Published in final edited form as:
Neuroethics. 2021 October ; 14(Suppl 1): 83–93. doi:10.1007/s12152-019-09417-4.

DBS and Autonomy: Clarifying the Role of Theoretical Neuroethics

Peter Zuk,
Baylor College of Medicine and Rice University, Center for Medical Ethics and Health Policy, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030

Baylor College of Medicine and Rice University, Department of Philosophy, Rice University, 6100 Main St, Houston, TX 77005

Gabriel Lázaro-Muñoz
Baylor College of Medicine, Center for Medical Ethics and Health Policy, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030

Abstract

Gilbert, Viaña, and Ineichen call for further empirical work on the effects of deep brain stimulation (DBS) on personality, identity, agency, authenticity, autonomy and self (PIAAAS) (Gilbert et al. 2018a). In particular, they emphasize the need for more sophisticated instruments measuring potential changes in PIAAAS. The development of such instruments, they argue, will provide a stronger empirical foundation for theoretical neuroethics work on DBS. We agree with this proposal. However, we believe that theoretical neuroethics has an important role to play in advancing empirical neuroethics that is not emphasized in Gilbert et al.'s remarks on the relationship between empirical and theoretical neuroethics. The development of instruments for more fully assessing changes in PIAAAS will require significant clarification of its component concepts. This task of clarification is the purview of theoretical neuroethics. In this article, we sketch how theoretical neuroethics can clarify the concept of autonomy. We hope that this can both serve as a model for the conceptual clarification of other components of PIAAAS and contribute to the development of the empirical measures that Gilbert and colleagues propose.

Keywords

Deep brain stimulation; autonomy; agency; theoretical neuroethics; empirical neuroethics

---

glazaro@bcm.edu, **Phone:** 713-798-2672.

CONFLICT OF INTEREST
The authors declare that they have no conflict of interest.

INTRODUCTION

Gilbert, Viaña, and Ineichen call for further empirical work on the effects of deep brain stimulation (DBS) on personality, identity, agency, authenticity, autonomy, and self (PIAAAS) (Gilbert et al. 2018a). In particular, they emphasize the need for more sophisticated instruments measuring potential changes in PIAAAS. The development of such instruments, they argue, will provide a stronger empirical foundation for theoretical neuroethics work on DBS. We agree with this proposal. However, we believe that theoretical neuroethics has an important role to play in advancing empirical neuroethics that is not emphasized in Gilbert et al.'s piece. The development of instruments for more fully assessing changes in PIAAAS will require significant clarification of its component concepts, and this task of clarification is the purview of theoretical neuroethics. Here, we sketch how theoretical neuroethics can clarify the concept of autonomy (relating, it at times, to the additional PIAAAS concept of *agency*). We hope that this will both serve as a model for the conceptual clarification of other components of PIAAAS in neuroethics and contribute to the development of the empirical measures that Gilbert and colleagues propose.

When neuroethicists suggest that DBS may impact autonomous agency, they are often appealing to quite different conceptions of autonomy. We identify three such conceptions in the neuroethics literature: a *traditional* conception, an *experiential* conception, and a *relational* conception. The first of these conceptions informs clinical understandings of autonomy used to assess an individual's capacity for informed consent to medical interventions. The remaining two offer understandings of autonomy in a broader, non-clinical sense. Because these three conceptions propose overlapping but distinct criteria for autonomy, they differ in their accounts of what is required to undermine or promote it. This can make it unclear what is meant, and so what exactly is ethically at stake, when DBS is claimed to impact autonomy.

We do not attempt to adjudicate which of these different conceptions, if any, best captures the nature of autonomy. DBS subjects themselves may well (rightly or wrongly) differ from one another in how they weight these criteria, complicating the prospects for practical implementation of a "one-size-fits-all" understanding of their autonomy. Our goal here is instead to make the different conceptions clear to set the stage for development of empirical measures that more fully capture the various phenomena that each conception claims to constitute autonomy. This may in turn help to clarify the *different senses* in which DBS might be said to impact autonomy, and so help to make ethical debates more tractable. In order to accomplish this, we describe how these conceptions apply to health conditions that can be broadly understood as *disorders of volition.* These rather extreme examples help to clarify from a theoretical standpoint the ways in which DBS may affect autonomy and how different conceptions of autonomy analyze these possible effects. This in turn allows for a more refined understanding of what to look for in empirical assessment of the effects of DBS on autonomy.

UNDERMINING AUTONOMY

As Gilbert et al. (2018a) note, *claims* that DBS may undermine autonomy are commonplace in the neuroethics literature. Commenting on a case study from Herzog et al. (2003) in which a patient undergoing DBS for Parkinson's experienced mania, disinhibition, and psychosis, Klaming and Haselager (2013, 534) claim that "although DBS can extend a patient's freedom considerably—enabling him to live a life independent from constant care—it may at the same time affect the patient's decisional capacity and hence his autonomy" by undermining basic mental and/or volitional competencies. Discussing work suggesting that pathological gambling is a possible side effect of DBS, Baylis (2013, 524) writes, "When direct brain manipulation explains a belief or behavior there is reason to think of this as a serious threat to agency." And Goering et al. (2017, 65), commenting on adaptive or closed-loop DBS (aDBS) in particular, write that "the agent using the device may… doubt whether or not she is the author of her action, given that the device may operate in ways that are not transparent to her."

More recent cases are also invoked in support of these kinds of claims. Gilbert et al. (2017, 99) provide several: a patient's compulsive attempt to reach her husband on foot, a journey that would have taken multiple days; another patient's attempt to move a pool table, narrowly avoiding permanent disability; and a third patient's self-described "emotional incontinence." Goering et al. (2017, 63) report a case at the other end of the spectrum: a patient experiencing a *lack* of appropriate emotion at a funeral, against the patient's own preferences. Older examples include a patient allegedly experiencing themselves as though under the sway of "remote control" (Schüpbach et al. 2006 in Goering et al. 2017, 63) and a patient who said, "I feel like an electric doll" quoted in Schüpbach et al. (2006) and Agid et al. (2006)[1].

As Gilbert et al. (2018a, 8) point out, however, these kinds of cases are few and far between in the DBS literature: "Out of 64 first-hand primary studies, 43 did not corroborate evidence that DBS leads to PIAAAS alteration. Out of the remaining 21 articles, 13 were marginal or single reports." This number is dwarfed by the overall number of patients who have been implanted with DBS, which stands at 150,000 for Medtronic devices alone, so the total is presumably a good deal higher (Gilbert et al. 2018a, 9). But because it is likely that DBS researchers and clinicians are not systematically examining or publishing on PIAAAS-related side effects, more extensive empirical investigation of the kind Gilbert et al. (2018a) call for will be required for the field to come to a confident judgment about the frequency of this issue.

PROMOTING AUTONOMY

On the other hand, some commentators point out that DBS may also promote autonomy. de Haan et al. (2015) describe a number of effects of DBS for obsessive-compulsive disorder (OCD) that might be interpreted as autonomy-promoting: greater self-confidence, self-reliance, and self-trust, and feelings of empowerment taking the form of a greater

---

[1] In Agid et al. (2006), however, the quote is apparently mistranslated as "I am an electric doll" (Gilbert et al. 2018a, 4).

degree of short- and long-term planning (8). Vukov (2017, 90) suggests that aDBS may promote autonomy understood as "the ability to set and achieve one's ends." And others have emphasized that because the disorders for which aDBS is used as an intervention often *themselves* undermine autonomy, aDBS may promote autonomy in serving to "correct or prevent neural dysfunction" associated with these disorders (Lázaro-Muñoz et al. 2017, 74).

This last point raises the issue of how to think about a suitable *baseline* against which to compare changes in patient autonomy in aDBS. The appropriate baseline level of autonomy may not be that of the average neurotypical adult, but instead the patient's own pre-aDBS state (Lázaro-Muñoz et al. 2017, 74). Arguably, we should be most concerned about whether DBS and aDBS promote or interfere with autonomy relative to this standard. Focusing on intrapersonal rather than interpersonal comparisons in this way has important implications for the future development and interpretation of empirical measures of autonomy in DBS and aDBS, as it may well affect whether an outcome is properly counted as a benefit or detriment vis-á-vis patient autonomy. In order to set the stage for such work, we now turn our attention to the three conceptions of autonomy most commonly invoked in the neuroethics literature.

THREE CONCEPTIONS OF AUTONOMY

A range of positions have been articulated regarding how great an interference (and how great a promise) DBS may involve for patient autonomy. Views about how DBS may affect autonomy vary not only in terms of their direction (promoting or undermining) and the magnitude of the possible effect, but also regarding *how* autonomy is affected. What explains the dramatically different ways of talking about autonomy that one encounters in the neuroethics literature on DBS? The answer, it seems, is that the parties to the debate are deploying different *conceptions* of autonomy—different understandings of its core features, or at the very least, different degrees of emphasis on, or privileging of, certain potential features over others. We now describe three such conceptions that help to explain how autonomy is deployed in the neuroethics literature and discuss their diverging implications.

The Traditional Conception

On what we term a *traditional* conception of autonomy, it consists fundamentally in the capacity to act (1) intentionally, (2) with understanding, and (3) without internal or external "controlling influences" (Beauchamp and Childress 2013). This is a somewhat minimal conception of autonomy, stating basic requirements that must be met for a person to give their consent. As such, this is the conception of autonomy regularly employed in medical research and clinical contexts. Foundational accounts of this conception in the philosophy literature (Frankfurt 1971, Dworkin 1976, Dworkin 1988, DeGrazia 2005) may be more demanding in their accounts of intention, understanding, and controlling influences, but they share with Beauchamp and Childress an understanding of autonomy rooted in an individual's attitudes—in particular, her desires or preferences (see Gibert 2017 for a discussion of these views and their potential relationship to DBS). Before describing how these conditions apply in DBS interventions, we situate them in the context of a rather extreme case of what can broadly be classified as a *disorder of volition.* This demonstrate vividly how a person can meet some of these conditions while failing to meet all of them.

Consider a case involving bilateral frontal lobe ablation reported by Eslinger and Damasio (1985). The individual in question experienced "dissociation between intact cognitive abilities measured by standardized tests, and the poor utilization of those abilities in the real environment" (Eslinger and Damasio 1985, 1737). In conversation, the individual was able to articulate what reasonable courses of action would be in a variety of hypothetical practical and social contexts. However, he was unable to put this understanding of appropriate behavior into action in his own life. He took inordinate amounts of time to make even simple decisions, e.g., sometimes deliberating for hours without resolution about which restaurant to dine at. This had devastating consequences for his work and home life.

As Eslinger and Damasio interpret this individual's situation, "He had learned and used normal patterns of social behavior before his brain lesion, and although he could recall such patterns when he was questioned about their applicability, *real-life situations failed to evoke them*" (Eslinger and Damasio 1985, 1737). This individual seems clearly to meet condition (2) of the traditional conception of autonomy, demonstrating clear *understanding* of options and how to realize them. But he does not meet condition (1): he is unable to reliably put his knowledge into practice through effective *intentional action.*

Cabrera et al. (2014, 39) articulate some of the possible worries that someone holding the traditional conception of autonomy might have about DBS. Potential patients or study participants considering DBS may believe that they *must* consent to DBS due to a lack of other options and may also face pressure to undergo DBS as its use becomes more widespread. If sufficiently strong, these pressures might rise to the level of external controlling influences. Potential patients or study participants will often have a psychiatric condition or some form of cognitive impairment, potentially affecting their ability to choose DBS with sufficient understanding. Finally, the effects of DBS themselves may change an individual's ability to consent and withdraw consent by changing the individual's mental state.

The Experiential Conception

While it remains highly influential, the traditional conception is not the only one invoked in the neuroethics literature. A comparatively more demanding *experiential* conception of autonomy also appears in the neuroethics literature. This is a conception of autonomy in a broader philosophical sense that goes beyond the capacities required for informed consent, reflecting a trend in the field of neuroethics to supplement understandings of clinical efficacy with non-clinical outcomes. What unifies different versions of this conception of autonomy is their insistence on the idea that conscious experience plays a crucial role in autonomous agency.

Recent empirically-informed work on the nature of human agency has reflected this approach. It is evident in Pacherie (2007, 2008)'s work, which makes an empirical case for a relationship between the experience of agency and agency itself, focusing in particular on "the experience of intentional causation, the sense of initiation and the sense of control"

(Pacherie 2007, 2). Bayne (2008) describes different varieties of *agentive experience,* stressing the experience of deliberation and action and relating this explicitly to autonomy. In particular, he suggests that certain forms of agentive experience may be essential to certain forms of agency—in particular, "complex" or "more autonomous" forms (Bayne 2008, 197). More recently, Kriegel (2015) has defended the irreducibility of some forms of agentive experience to other forms of experience.

An influential employment of this approach in normative ethics is Korsgaard (2009)'s. While Korsgaard may be most fundamentally concerned with the question of what it is to be an *agent,* the Kantian framework she employs affirms this question as inseparable from the idea of autonomy.[2] Korsgaard (2009, 19–20), describes it well when she writes that:

> When you deliberately decide what sorts of effects you will bring about in the world [by acting], you are also deliberately deciding what sort of a cause you will be. And that means you are deciding who you are. So we are each faced with the task of constructing a peculiar, individual kind of identity—personal or practical identity… It is as the possessor of personal or practical identity that you are the author of your actions, and responsible for them.

Korsgaard attempts to vindicate the idea that there is an important sense in which we constitute—make or create—ourselves. In doing so, she is not committing to an obscure metaphysical view, but rather (in decidedly Kantian fashion) drawing attention to the role of a particular kind of *experience* of deliberation and subsequent action in autonomy. Elsewhere, she writes, "When you deliberate, it is **as if** there were something over and above all of your desires, something which is *you,* and which *chooses* which desire to act on" (Korsgaard 1996, 100, our bolding).

The "as if" clause of this claim is important, marking this Kantian view as most fundamentally concerning the nature of *practical reasoning* (the way in which one must understand oneself and structure one's deliberations in order to act rationally, and thereby autonomously) rather than the *metaphysics* of agency (which involves, e.g., questions about the extent to which our actions are causally determined, questions about whether there is in fact an agential *you* over and above your various motives on the basis of which you choose between them, and so on). The key insight of the Kantian view is that autonomous action requires that one *experience* the future as though causally open and *experience* oneself as though deciding which of one's various motives to act upon.

This process of reflectively endorsing and acting out your own motives, in a distinctively experiential way, yields a distinctive conception of autonomy that privileges this form of experience. On this experiential conception, we shift our theoretical attention from the phenomena central to the traditional conception (intention, understanding, and absence of controlling influences) to the experience of deliberation that *generates* informed intentions and then *guides* the intended action to completion. The experiential conception goes beyond the traditional conception by locating paradigmatic or ideal exercises of autonomy in the

---

[2] On Korsgaard's view, autonomy is one of two "essential characteristics of an agent," with the other being efficacy (the degree to which one is successful in realizing the effects one intended to cause in action) (Korsgaard 2009, 82). We thank an anonymous reviewer for raising this issue.

reflective, experientially self-aware form of deliberation in which human beings sometimes engage, together with its realization in action.

In order to see this more clearly, consider a disorder of volition that interferes with this kind of deliberation: *anosognosia,* a lack of recognition that one is affected by neurological and/or neuropsychological symptoms (Prigatano 2009). One particularly striking instance of anosognosia is Anton's syndrome, in which a person with cortical blindness (vision loss due to occipital cortex damage) insists, even against strong evidence, that she can see. This situation greatly complicates an individual's prospects for rehabilitation or adaptation. As Prigatano (2009, 609) writes, "In neurorehabilitation, it is a common practice to have patients identify treatment goals and to develop a treatment program around those goals. If the patients' perceptions of themselves are unrealistic, then their goals will be unrealistic."

Such a situation inhibits autonomy on the traditional conception by interfering with understanding, to be sure, but is even *more* inhibiting on the experiential conception. Often, a medical situation calling for lengthy and uncertain rehabilitation might occasion self-reflection on one's values and their relative priority, as well as whether to reevaluate them in light one's changed circumstances. Anosognosia impairs this self-reflection by precluding an individual from fully understanding her circumstances. In doing so, it impairs the kind of self-creation (or perhaps *self-restoration*) in which neurorehabilitation would otherwise allow her to engage. On the experiential conception, this is an especially great blow to an individual's autonomy.

Unlike the traditional conception, the experiential conception is seldom appealed to *explicitly* in the neuroethics literature. However, its emphasis on the experience of deliberation and action is evident in the work of several commentators on the neuroethics of DBS, suggesting an implicit adherence to this conception. For this reason, making this role of this conception in the neuroethics literature more explicit has the potential to drive important debates forward. de Haan et al. (2013) invoke individuals' "sense of agency," suggesting changes to this experience as a potential explanation for symptom improvement reported in a cohort of patients receiving DBS for OCD. Indeed, de Haan et al. (2013, 2015) elucidate the effects of DBS on symptoms of OCD using Gibson (1986 [1979])'s concept of *affordances:* possibilities for action in and interaction with one's environment. de Haan et al. propose that symptom relief following DBS for OCD can be explained in terms of patients' experience of affordances in the physical and social environment. In their view, OCD symptoms such as compulsions manifest in some affordances being impressed so strongly in experience that other affordances are crowded out of one's perceptual attention. DBS, they argue, can be interpreted as mitigating this phenomenon, restoring an individual's experience of affordances to more closely resemble those of neurotypical subjects.[3] Kellmeyer et al. (2016, 624), writing in a neuroethics context, explains autonomy in terms of experience: "personal autonomy arises from the subject's experience of congruence of motive and action, which gives rise to the feeling of individual agency." Goering et al. (2017, 64) write, "One rationale for closed-loop technology is that it

---

[3] de Haan et. al.'s emphasis on the significance affordances in this context is in line with recent experimental work in other neuropsychiatric contexts. See, for example, Thill et. al. (2013) and McBride et al. (2013).

could enhance a sense of agency," where the sense of agency is understood in terms of one's experience of interacting with other people and the environment.[4] And a defining feature of the influential work of Gilbert (2015), Gilbert and Cook (2015), Gilbert et al. (2017), and Gilbert et al. (2018b) is careful attention to the experience of autonomous agency, which they take to have significant consequences for autonomous agency itself.

Relational Autonomy

The third major conception of autonomy reflected in the neuroethics literature is the *relational* conception. The relational conception claims that an individual's autonomy depends importantly on the relations in which she stands to other individuals, and sometimes also to broader classes of social factors. On this view, other people do not merely contribute to, but are in fact part of what *constitutes* (that is, makes up), a person's autonomy (Christman 2004, Westlund 2009). While a number of relational views of autonomy have been developed, the various developments of the view have in common a *distributed* conception that extends autonomous agency beyond the bounds of the traditional Western conception of the self.[5]

The case of *locked-in syndrome* (LIS) can help us to better understand this conception. In *classic* LIS, an individual lacks most voluntary bodily movement while retaining some cognitive function (Vidal 2018, 2). Such individuals can often communicate in technologically mediated ways, such as with word boards that respond to eye movements. In *complete* locked-in syndrome (CLIS), however, even the small movements that enable communication in classic LIS are absent. For those experiencing CLIS, advances in brain imaging (Monti et al. 2010, Fernández-Espejo and Owen 2013) and brain-computer interface (Naci et al. 2012, Schnakers et al. 2015) technologies may eventually allow for reliable communication based on the detection of neural activity. This could potentially allow those experiencing CLIS to express medical decisions, and so to exercise their agency.

In such a case, the social environment would provide the very *possibility* of the patient exercising her autonomy.[6] But despite the highly mediated nature of her actions, it would fundamentally be the *patient* who was autonomously expressing her wishes, and the *patient* whose decisions were being enacted. This helps us to understand the relational or distributed conception more clearly by providing a clear instance in which one individual's autonomy depends on the agency of others in order to be realized. By making this especially vivid by considering examples like CLIS, we can start to see subtler ways in which one individual's autonomy depends crucially on that of others.

Consider my paying taxes this year. We think of this as a clear-cut case of autonomous action, one which performing (or failing to perform) is attributable to *me* as an agent. But

---

[4] They cite Goering (2015), Gilbert and Cook (2015), and Glannon and Ineichen (2016) as supportive of this rationale.

[5] Of particular significance is the distinction between substantive and procedural relational views of autonomy (Stoljar 2018, §3). Substantive views place normative constraints on an agent's point of view, requiring that she have (or not have) certain attitudes in order to count as autonomous. Procedural views, by contrast, require only that an agent's attitudes were (or could have been) arrived at through an adequate process of reflection. We do not attempt to decided between these types of views, as our aim is not to adjudicate between different specifications of the conceptions that we have identified. We thank an anonymous reviewer for raising this point.

[6] …Or at least, of her exercising it beyond the confines of her own mind. Some views of the mind conceptualize states of consciousness as mental acts (Brentano 2012 [1874]).

now consider all of the other acts and agents required for the state of affairs consisting in *my paying taxes this year* to actually come to pass (e.g., my employer must have supplied me with accurate tax documents, the folks at TurboTax need to have designed and maintained a reliable tax calculation software, and the IRS must receive and process my return correctly). One might respond to an example like this by giving up on the idea that the act really is attributable to me after all. But on the relational or distributed view, we should say instead that the action *is* attributable to me. We do so by expanding our conception of what counts as *my* agency to include the relevant actions of others in furtherance of my aim. This is the essence of the relational conception.

In the aDBS literature, Goering et al (2017, 67) suggest that this conception of autonomy may allow us to think of aDBS devices as playing a role in patients' autonomy analogous in important respects to that of other people. aDBS might in this way support autonomy rather than threaten it. Some of their interlocutors (Morar and Skorburg 2017) in turn point out potential drawbacks of the relational conception. In their view, "any relational conception of agency or autonomy creates as much of a possibility for empowerment as disempowerment" due to implying that individuals are significantly physically, psychologically, and socially vulnerable to others (Morar and Skorburg 2017, 84). There is an active and spirited debate in the literature about the role of the relational conception in the neuroethics of DBS.

HOW TO ADDRESS THIS PLURALITY

These examples help illuminate the different conceptions of autonomy in the neuroethics literature and how they apply in practical contexts. We do not know the frequency of changes in autonomy as a result of DBS, but reflection on these cases may at least give us some idea side effects do not seem typically to involve the extreme diminution of autonomy involved in cases like the ones that we have described. Nonetheless, potential DBS-related effects on autonomy may have significant impact on the lives of patients and their families. It is therefore vital to develop more sophisticated ways of measuring how DBS or other neurotechnologies impact autonomy. This will require the further conceptual work to which we now turn.

Having described the three major conceptions of autonomy reflected in the literature on DBS, we are now in a position to state our central question: *How should neuroethicists address this plurality of conceptions in order to make progress*?

In our view, each of the three conceptions offers important insights about the nature of autonomy in the context of DBS. It therefore will not do to simply adopt one of the conceptions and jettison the others. We need some method of synthesizing the virtues of each conception in a way that allows us to draw upon the insights of each for empirical and theoretical work. Below, we describe three potential answers to the question of how to make progress of this kind. We then describe how the final one, which we endorse, can help drive further empirical research on DBS's possible effects on autonomy.

Reductionism

According to this proposal, we should recast allegedly "metaphysical" debates about autonomy in other terms (Müller et al. 2017). We might, for example, take each conception of autonomy to express a different conception of the "good life." This is a strategy that Levy (2011) appears at one point to endorse (and attributes to Parens 2005) on the related question of *authenticity* in the context of neurointervention. Discussing two rival views of authenticity, Levy writes that they "are better seen as outlooks on human life; conceptions of how we best live" (Levy 2011, 312). One might say something similar about autonomy in order to make the debate over DBS and autonomy more tractable.

In our view, this kind of proposal would be too minimalist in the present context. In attempting to make debates about autonomy more tractable by explaining it in terms of something else, it threatens to explain autonomy *away*. Autonomy appears to be too central a value in bio- and neuroethics to reduce in this way. We need to give autonomy its due, and that requires a more nuanced approach.

Pluralism

One such approach takes the form of *pluralism* about autonomy. On this approach, autonomy is *multi-faceted,* and each of the major conceptions of autonomy describes one of these irreducibly different facets. On this approach, the full truth about autonomy is therefore expressed by taking key elements of each conception *in combination.* Mackenzie (2014, 15) offers a view of this kind, identifying "three distinct, but causally interdependent, dimensions of autonomy." Mackenzie carves up the conceptual terrain differently than we have here, doing so in terms of *self-determination, self-governance,* and *self-authorization.* But these three categories do collectively capture the core elements of the conceptions that we have identified.[7]

Pluralist analyses of ethically-relevant concepts have recently appeared in the neuroethics literature. Gallagher (2013, 2018) and (Dings and De Bruin 2016) defend a pluralist view of personal identity. But this view has been subject to the objection "that it is just a list; a heaping of aspects, without an account of how they relate" (de Haan et al 2017, 5). In one important sense, Mackenzie's view is clearly not "just a list" because it posits *causal interdependence* between its elements. But at the conceptual level, it is by design a *multidimensional* conception as opposed to *unitary* one (Mackenzie 2014, 15). This means that it contains irreducibly different elements, giving rise to the questions about how to weight the various elements in an overall assessment of an individual's degree of autonomy, and about whether such an overall assessment is even possible. One might call this a *calculation problem* for pluralist views. But this problem, we think, is not insurmountable. We now introduce a nearby view that preserves the insights of pluralism while offering the possibility of progress through more detailed empirical work.

---

[7] Self-determination pertains to "external, structural conditions" involving freedom and opportunity (Mackenzie 2014, 17), and so captures important aspects of the relational conception of autonomy. Self-governance pertains to the ability decide and act in accordance with a Korsgaardian practical identity or self-conception (17–18), and so encompasses important aspects of both the traditional and experiential conceptions of autonomy. Self-authorization pertains to taking oneself to possess the "normative authority to be self-determining and self-governing" (18) in way related to, among other things, practices of accountability and answerability (19), and so also involves the experiential and relational conceptions of autonomy.

Expansionism

The three conceptions are incompatible as views of what autonomy *fundamentally is.* But they are—at least in the basic form in which we have presented them—compatible as views of what desirable exercises of human autonomy *characteristically involve.* As pluralism claims, each has an important insight to offer. But our proposed expansionism views the three conceptions as more deeply interrelated than pluralism does—not only causally, but conceptually as well. While none is straightforwardly reducible to the others, some are essential preconditions for or contributors to others.

For instance, the experiential conception plausibly *includes* the phenomena picked out by the traditional conception. On the experiential conception, action that is i) intentional, ii) involves understanding, and is iii) free of controlling influences is *necessary but not sufficient* for paradigmatic exercises of autonomy (as this conception understands it). This basic ability to act is required for the *realization* of the experiential conception's distinctively experiential form of reflection and choice. So we can speak of the experiential conception *subsuming* the more basic capacities picked out by the traditional conception.

The relational conception in turn plausibly subsumes the phenomena picked out by the experiential conception, taking them to be necessary but not sufficient for paradigmatic exercises of autonomy. This is because, at least for most prominent developments of the relational conception, an individual's self-understanding is a key ingredient in her relational *identity,* which in turn undergirds her capacities to act autonomously.

Meanwhile, the traditional view does not take the phenomena picked out by the other two conceptions to themselves count as *part* of autonomy. But it can, and should, take them to be plausible *causal contributors* to the exercise of autonomy. Proponents of the traditional conception are unlikely to dispute that we undergo the distinctive experience of deliberation and choice described by the experiential conception. Instead, they may deny that these experiences themselves count as part of autonomy. But that is compatible with these things *contributing* to greater autonomy as the traditional conception understands it, playing a causal role in firmer intentions, greater understanding, and a higher degree of freedom from controlling influences. And this implies that the *lack* of these things can *diminish* autonomy as the traditional conception understands it.

Similarly, the experiential conception does not take the phenomena picked out by the relational conception to themselves count as part of autonomy. But it can, and should, take them to be plausible causal contributors to the exercise of autonomy. That is because of the role of other people and broader social phenomena in shaping a self-conception that promotes, or hinders, the experience of deliberation and choice that characterize the experiential conception (see Raz 1994, 177). And if that is so, the phenomena picked out by the relational conception also causally contribute, transitively, to autonomy as understood by the traditional conception.

We see now that Mackenzie's thesis of the causal interdependence of different aspects of autonomy turns out to be a deep insight indeed. The conceptions can be visualized as a set of concentric circles. Starting from the traditional conception and moving outward, we have more basic capacities that serve as necessary conditions for more complex ones. Starting from the relational conception and moving inward, we have capacities that do not themselves count as part of autonomy but causally contribute to its promotion or diminution.

Admittedly, this tidy description of the relationships between the different conceptions is an oversimplification. There are a number of more particular ways of developing the different conceptions as detailed sets of conditions for autonomy, and that complexity is not reflected here. Just as our aim is not to settle debates about the comparative merits of the three overall conceptions, it is also not our aim to adjudicate debates about how best to specify each conception in detail. We aim instead to shed greater light on how the three conceptions relate at a higher level of abstraction. This allows for potential advances in empirical measures of autonomy, which will in turn help to clarify such theoretical issues.

IMPLICATIONS FOR MEASURING AUTONOMY

Proponents of the various conceptions disagree on the fundamental *constituents* of autonomy. But that need not prevent them from agreeing on empirical *indicators* autonomy. If the constituents put forward by each conception are causally interdependent, then we can expect overlap in their indicators. Indeed, what one conception counts as constituents will be indicators relative to another. The chart below summarizes what proponents of one conception can most plausibly say about the phenomena picked out by another conception, given our analysis:

These relations between the conceptions have significant implications for empirical neuroethics. They make it possible to construct measures of autonomy that have a distinctively ecumenical character. We envision a measure that would incorporate the different conceptions as domains each with their own subscale(s). An overall score would be a function of the different subscales, but the function applied could be variable depending on one's particular view of autonomy. Those with highly specified accounts of autonomy might wish to assign particular weights to favored factors. Those who favor a more integrative approach would have a measure that combines the insights of each of the different conceptions.

Perhaps even more importantly, an ecumenical measure like this could be useful in clinical and research settings. Patients and clinicians might discuss those items for which there is a change from the perspective of the patient pre- and post-DBS surgery, and then consider whether the patient approves or disapproves of this change, and how strongly. If the patient has trouble with a change on a particular item, the clinical team and patient could work together to identify potential ways of managing perceived negative impacts regarding that element of autonomy.

In principle, such a measure could even help to make conceptual debates about autonomy more tractable. Because conceptions of autonomy are often different philosophical interpretations of the same empirical facts, discovering more empirical facts about individuals' self-perceptions of autonomy is likely to help shape these conceptions. Gilbert et al. (2018a) characterize theoretical neuroethics in the absence of rigorous empirical data as speculative. This is surely an accurate characterization when theoretical neuroethics arguments are made in support of claims that require empirical data, and Gilbert et al. have done a great service in drawing attention to such a potential trend. But it must also be emphasized that the deployment of theoretical neuroethics in service of conceptual clarification is not a speculative exercise. In fact, perhaps one thing that has contributed to persistent concerns about the impact of DBS on autonomy without sufficient empirical data to support this, is that there has sometimes been insufficient conceptual clarification of PIAAAS in the neuroethics literatures. This leads to difficulty interpreting what each author's concerns are and developing coherent empirical measures to examine the actual frequency, magnitude, and type of impact that DBS has on PIAAAS. Good conceptual clarification, of course, will itself sometimes draw on empirical knowledge in order to show how the concepts at issue apply to the world. In our case, we have drawn upon findings about human volition in general, outside the context of DBS in hopes of showing how they might apply here. The relationship between empirical and theoretical neuroethics, we think, must therefore be characterized as a deeply reciprocal one. It is in bringing them together that we arrive at the fullest possible picture of the implications of emerging neurotechnologies like DBS on PIAAAS concepts and other matters of ethical significance.

ACKNOWLEDGEMENTS

Research for this article was funded by a BRAIN Initiative grant from the U.S. National Institute of Mental Health of the National Institutes of Health, Award Number R01MH114854 (to G.L.M.). The views expressed are those of the authors alone and do not necessarily reflect views of the NIH, Baylor College of Medicine, or Rice University.

References

Agid Y, Schüpbach M, Gargiulo M, Mallet L, Houeto JL, Behar C, … Welter ML (2006). Neurosurgery in Parkinson's disease: The doctor is happy, the patient less so? Journal of Neural Transmission, 70 [Suppl], 409–414. 10.1007/978-3-211-452950_61.

Baylis F (2013). "I am who I am": On the perceived threats to personal identity from deep brain stimulation. Neuroethics, 6(3), 513–526. 10.1007/s12152-011-9137-1. [PubMed: 24273621]

Bayne T (2008). The Phenomenology of Agency. Philosophy Compass 3(1), 182–202. 10.1111/j.1747-9991.2007.00122.x.

Brentano F (2012 [1874]). Psychology from an Empirical Standpoint. Edited by Kraus O, McAlister LL Translated by Rancurello AC, Terrell DB, & McAlister LL London and New York: Routledge.

Cabrera LY, Evans EL, & Hamilton RH (2014). Ethics of the electrified mind: Defining issues and perspectives on the principled use of brain stimulation in medical research and clinical care. Brain Topography, 27(1), 33–45. 10.1007/s10548-013-0296-8. [PubMed: 23733209]

Christman J (2004). Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves. Philosophical Studies 117(½), 143–164. https://doi.org/www.jstor.org/stable/4321441.

DeGrazia D (2005). Human Identity and Bioethics. New York: Cambridge University Press.

de Haan S, Rietveld E, Stokhof M, & Denys D (2017). Becoming more oneself? Changes in personality following DBS treatment for psychiatric disorders: Experiences of OCD patients and general considerations. PLoS ONE, 12(4), 1–27. 10.1371/journal.pone.0175748.

de Haan S, Rietveld E, Stokhof M, & Denys D (2015). Effects of deep brain stimulation on the lived experience of obsessive-compulsive disorder patients: In-depth interviews with 18 Patients. PLoS ONE, 10(8), 1–29. 10.1371/journal.pone.0135524.

de Haan S, Rietveld E, Stokhof M, & Denys D (2013). The phenomenology of deep brain stimulation-induced changes in OCD: an enactive affordance-based model. Frontiers in Human Neuroscience, 7, 1–14. 10.3389/fnhum.2013.00653. [PubMed: 23355817]

Dings R, & de Bruin L (2016). Situating the self: understanding the effects of deep brain stimulation. Phenomenology and the Cognitive Sciences, 15(2), 151–165. 10.1007/s11097-015-9421-3.

Dworkin G (1988). The Theory and Practice of Autonomy. Cambridge: Cambridge University Press.

Dworkin G (1976). Autonomy and behavior control. Hastings Center Report 6(1):23–28. 10.2307/3560358.

Eslinger PJ, & Damasio AR (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. Neurology, 35(12), 1731–1741. [PubMed: 4069365]

Fernández-Espejo D, & Owen AM (2013). Detecting awareness after severe brain injury. Nature Reviews Neuroscience, 14, 801–809. [PubMed: 24088810]

Frankfurt HG (1971). Freedom of the Will and the Concept of a Person. The Journal of Philosophy 68(1), 5–20. 10.2307/2024717.

Gallagher S (2018). Deep Brain Stimulation, Self and Relational Autonomy. Neuroethics, 1–13. 10.1073/pnas.0913016107.

Gallagher S (2013). A Pattern Theory of Self. Frontiers in Human Neuroscience, 7, 1–7. 10.3389/fnhum.2013.00443. [PubMed: 23355817]

Gibert SH (2017). Closed-Loop Deep Brain Stimulation and Its Compatibility With Autonomous Agency. AJOB Neuroscience, 8(2), 88–90. 10.1080/21507740.2017.1320326.

Gibson JJ (1986 [1979]). The Ecological Approach to Visual Perception. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gilbert F (2015). A Threat to Autonomy? The Intrusion of Predictive Brain Implants. AJOB Neuroscience, 6(4), 4–11. 10.1080/21507740.2015.1076087. [PubMed: 26740906]

Gilbert F, Viaña JNM, & Ineichen C (2018a). Deflating the "DBS causes personality changes" bubble. Neuroethics, 1–17. 10.1007/s12152-018-9373-8.

Gilbert F, O'Brien T, & Cook M (2018b). The Effects of Closed-Loop Brain Implants on Autonomy and Deliberation: What are the Risks of Being Kept in the Loop? Cambridge Quarterly of Healthcare Ethics, 27(2), 316–325. 10.1017/S0963180117000640. [PubMed: 29509128]

Gilbert F, Goddard E, Viaña JNM, Carter A, & Horne M (2017). I Miss Being Me: Phenomenological Effects of Deep Brain Stimulation. AJOB Neuroscience, 8(2), 96–109. 10.1080/21507740.2017.1320319.

Gilbert F, & Cook M (2015). Are predictive brain implants an indispensable feature of autonomy? Bioethica Forum 8(4), 121–27. http://www.bioethica-forum.ch/content/e_Archive.php.

Glannon W, & Ineichen C (2016). Philosophical aspects of closed-loop neuroscience. In Closed Loop Neuroscience (pp. 259–70). Edited by Hady A. El. New York: Academic Press.

Goering S, Klein E, Dougherty DD, & Widge AS (2017). Staying in the Loop: Relational Agency and Identity in Next-Generation DBS for Psychiatry. AJOB Neuroscience, 8(2), 59–70. 10.1080/21507740.2017.1320320.

Goering S (2015). Stimulating Autonomy: DBS and the Prospect of Choosing to Control Ourselves Through Stimulation. AJOB Neuroscience 6(4), 1–3. 10.1080/21507740.2015.1106274.

Herzog J, Reiff J, Krack P, Witt K, Schrader B, & Mu D (2003). Manic episode with psychotic symptoms induced by subthalamic nucleus stimulation in a patient with Parkinson's disease. Movement Disorders, 18(11), 1382–1384. [PubMed: 14639687]

Kellmeyer P, Cochrane T, Müller O, Mitchell C, Ball T, Fins JJ, & Biller-Andorno N (2016). The effects of closed-loop medical devices on the autonomy and accountability of persons and systems. Cambridge Quarterly of Healthcare Ethics, 25(4), 623–633. 10.1017/S0963180116000359. [PubMed: 27634714]

Klaming L, & Haselager P (2013). Did My Brain Implant Make Me Do It? Questions Raised by DBS Regarding Psychological Continuity, Responsibility for Action and Mental Competence. Neuroethics, 6(3), 527–539. 10.1007/s12152-010-9093-1. [PubMed: 24273622]

Korsgaard CM (2009). Self-Constitution: Agency, Identity, and Integrity. New York: Oxford University Press.

Korsgaard CM (1996). The Sources of Normativity. Edited by O'Neill O Cambridge: Cambridge University Press.

Kriegel U (2015). The Varieties of Consciousness. New York: Oxford University Press.

Lázaro-Muñoz G, McGuire AL, & Goodman WK (2017). Should We Be Concerned About Preserving Agency and Personal Identity in Patients With Adaptive Deep Brain Stimulation Systems? AJOB Neuroscience, 5(2), 73–75. 10.1080/21507740.2017.1320337.

Levy N (2011). Enhancing authenticity. Journal of Applied Philosophy, 28(3), 308–318. 10.1111/j.1468-5930.2011.00532.x.

Mackenzie C (2014). Three Dimensions of Autonomy: A Relational Analysis. In Veltman A & Piper M (Eds.), Autonomy, Oppression, and Gender (pp. 15–41). New York: Oxford University Press.

McBride J, Sumner P, Jackson SR, Bajaj N, & Husain M (2013). Exaggerated object affordance and absent automatic inhibition in alien hand syndrome. Cortex 49(8), 2040–2054. 10.1016/i.cortex.2013.01.004. [PubMed: 23433243]

Monti MM, Vanhaudenhuyse A, Coleman MR, Boly M, Pickard JD, Tshibanda L, … Laureys S (2016). Willful Modulation of Brain Activity in Disorders of Consciousness, 579–589. 10.1056/NEJMoa0905370.

Morar N, & Skorburg JA (2017). Relational Agency: Yes—But How Far? Vulnerability and the Moral Self. AJOB Neuroscience, 8(2), 83–85. 10.1080/21507740.2017.1326409.

Müller S, Bittlinger M, & Walter H (2017). Threats to Neurosurgical Patients Posed by the Personal Identity Debate. Neuroethics, 10(2), 299–310. 10.1007/s12152-017-9304-0.

Naci L, Monti MM, Cruse D, Kübler A, Sorger B, Goebel R, … Owen AM (2012). Brain-computer interfaces for communication with nonresponsive patients. Annals of Neurology, 72(3), 312–323. 10.1002/ana.23656. [PubMed: 23034907]

Pacherie E (2008). The phenomenology of action: A conceptual framework. Cognition 107(1), 179–217. 10.1016/j.cognition.2007.09.003. [PubMed: 17950720]

Pacherie E (2007). The Sense of Control and the Sense of Agency. Psyche 13(1), 1–30. https://jeannicod.ccsd.cnrs.fr/ijn_00352565.

Parens E (2005). Authenticity and Ambivalence: Toward Understanding the Enhancement Debate. The Hastings Center Report, 35(3), 34–41.

Prigatano GP (2009). Anosognosia: Clinical and ethical considerations. Current Opinion in Neurology, 22(6), 606–611. 10.1097/WCO.0b013e328332a1e7. [PubMed: 19809315]

Raz J (1994). Multiculturalism: A Liberal Perspective. In Ethics in the Public Domain: Essays in the Morality of Law and Politics (pp. 170–191). Oxford: Clarendon Press. 10.1093/acprof:oso/9780198260691.003.0008.

Schnakers C, Giacino JT, Løvstad M, Habbal D, Boly M, Di H, … Laureys S (2015). Preserved covert cognition in noncommunicative patients with severe brain injury? Neurorehabilitation and Neural Repair, 29(4), 308–317. 10.1177/1545968314547767. [PubMed: 25160566]

Schüpbach M, Gargiulo M, Welter ML, Mallet L, Behar C, Houeto JL, … Agid Y (2006). Neurosurgery in Parkinson disease: a distressed mind in a repaired body? Neurology, 66(12), 1811–1816. [PubMed: 16801642]

Stoljar N (2018). Feminist Perspectives on Autonomy. The Stanford Encyclopedia of Philosophy (Winter 2018 Edition), ed. Zalta EN https://plato.stanford.edu/archives/win2018/entries/feminism-autonomy/.

Thill S, Caligiore D, Borghi AM, Ziemke T, & Baldassarre G (2013). Theories and computational models of affordance and mirror systems: An integrative review. Neuroscience and Biobehavioral Reviews 37(3), 491–521. 10.1016/j.neubiorev.2013.01.012. [PubMed: 23333761]

Vidal F (2018). Phenomenology of the Locked-In Syndrome: an Overview and Some Suggestions. Neuroethics 10.1007/s12152-018-9388-1.

Vukov JM (2017). Three Kinds of Agency and Closed-Loop Neural Devices. AJOB Neuroscience, 8(2), 90–91. 10.1080/21507740.2017.1320324.

Westlund AC (2009). Rethinking Relational Autonomy. Hypatia, 24(4), 26–49. 10.1111/j.1527-2001.2009.01056.x.

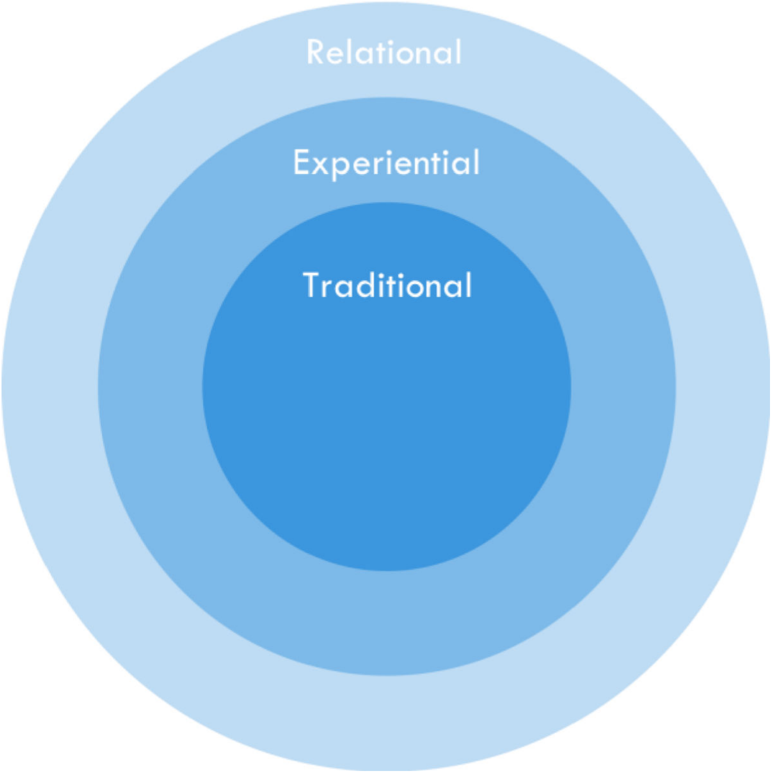
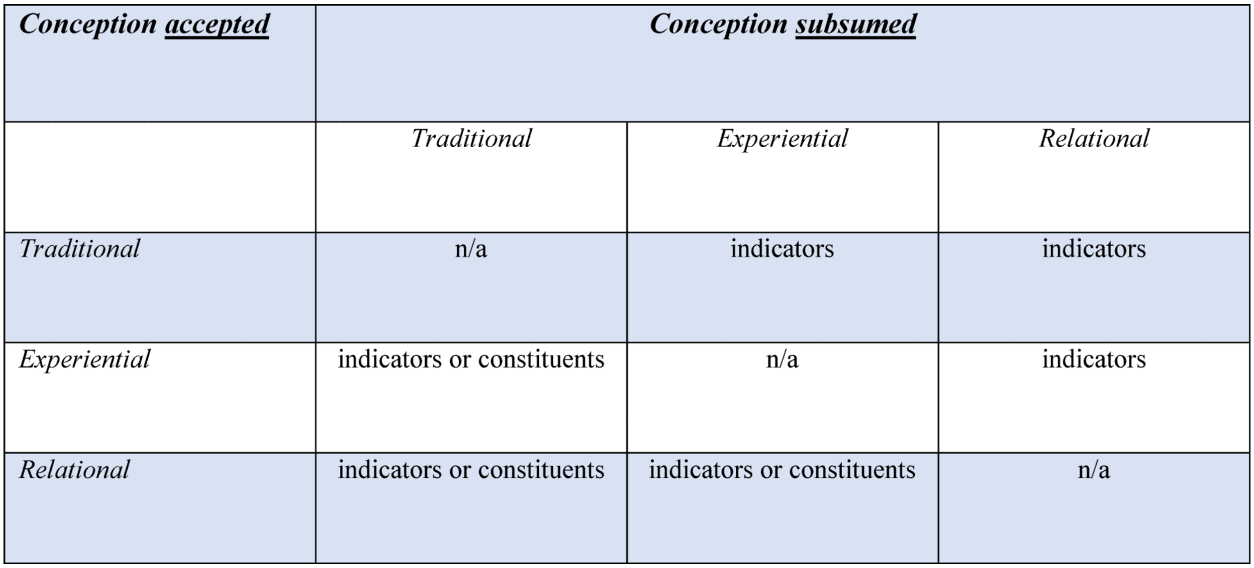


Figure 1.

Visualization of relationships between the phenomena picked out each conception of autonomy



Conception accepted	***Conception subsumed***		
	Traditional	*Experiential*	*Relational*
Traditional	n/a	indicators	indicators
Experiential	indicators or constituents	n/a	indicators
Relational	indicators or constituents	indicators or constituents	n/a

Figure 2.

Options for subsuming other conceptions within one’s favored conception. Depending on which conception one accepts, one will see the phenomena picked out by the other conceptions as related to autonomy in different ways. But proponents of each conception can allow that the phenomena of other conceptions are, at the very least, indicators of autonomy.