires attention to a person's values and how they came about.

Moving on to *Subsequent Scenario 2*, the problems are similar. Yet, in clinical practice, the withdrawal of consent is normally even less regulated than the monitoring of continued consent. When people want to withdraw their consent, they often need not give any reason at all and it is even deemed inappropriate (sometimes justifiably so) if the medical teams insist on a conversation. But this currently wide-spread *laissez-faire* model of consent withdrawal becomes problematic in Victor's case. We suggest that *Subsequent Scenario 1* and *Subsequent Scenario 2* are on a par, morally speaking, so that accepting greater scrutiny in the former case commits us to greater scrutiny in the later. Thus, we think that it would also be reckless of the medical team to accept Victor's withdrawal without further discussion. Ultimately, they may not be able to proceed to compulsory treatment, but they are obliged to inform Victor about the risks of withdrawal and offer decision-making support throughout. We are inclined to think that, in a case like this, treatment should be withdrawn only after several discussions, over a period of time, taking into account the fluctuating influence of EmoTune, or after weakening or pausing EmoTune for the duration of Victor's treatment decisions. Alternatively, the medical team may have to obtain *meta*-consent earlier, i.e. the patient's consent (ahead of any treatment by a neural implant) to how his continued consent ought to be handled.

In conclusion, then, use of neural implants may result in a failure to promote autonomy if (as in *Subsequent Scenario 1*) they rely on their own psychological effects to sustain continued consent from the

recipient, or if (as in *Subsequent Scenario 2*) they produce psychological effects that make a patient withdraw consent. We suggest the term *consent kinetics* for these considerations. Kinetics concerns the physical forces acting upon bodies and the motion produced by those bodies. *Consent kinetics*, as we understand it, concerns the manifold normative forces acting upon (continued or withdrawn) consent and the moral implications of these forces. According to our proposal, consent kinetics are highly relevant when assessing the autonomy implications of neural implants.

Conclusion

In this paper, we introduced TRICK as a novel framework to evaluate autonomy-related concerns relating to neural implants. TRICK stands for Transparency, Rationality, Irresistibility, and Consent Kinetics. It explores the complex interplay between technology and autonomy, particularly in how neural implants, while potentially enhancing autonomy, for example, by mitigating diseases, can also introduce new limitations on autonomy. These limitations include potential impacts on oversight, rational and emotional capabilities, and the introduction of influences that could be psychologically irresistible. TRICK aims to make a meaningful contribution to clinical practice, but we have not argued that it provides an exhaustive solution to the myriad of emerging challenges. Nevertheless, TRICK could be an important step towards a more robust foundation for patient autonomy in the context of biomedical technologies.

References

Bovens, L. 2009. "The Ethics of Nudge." In *Preference Change Approaches from Philosophy, Economics and Psychology.*, edited by T. Grüne-Yanoff and S. Hansson, 207-219. Springer.

Charland, Louis. 1998. "Appreciation and Emotion: Theoretical Reflections on the MacArthur Treatment Competence Study."

Kennedy Institute of Ethics Journal 8 (4): 359-376. https://doi.org/ https://doi.org/10.1353/ken.1998.0027.
Cho, Younguk, Sanghoon Park, Juyoung Lee, and Ki Jun Yu. 2021. "Emerging materials and technologies with applications in flexible neural implants: a comprehensive review of current issues with neural devices." *Advanced Materials* 33 (47): 2005786.
Clausen, Jens, Eberhard Fetz, John Donoghue, Junichi Ushiba, Ulrike Spörhase, Jennifer Chandler, Niels Birbaumer, and Surjo R Soekadar. 2017. "Help, hope, and hype: Ethical dimensions of neuroprosthetics." *Science* 356 (6345): 1338-1339.
Cohen, Shlomo. 2013. "Nudging and informed consent." *The American Journal of Bioethics* 13 (6): 3-11.
Conly, Sarah. 2013. *Against autonomy: justifying coercive paternalism*. Cambridge: Cambridge University Press.
Damasio, Antonio R. 2006. *Descartes' error : emotion, reason and the human brain*. London: Vintage.
De Marco, Gabriel, and Thomas Douglas. *Forthcoming*. "Nudge Transparency Is Not Required for Nudge Resistibility." *Ergo*.
Engelen, Bart. 2019. "Nudging and rationality: What is there to worry?" *Rationality and Society* 31 (2): 204-232.
Gilbert, Frederic. 2015. "A threat to autonomy? The intrusion of predictive brain implants." *Ajob Neuroscience* 6 (4): 4-11.
Gilbert, Frederic, Mark Cook, Terence O'Brien, and Judy Illes. 2019. "Embodiment and estrangement: results from a first-in-human "intelligent BCI" trial." *Science and engineering ethics* 25: 83-96.
Gilbert, Frederic, Marcello Ienca, and Mark Cook. 2023. "How I became myself after merging with a computer: Does human-machine symbiosis raise human rights issues?" *Brain Stimulation* 16 (3): 783-789.
Gippert, Sabrina M, Christina Switala, Bettina H Bewernick, Sarah Kayser, Alena Bräuer, Volker A Coenen, and Thomas E Schlaepfer. 2017. "Deep brain stimulation for bipolar disorder—review and outlook." *CNS spectrums* 22 (3): 254-257.
Glannon, Walter. 2016. "Ethical issues in neuroprosthetics." *Journal of Neural Engineering* 13 (2): 021002.
Greenspan, Patricia. 2004. "Practical Reasoning and Emotion." In *The Oxford Handbook of Rationality*, edited by Alfred R Mele and Piers Rawling. New York: Oxford University Press.
Grubb, Andrew, Judith M. Laing, Jean V. McHale, and Ian Kennedy. 2010. *Principles of medical law*. 3rd ed. Oxford: Oxford University Press.
Hansen, Pelle Guldborg, and Andreas Maaløe Jespersen. 2013. "Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy." *European Journal of Risk Regulation* 4 (1): 3-28.
Holtzheimer, Paul E, Mary E Kelley, Robert E Gross, Megan M Filkowski, Steven J Garlow, Andrea Barrocas, Dylan Wint, Margaret C Craighead, Julie Kozarsky, and Ronald Chismar. 2012. "Subcallosal cingulate deep brain stimulation for treatment-resistant unipolar

and bipolar depression." *Archives of general psychiatry* 69 (2): 150-158.
Hughes, MA. 2014. "Engineering brain-computer interfaces: past, present and future." *Journal of neurosurgical sciences* 58 (2): 117-123.
Hursthouse, Rosalind. 1991. "Arational Actions." *The Journal of philosophy* 88 (2): 57-68. https://doi.org/10.2307/2026906.
Ienca, Marcello, and Effy Vayena. 2021. "Digital Nudging: Exploring the Ethical Boundaries." In *Oxford Handbook of Digital Ethics*, edited by Carissa Véliz, 356-377. Oxford: Oxford University Press.
Kawala-Sterniuk, Aleksandra, Natalia Browarska, Amir Al-Bakri, Mariusz Pelc, Jaroslaw Zygarlicki, Michaela Sidikova, Radek Martinek, and Edward Jacek Gorzelanczyk. 2021. "Summary of over fifty years with brain-computer interfaces—a review." *Brain Sciences* 11 (1): 43.
Kiener, Maximilian. 2021a. "Artificial intelligence in medicine and the disclosure of risks." *Ai & Society* 36 (3): 705-713.
---. 2021b. "When do nudges undermine voluntary consent?" *Philosophical Studies*: 1-26.
Klein, Eran. 2015. "Are brain-computer interface (BCI) devices a form of internal coercion?" *AJOB Neuroscience* 6 (4): 32-34.
---. 2016. "Informed consent in implantable BCI research: identifying risks and exploring meaning." *Science and Engineering Ethics* 22: 1299-1317.
Klein, Eran, Sara Goering, Josh Gagne, Conor V Shea, Rachel Franklin, Samuel Zorowitz, Darin D Dougherty, and Alik S Widge. 2016. "Brain-computer interface-based control of closed-loop brain stimulation: attitudes and ethical considerations." *Brain-Computer Interfaces* 3 (3): 140-148.
Klein, Eran, and Jeffrey Ojemann. 2016. "Informed consent in implantable BCI research: identification of research risks and recommendations for development of best practices."
Kremen, Vaclav, Benjamin H Brinkmann, Inyong Kim, Hari Guragain, Mona Nasseri, Abigail L Magee, Tal Pal Attia, Petr Nejedly, Vladimir Sladky, and Nathanial Nelson. 2018. "Integrating brain implants with local and distributed computing devices: a next generation epilepsy management system." *IEEE journal of translational engineering in health and medicine* 6: 1-12.
Lidz, Charles W, Paul S Appelbaum, and Alan Meisel. 1988. "Two models of implementing informed consent." *Archives of Internal Medicine* 148 (6): 1385-1389.
Lillehammer, H. . 2012. "Autonomy, Value, and the First Person." In *Autonomy and Mental Disorder*, edited by L. Radoilska 192–213. Oxford University Press.
MacKay, Douglas, and Alexandra Robinson. 2016. "The ethics of organ donor registration policies: Nudges and respect for autonomy." *The American Journal of Bioethics* 16 (11): 3-12.
Maclean, Alasdair. 2009. *Autonomy, informed consent and medical law: a relational challenge*. Cambridge, UK: Cambridge University Press.

Mutz, Julian. 2023. "Brain stimulation treatment for bipolar disorder." *Bipolar Disorders* 25 (1): 9-24.
Noggle, Robert. 2017. "Manipulation, salience, and nudges." *Bioethics* 32 (3): 164-170.
Nune, George, Christopher DeGiorgio, and Christianne Heck. 2015. "Neuromodulation in the treatment of epilepsy." *Current treatment options in neurology* 17: 1-6.
Nussbaum, Martha C. 2001. *Upheavals of thought : the intelligence of emotions*. Cambridge: Cambridge University Press.
Palacios-González, César. 2015. "Epilepsy, Decisional Vulnerability, and the Nature of Predictive Brain Implants." *American Journal of Bioethics Neuroscience* 6 (4).
Railton, Peter. 2014. "The Affective Dog and Its Rational Tale: Intuition and Attunement." *Ethics* 124 (4): 813-859. https://doi.org/10.1086/675876.
Romanelli, Pantaleo, Marco Piangerelli, David Ratel, Christophe Gaude, Thomas Costecalde, Cosimo Puttilli, Mauro Picciafuoco, Alim Benabid, and Napoleon Torres. 2018. "A novel neural prosthesis providing long-term electrocorticography recording and cortical stimulation for epilepsy and brain-computer interface." *Journal of Neurosurgery* 130 (4): 1166-1179.
Saghai, Yashar. 2013. "Salvaging the concept of nudge." *Journal of medical ethics* 39 (8): 487-493.
Scarantino, Andrea , and Ronald de Sousa. 2018. "Emotion." *Stanford Encyclopedia of Philosophy*.
Schmidt, Andreas T. 2019. "Getting real on rationality—Behavioral science, nudging, and public policy." *Ethics* 129 (4): 511-543.
Thaler, R.H., and C.R. Sunstein. 2009. *Nudge: Improving Decisions about Health, Wealth and Happiness*. Penguin Books.
Wallace, Susan E, Elli G Gourna, Graeme Laurie, Osama Shoush, and Jessica Wright. 2016. "Respecting autonomy over time: Policy and empirical evidence on re-consent in longitudinal biomedical research." *Bioethics* 30 (3): 210-217.
Wilkinson, T Martin. 2013. "Nudging and manipulation." *Political Studies* 61 (2): 341-355.