This is a submission version of a paper now published as

Walker MJ, Mackenzie C. 2020. Neurotechnologies, relational autonomy, and authenticity. *International Journal of Feminist Approaches to Bioethics* 13(1):98-119, <https://www.muse.jhu.edu/article/751378>

**Neurotechnologies, relational autonomy, and authenticity**

**Abstract**

The ethical debate about neurotechnologies has largely been framed around their effects on authenticity. In this paper we investigate the concept of authenticity and associated conceptions of the self. We develop a conception of authenticity that eschews problematic essentialist or existentialist views of the self, and the assumption that the authentic self transcends socialisation. In our view, authenticity is a condition for self-governance and can involve either endorsement or acknowledgement. Revisiting the debate about neurotechnologies, we show why framing the ethical debate in terms of authenticity is unhelpful and that these ethical concerns are better understood as concerns about autonomy.

**Keywords**: neurotechnologies, deep brain stimulation, authenticity, autonomy, self-change, self-governance

The ethical debate about neurotechnologies – including both drugs and implanted devices ­– has largely been framed around the questions of whether and when these technologies could damage or promote authenticity. Patients can experience changes in mood, behaviour, emotion, or preferences – seemingly, changes in character or personality. Some describe such changes by saying they feel like a ‘different person’; that they have become either more or less ‘themselves’; or that they feel as though some of their moods, behaviour, emotions or preferences ‘aren’t their own’. These kinds of statements are used to describe both changes that a drug or device is introduced to achieve (e.g., mood changes relating to anti-depressant treatments) and changes incidental to treatment (e.g., feeling more assertive as a result of deep brain stimulation to treat motor symptoms of Parkinson’s disease). In other words, they are used to describe changes that constitute both treatments and enhancements. Interestingly, such changes can be experienced as either positive or negative, and the evaluation does not necessarily match up with a sense of being more or less oneself, or with whether the clinical outcomes are primarily positive or negative.

This kind of language invokes the concept of authenticity.[[1]](#endnote-1) The concept of authenticity is understood in a variety of ways in the literature, as discussed below. However, loosely speaking, to be authentic is to be true to oneself, or to act on motives, desires, preferences and other reasons for action (henceforth ‘reasons’) that are ‘one’s own’.[[2]](#endnote-2) Authenticity thus provides a way of conceptualising psychological dysfunction or oppressive socialisation as problematic because they prevent us from being our real selves (Guignon 2008, 278; 2004, 4; Friedman 1986; Meyers 1989; Christman 1991). If neurotechnologies affect authenticity, this could be a reason for ethical concern, given that authenticity is something many people value.

Recently several authors have questioned the usefulness of the concept of authenticity for understanding the ethical issues at stake in neurotechnological interventions (e.g., Mackenzie and Walker 2015; Goddard 2017).[[3]](#endnote-3) These authors argue that the concept of authenticity as it is used in the ethical debate relies on an implausible view of the self and blurs the distinction between identity and autonomy. They suggest that autonomy is actually the relevant ethical concept. We agree with these claims. Our aim here, however, is not to revisit these arguments but to examine several questions they raise. Is there a way of conceptualising authenticity that is consistent with more plausible theories of the self? Since most theories of autonomy include authenticity among the conditions for autonomy, can the autonomy literature, and especially the literature on relational autonomy, help to clarify the relationship between autonomy and authenticity? And, might reconceptualisation of authenticity help us sort out ethical intuitions about neurotechnologies?

To answer these questions, we begin in section 1 with an overview of the ethical concerns relating to authenticity that are raised about neurotechnologies. In section 2, we argue that the central feature of the concept is that to be authentic is to act in accordance with values or preferences that are one’s own. We also develop a taxonomy of different conceptions of authenticity as substantive, non-substantive, or existentialist. The analysis identifies three implausible ideas about the self that tend to be implied in the concept of authenticity: essentialist views, existentialist views, and views that position the self as opposed to socialisation. In section 3, we draw on this analysis to develop a new way of conceptualising authenticity. We regard it as one condition for self-governance (a dimension of autonomy) and argue that an agent’s reasons are more authentic when that agent would endorse them upon (historically sensitive) reflection, or when the agent acknowledges and takes responsibility for them. We then show why this conception of authenticity avoids implausible conceptions of the self and is consistent with understanding the self as socially embedded and relational.

In section 4, we re-interpret and reassess the ethical concerns about neurotechnologies in light of this revised conception of authenticity. We show that the use of the concept of authenticity is unlikely to be fruitful, because concerns about authenticity tend to be premised on essentialist conceptions of the self, and technology can be positioned as either preventing or enabling the discovery of this self. However, assessing each concern in terms of autonomy – as an overarching value which may be supported by authenticity – provides a better way to capture the ethical force of the concerns.

# 1. Ethical concerns about neurotechnologies

Concerns about neurotechnologies and character change gained prominence in debates about psychopharmaceuticals during the 1980s and 90s. Kramer’s *Listening to Prozac* (1993) presented cases of patients who experienced alterations in character when using Prozac, some of whom interpreted these changes in terms of finding a truer self. One patient who was able, on Prozac, to transform her character into a new, preferred one, returned to treatment after being eased off medication to report: “I am not myself” (1993, 18). Kramer presents further typical examples: “’I am myself without the lead boots,’ ‘myself without swimming through Jell-O,’ ‘myself on a good day, although I never had days this good,’ ‘myself without fears’” (1993, 195-6).

Although Kramer emphasises that there is an alternative reading – that the availability of medication that can aid self-transformation can *alter* what traits we interpret as ‘our own’ – the idea that psychotherapeutics offer a route to authenticity is not uncommon. This idea is implied in advertisements by pharmaceutical companies for treatments: “GlaxoSmithKline … advertised Paxil with the slogan, ‘Relieve the anxiety; reveal the person’” (Elliot 2011, 368). Other influential works that examine not only neurotechnologies but also other forms of enhancement further strengthened the framing of these issues in terms of authenticity.[[4]](#endnote-4) Elliot (2003) relates self-reports of becoming more or less oneself on psychopharmaceuticals to similar self-reports related to other kinds of change, from dieting and exercise through to sex-reassignment surgery, presenting them as practices pursued in accordance with the modern Western ideal of authenticity (2003, Ch. 2). DeGrazia (2005) argues that neurotechnologies (and other interventions) can either support or compromise authenticity, depending in part on whether the choice to have the intervention is autonomous (2005, 102-112).

The latter argument reflects that experiences related to psychopharmaceuticals are not uniformly of finding a *more* authentic self – some people report feeling less themselves, even where the associated changes are positive (in clinical and everyday senses). An example reported by Kramer is the case of Philip, a moderately depressed young man struggling with difficult childhood experiences:

… his moodiness and irritability were comfortable to him, because they represented his legitimate suffering and rage […] On Prozac, Philip felt better than well, and he hated it. He had been prematurely robbed of his disdain, his hatred, his alienation [...] He felt phony, he did not trust himself. (1993, 291)

More recently, discussion has focused on implantable brain devices, particularly DBS. Similar patterns are found in DBS patient statements.[[5]](#endnote-5) The pivotal study by Schüpbach et al (2006, 1812) found that some patients reported feelings of strangeness or self-alienation: “I don’t feel like myself any more” and “I don’t recognize myself anymore”. This occurred even in cases where clinical outcomes were positive. A typical pattern in DBS studies is that some patients report becoming ‘more themselves’ post-treatment, while others interpret seemingly new post-treatment desires or traits as alien (e.g., de Haan et al 2017, 11-12; Dings and de Bruin 2016; Gilbert et al 2017; Johansson et al 2011; Kraemer 2013; Witt et al 2011). The suddenness of the change, the patient’s interpretation of the disease DBS aims to treat as part of or alien to them, and whether or not changes are part of or incidental to the aim of treatment, may all make a difference in how changes are interpreted (de Haan et al 2017, Gilbert et al 2017, Schechtman 2010).

Although the concerns about authenticity seem to be evoked by patient statements, the concept of authenticity carries with it the implication that we might be mistaken about whether or not we are authentic (as we shall see further below). Thus, the ethical concerns about neurotechnologies and authenticity do not always require that patients actually feel less authentic. The concerns include whether neurotechnologies: (1) are a problematic method of personal change; (2) encourage quietism; (3) encourage conformism; and (4) involve a loss related to ‘respect for the given’.

First, the concern about neurotechnologies as a method of change often contrasts it with other methods such as education or psychotherapy. Such methods may be thought superior because they are actively achieved, capable of being meaningfully first-personally understood, and involve emotional growth and increased self-knowledge. In contrast, change achieved technologically (it is claimed) is passive, cannot be made meaningful in the same way, and blocks emotional growth and self-knowledge. As a result, such change cannot be fully owned by the person (President’s Council on Bioethics 2003, 293; Sveneaus 2009; Hoffman 2013, 169; Kramer 1993, 254).

Second, it has been argued that although neurotechnological interventions might make people happier or more functional, the negative emotions they remove may sometimes be entirely appropriate to the situations people are in. People may be depressed or anxious, or develop dysfunctional behaviours, due to oppressive social norms or dysfunctional personal relationships (Kramer 1993, 254-5; 278; Kraemer 2011, 13 of 15; Schermer 2013). If neurotechnologies cause a loss of authentic responses to such situations, this may lead to quietism (Elliot 2011; Levy 2011, 310).

Third, some discuss the possibility that social pressures to have personality traits that can help to achieve success in work and social life might drive the use of neurotechnologies (President’s Council on Bioethics 2003, 284). Increased desires to use neurotechnologies might occur via individuals’ internalisation of social values and/or desires to avoid the impacts of stigma associated with noncomformity (Elliot 2011, 367). The plausibility of this idea is underlined by the way that a neurotechnology can seemingly be sold to people in the name of achieving authenticity, noted above. This could enforce social conformism, leading to losses of individuality, and so of individuals’ authenticity (President’s Council on Bioethics 2003, 284; Kramer 1993, 249; 269).

Fourth, some argue that increased capacities to choose or control what our characters and experiences are like involves a loss of value. Such views suggest that there is some value in accepting, or being capable of accepting, the way things are. The President’s Council on Bioethics for instance talks of respect for ‘the given’, or for what is “naturally human” (2003, 286) – which is sometimes conceptualised as how we authentically are, or should be. If we no longer need to accept but rather can change ‘the given’, we lose reverence for a world that is not neurotechnologically mediated, and the value associated with capacities to cope with such a world.

Before proceeding to examine the concept of authenticity, we note one concern that we will leave aside in what follows. The loss of ‘felt’ authenticity can itself be regarded as an ethical concern, since it is sometimes experienced as aversive (Gilbert 2013, 475; 2015, Johannson et al 2014, 29; Goering et al 2017; Kraemer 2011), and there is evidence that it correlates with other harms such as (in relation to DBS) post-operative suicide (Gilbert 2013). We agree this is an important clinical concern requiring sensitive management, but leave it aside for several reasons. It is unlikely to be informed by philosophical analysis of the meaning of authenticity, since it relies on patients’ self-interpretations. There may be as many patients who come to feel more rather than less themselves on using neurotechnologies, and those who find changes initially difficult may adjust post-treatment, reinterpreting the changes over time. Thus, it is important not to overstate the possibility of loss of felt authenticity upon treatment, not least because it might discourage people who could benefit from treatments (Gilbert 2018).

# 2. What is authenticity?

While the concept of authenticity has framed debates about neurotechnologies, little attention has been paid to the meaning of the concept, making the bioethical concerns less clear and leaving open the possibility that interlocutors in these debates are talking past each other (Parens 2005). In this section we examine the concept drawing on historical analyses of its development and contemporary debate about autonomy. We argue that the main feature of the concept of authenticity is that to be authentic is to live one’s life in accordance with values or preferences that are ‘one’s own’, rather than being imposed upon one ‘externally’. We then examine three ways the concept of authenticity is fleshed out, in what we call substantive, procedural, and existentialist views of authenticity (drawing on Mackenzie and Stoljar 2000).

Running through these ways of conceptualising authenticity are a number of partly overlapping distinctions, which reflect different ways of distinguishing authentic from inauthentic reasons, recommendations for achieving authenticity, and conceptions of the self. These include: self-discovery and self-creation; self-expression and self-control; self-knowledge and self-definition; passivity and activity; immutability and dynamism; and acknowledgement and endorsement. In general, the cluster of the first term of each of these distinctions leads to a conception of the self as fixed or pre-existing, which we must first uncover, and then express or enact. On this picture, we are passive in the face of the authentic self, limited to activities of discovering and expressing it. On the opposing picture, associated with the cluster of the second term in each distinction, the self is dynamic, created by our activities and decisions over time.

What we call substantive views tend towards the former cluster, and existentialist views toward the latter. Some (endorsement-based) procedural views are also closer to the former cluster, and other (acknowledgement-focused) procedural views to the latter. Where the meaning of authenticity is discussed in relation to neurotechnologies, authors have distinguished these two main conceptions, usually dubbed either self-discovery and self-creation, or essentialist and existentialist views (e.g., Levy 2011, Mackenzie and Walker 2015). However, the various distinctions may not precisely line up, and it is possible to conceptualise authenticity in ways that draw on different parts of different distinctions, or to combine aspects of these clusters. We agree with other analyses (Levy 2011; Schechtman 2004) that more plausible understandings of authenticity will be nuanced enough to recognise both ‘sides’ of this cluster. Our analysis will, in addition, reveal the need to avoid positioning authenticity as opposed to social influence, and that the ideal of authenticity needs to be conceptualised in a way that is consistent with accepting *some* values or criteria for living.

### 2.1. Living in accordance with values or criteria that are ‘one’s own’

The modern notion of authenticity can be explained via a contrast with a pre-modern value of being ‘true to oneself’ (Trilling 1971; Taylor 1989; 1991; Guignon 2004). This pre-modern value was a call to be true to oneself *in order* to be true to something outside of, and greater than oneself. Being true to oneself meant aligning oneself with an external source of value and meaning, which provided criteria for deciding how one should live, and gave meaning and value to one’s pursuits.

Various concepts may play the role of this external source. One such source is God: one is to align oneself properly with God, and this requires being true to oneself because knowledge of God comes from looking within – listening to conscience, the voice of God within us (Taylor 1991, 26). Another is other people: one should be true to oneself in order to be true to others; to meet one’s social and moral obligations (Trilling 1971, 9). In pre-modern societies, how individuals were to achieve this was largely set by the social hierarchies of their time and place, which provided them with an understanding of their role in society and their proper relationships with others (Taylor 1994). A third source was sets of beliefs associated with teleological conceptions of human nature and of the cosmos, which similarly provided normative direction (Guignon 2004, 12).

The belief sets associated with the pre-modern notion of being true to oneself declined over the modern period. The scientifically, mechanistically understood natural world provides no meaning or normative direction (Taylor 1989, 417). Social hierarchies come to be understood as illegitimate sources of inequality, rather than legitimately defining people’s roles, and begin to change. Conceptions of the good become increasingly securalised (Taylor 1991, 44). As a result, the modern ideal of authenticity develops: the idea of being true to oneself is retained but the external source(s) of value are dropped. Authenticity thus becomes an end in itself rather than a means to some other, valued end (Guignon 2008, 278; Varga and Guignon 2014). The self, “with no defining connections to anything distinct from itself”, becomes the source of meaning (Guignon 2008, 279; Taylor 1991, 29; Trilling 1971, 25).

Different conceptions of authenticity thus arise from different ways of conceptualising what is within, and what external to the self. Further, externally-derived criteria or values are sometimes conceived as potential sources of oppression or distortion of the self. In the next three subsections we consider different ways that authenticity may be fleshed out from this point: substantive views, on which authenticity aligns with something of value that is common to all, but not external to the self; non-substantive views, on which authenticity involves procedural or formal constraints; and existential views, on which authenticity requires strict rejection of any values or meaning-making processes that are not self-originated.

### 2.2 Substantive authenticity

Although authenticity calls on us to reject values or preferences that are ‘external’, the way the distinction between internal and external is understood leaves room for authenticity to be aligned with something common to all persons. Two examples of ‘substantive’ views associate authenticity with nature, and with sentiment.

The paradigmatic view aligning authenticity with nature is from Rousseau. Rousseau equated authenticity with genuine self-regard (amour-de-soi), moral, physical and intellectual independence from others, and self-mastery (Rousseau, [1755]/1973). Achieving this kind of authenticity requires us to reconnect with our true and proper natural selves and with the voice of conscience (Taylor 1989, 358). In Rousseau’s view, this could only be achieved by distancing ourselves from the corrupting influences of society, and dependence on the opinion of others, both of which give rise to egoistic forms of self-love (amour-propre) (Rousseau, [1755]/1973, Trilling 1971, 30-9). This makes the task of authenticity into one of finding our reasons for action within ourselves (Taylor 1989, 361), where the boundary between reasons from within and from without is understood in terms of rejection of the corrupting influences of sociality.

This idea that the authentic person transcends the ills of socialisation is still evident in modern thinking (Trilling 1971, 167-72), even in strands of thought where the evaluation of the authentic self as fundamentally good is altered. Some post-romantic thought suggests that our natural, unsocialised selves might rather be aggressive, cruel, destructive, and uncontrolled (Trilling 1971, 56-7; Taylor 1989, 434-8; Guignon 2004, 95). Such views may position society and internalised socialisation as positive influences, enabling us to overcome our nature, or more moderately, as a necessary check on our natural state that enables cooperation with others (Trilling 1971, 110). Perhaps the most clearly formalised example is found in Freudian theory, according to which socialisation does indeed suppress and distort what is natural to us, sometimes in ways that are psychologically difficult, but is also required for civilisation (Trilling 1971, 157). But, even if the natural or non-socialised self is not intrinsically good, recognition of it may be thought to have value (Trilling 1971, 110).

A second substantive view of authenticity associates it with sentiment. This gives substance to the notion of authenticity while endorsing individuality because, although the fact of having sentiments is universal, the content of sentiments is specific to each individual. For the romantics, our sentiments are an inner voice that can point us toward truth (Taylor 1989, 371; 368-9). Expressing one’s real feelings thus became in romantic thinking vitally important. Since one’s authentic feelings might also have been distorted by society, this may be difficult. Authenticity becomes associated with the creation of art, which provides a process for expressing, but also for giving shape to, one’s sentiments. Such self-expression enabled both self-discovery and self-constitution (Taylor 1991, 61-3; Trilling 1971, 100).

### 2.3 Non-substantive authenticity

Alternatively, we might regard authenticity as formal or procedural. Here we turn to the contemporary autonomy literature, which understands what makes a reason one’s own as a formal feature of a person’s motivational structure, or as secured procedurally by the activity of the will.[[6]](#endnote-6)

In this literature, the most influential way of distinguishing authentic and inauthentic reasons appeals to endorsement following critical reflection, drawing on an hierarchical understanding of the will (Frankfurt 1971). This view implies that one’s authentic reasons are internally coherent, and that authenticity involves evaluative approval or ‘identification’. Critics have argued, however, that sometimes the ‘real’ self is revealed, not in what a person chooses, but in what occurs when they are out of control (Schechtman 2004), or in desires that an agent might regard as wayward (Mackenzie 2002, 190). We might be more limited in our ability to decide and control what we’re like than the endorsement picture suggests; in fact, understanding authenticity as endorsement might lead us to self-deception, to ignoring and repressing aspects of ourselves that we dislike (Velleman 2002a). Further, the endorsement view of authenticity overlooks that the viewpoint from which we reflect might be subject to distorting influences such as social oppression (Mackenzie 2002, 192).

In response to these issues, some have suggested revisions to our understanding of endorsement.[[7]](#endnote-7) We examine an alternative way of conceptualising authenticity conditions derived from discussions by Benson and Oshana. We refer to these as ‘acknowledgement’ views. Benson begins from the idea that we can sometimes seemingly act autonomously from motives that we do not endorse (2005, 105). An autonomous person can be defined, even centrally, by aspects of their character with which they do not identify. This suggests one might not need authenticity in the ‘endorsement’ sense for autonomy, and (although he does not call this a sense of authenticity) Benson argues that we may claim normative authority over some action of ours even when it is not motivated by an endorsed reason. As he puts it, “I can take ownership of my actions even when they do not align with who I am or what I stand for” (Benson 2005, 104). Similarly, Oshana distinguishes what she calls an ‘epistemic’ conception of authenticity from an endorsement-focused view. This involves acknowledging and taking responsibility for one’s reasons even if one is alienated from them (2007, 14-15). Authenticity, on this view, is not a matter of choosing what reasons we identify with, but a matter of correctly acknowledging and taking responsibility for who we are, including aspects of ourselves that we don’t endorse.

An example may help clarify; consider the unwilling addict, who has a first-order desire for a drug, but a second-order volition not to act on the first-order desire (Frankfurt 1971, 12). We can imagine such an addict saying that the first-order desire is ‘not really his’, or ‘not part of his real self’. This expression of disavowal can be regarded as a decision, an act of the will that defines what sort of person he wishes to be, and (we can imagine) this way of thinking about the first-order desire may help him to succeed in not acting upon the first-order desire (cf. Velleman 2002b). But it might also be a kind of self-deception, leading him to repress the desire rather than coping with it, or even to dodge responsibility for acting on the first-order desire should he give in to it. An alternative to disavowing the desire may be to acknowledge it as a desire of his own (perhaps by saying things like ‘I’m an addict’ at Narcotics Anonymous meetings). This could be a way of saying that he takes responsibility for coping with the desire, including resisting acting upon it.

Benson’s and Oshana’s discussions suggest that endorsement is not, or is not the only way we can understand what it means for a reason for action to be one’s own. It might involve acknowledgement and taking responsibility for who one is (Oshana 2007, 14-5), rather than choosing who one will be. This helps to make sense of the critiques of endorsement views noted above, emphasising that aspects of ourselves we have not chosen and perhaps do not approve of play a role in who we are; and recalls the post-romantic view that there is value in recognising and acknowledging what we are like, even if that is to recognise destructiveness or cruelty.

### 2.4 Existentialist views

Some existentialist views propose that authenticity is incompatible with accepting any constraints on the self, either substantive or procedural. Acceptance of any constraints essentialises the self and leads to bad faith and inauthenticity. What is needed for authenticity is the recognition that one lacks any essence and is radically free (Varga and Guignon 2014). The task of authenticity becomes that of creating oneself and generating meaning. This conception of authenticity tends to be quite unstable (Levy points out that even Sartre, who presents its paradigmatic statement, held it only for a short time and relied on an “extravagant” metaphysics (2011, 312)).[[8]](#endnote-8) The authentic individual, insofar as s/he transcends any externally-given values, would seem to be subject to an ongoing existential crisis, but is at the same time regarded atomistically, as separable from its surrounds.

The instability of such a view connects to criticisms that the ideal of authenticity is self-defeating or vacuous. The insistence that no external criteria should guide the self is either incoherent, if we think no criteria are available any other way (Trilling 1971, 164); or it is a way to justify self-indulgence and self-centredness, if it reduces criteria for choice to contingent individual preference or whim (Taylor 1991, 14-15).

### 2.5. Authenticity and the social self

These criticisms suggest that authenticity might instead be conceptualised in a way that is consistent with recognition of the socially constituted and situated nature of the self. Trilling notes various attempts to develop this thought (1971, Ch.2) and argues for the need to develop a more nuanced notion of the self (165-71). Taylor argues that the worth of the ideal of authenticity can be retrieved by recognising that value does not arise from each individual (not even for those individuals), but rather from intersubjective processes that rely on recognising oneself as part of a community (1991, 39). He regards the self as “dialogical”, created in a negotiation with others within a shared horizon of meaning (1991, 35). This does not mean that an individual’s values are ‘externally’ set; we play active roles in meaning-making, and authenticity involves both the discovery and the creation of values (1991, 39). Authenticity thus requires not disavowal but acceptance of unconditional relationships and moral demands beyond the self (1991, 72). Though the specifics differ, we can see similarities between this approach to authenticity and relational approaches to autonomy. Since persons develop within cultures and through relationships with others, it is both descriptively implausible and normatively undesirable to regard them as isolated or isolatable, or to align autonomy with independence.

This analysis provides direction for thinking about what a conception of authenticity that avoids implausible conceptions of the self will look like. As others have noted, it should avoid the extreme existentialist view, which implausibly denies the extent to which our selves are created through processes with respect to which we are passive, or that are relatively fixed (e.g. our pasts, and perhaps natural capacities or dispositions). It should also avoid an essentialising view which goes too far in the opposite direction, to deny our capacities to change or align acceptable change with some pre-existing self, awaiting discovery. It should avoid implying that authenticity is to be aligned solely with activity and self-control, or solely with passivity and self-expression. Our analysis has two further implications. First, a plausible conception of authenticity should avoid regarding the self as separate from or prior to social relationships and socialisation. Second, to avoid becoming empty and incoherent, the ideal of authenticity must be consistent with accepting *some* values or criteria for living. In the next section, we propose a procedural understanding of authenticity which aims to meet these requirements.

# 3. Authenticity and autonomy

Autonomy is typically equated with self-governance. To be an autonomous agent is to be self-governing: to make decisions and act on the basis of one’s own reflective preferences, values or commitments, rather than under the direction or control of others. This conception of autonomy, however, seems virtually indistinguishable from the analysis of authenticity as living one’s life in accordance with values or preferences that are ‘one’s own’.[[9]](#endnote-9) Moreover, in the mainstream literature it is typically understood in ways that presuppose an individualistic rather than social conception of the self. In this section, we first argue that autonomy is a broader concept than authenticity, drawing on Mackenzie’s (2014) multi-dimensional and relational analysis of autonomy. We then explain the relationship between the concepts of authenticity and self-governance. With these conceptual clarifications in place, we propose an interpretation of authenticity as involving either endorsement or acknowledgment, and suggest that this interpretation can avoid the extremes of essentialism and existentialism as well as being consistent with a social conception of identity.

Mackenzie (2014) argues that autonomy is a complex concept, which is conceptually connected to a range of other concepts, and comprises three distinct but causally interconnected dimensions or axes: self-determination, self-governance and self-authorization. To be self-determining is to have the authority and power to exercise both *de jure* and *de facto* control over important domains of one’s life. This authority and power is a function of the freedoms and opportunities available to a person in her social environment and of her social standing or status. The self-determination axis illuminates the conceptual connections between autonomy and freedom that have been highlighted by socio-relational theorists of autonomy (such as Oshana 2006). It helps explain why it can be difficult to lead an autonomous life in contexts of social oppression: because these social environments constrain individuals’ freedom and opportunities and deny them social standing as autonomous agents.

To be self-governing is to possess and exercise the competencies required to make decisions and act in ways that express or cohere with one’s identity, values and commitments. The *self-governance* axis illuminates the conceptual connections between autonomy, authenticity and competence. Importantly, on Mackenzie’s view, both competence and authenticity are understood relationally. Authenticity is relational because individuals’ identities, values and commitments are shaped by social relationships and by intersecting social determinants. Competence is relational because the skills and capacities required for governing the self are developed and scaffolded (or not, in oppressive social contexts) by the social environment.

To be self-authorizing is to regard oneself as having the authority to take ownership of, or responsibility for, who one is, and for one’s values, commitments and decisions. It also involves regarding oneself as an equal participant in reciprocal accountability relations – as able to account for oneself to others, and also to hold others to account. To be self-authorizing requires that a person hold appropriate attitudes of self-respect, self-trust and self-esteem. These attitudes are inherently relational and social, because they are developed and sustained in social relations and in the context of normative structures and practices of socialrecognition.

This analysis of autonomy clarifies the conceptual connections between autonomy and authenticity. Authenticity and autonomy are not equivalent. Authenticity is a condition for one dimension of autonomy: self-governance. However, it is not sufficient for self-governance, which also requires the exercise of a range of socially scaffolded competences, including rational, emotional, and imaginative skills, as well as capacities for self-control. The analysis also helps to clarify the relationship between authenticity and socialisation. Since who we are, as well as our values and commitments, are shaped by our social relationships, it is a mistake to define authenticity in opposition to social influence. This is not to deny, however, that oppressive forms of socialisation that demand conformism, inhibit exploration and reflection, and limit individuals’ imaginative horizons can thwart authenticity.

With this analysis in place we now want to argue in favour of extending the notion of authenticity to include both endorsement and acknowledgement. Endorsement-based accounts assume that only reasons for action with which we identify are authentic. However, as we argued above, this view fails to account for the importance for our identities and motivational structures of some non-endorsed reasons for action or characteristics (Oshana 2005, Mackenzie and Poltera 2009, Walker 2018). These non-endorsed aspects of ourselves may also be authentically our own, we suggested, if we acknowledge, and take ownership of or responsibility for them.

Understanding that acknowledgment may be a route to authenticity avoids the pitfalls of both the essentialist and existentialist versions of authenticity. The view allows that both chosen and unchosen features of one’s identity, and both chosen and unchosen social relationships can be authentic. Acknowledgement is not merely a matter of forming correct beliefs about what one is like (in respects one may be passive towards or unable to change). Nor should it be understood as a matter of discovering pre-existing or fixed traits, preferences, and desires. It is an active process of interpreting and evaluating our identities, values and commitments and making decisions about whether and how we will act on them. It is thus a process of constituting ourselves while recognising the ways in which the process of self-constitution is shaped by features of our identities and relationships that are unchosen or difficult to change. This attention to the unchosen or difficult-to-change aspects of who one is, and to the reasons one already has, also explains how the acknowledgement account of authenticity avoids the pitfalls of the existentialist view. Being authentic is not just a matter of being whoever one chooses to be, since our identities are shaped by characteristics, social circumstances and relationships that are often not of our own making.

Both endorsement and acknowledgement can be routes to authenticity in different circumstances or for different persons. Thus we do not need to choose between them. For example, recalling the ‘unwilling addict’ example, empirical studies confirm that some people who are recovering from addiction do in fact benefit from the endorsement approach: they distinguish a ‘real self’ from an ‘addictive self’, and attribute addictive desires to the addictive self as distinct from the real self. This form of disavowal can support attempts to change behaviour (Larkin and Griffiths 2002). Others reportedly benefit more from an acknowledgement approach, owning and identifying with their addictive desires as part of regarding themselves to have a responsibility not to act on them (Walker 2015, 36-8). Moreover, acknowledgement and endorsement can interact. Acknowledgement requires certain kinds of self-knowledge, recognising traits, characteristics and preferences as one’s own. These might be endorsed or disendorsed, or altered as they are reflected upon and interpreted, and sometimes new reasons to act can be created in these processes. The unwilling addict who repudiates his first-order desire for a drug, but acknowledges it to be *really his* desire, not only disavows the first-order desire as a reason to act but generates other reasons to act, i.e., provides himself with a reason not to take the drug. That is, disendorsement of a reason can generate other reasons to act, or change one’s overall set of reasons in a positive, and not only a negative way.

It is important to stress that the kind of self-knowledge and authenticity that is achieved through processes of endorsement and acknowledgment involves not just introspection but interactions with others, in the world, over time. We cannot understand who we are – our relatively stable commitments, preferences, values and desires, our patterns of emotional, intellectual and evaluative response – simply by introspection. Rather we come to understand ourselves, and what activities and interactions we find fulfilling and meaningful, through action and interaction with others over time. This account of authenticity is thus consistent with a relational and social conception of the self and of self-governing agency. It also recognises that authenticity is a scalar notion. A person is not either authentic or inauthentic. Authenticity is achieved in a piecemeal way over time and we can be more or less authentic at different times and in different contexts.

By clarifying the relationship between authenticity and autonomy we hope to have shown that the value of authenticity derives from its role as a condition of or component of self-governance. In arguing that acknowledgement can be a route to authenticity, we have tried to develop an account of authenticity that avoids the pitfalls of both essentialist and existentialist interpretation of authenticity. And we have argued that it is a mistake to present socialisation as inimical to authentic individuality. An adequate account of authenticity must be based on recognition that individual identity and agency are shaped and scaffolded by social relationships and the social environment.

# 4. Revisiting the ethical concerns

In light of the preceding analyses, in this section we reinterpret and reassess the ethical concerns about neurotechnologies. In relation to each concern, we examine implicit assumptions about authenticity, showing that the concerns tend to rely on essentialist conceptions. We further show that the concerns relating to the method of change and respect for the given are implausible on our view of authenticity, but that the concerns relating to quietism and conformism can be reinterpreted as concerns about autonomy. This reinterpretation better captures the issues at stake, without implying an implausible conception of the self, and enables more refined ethical consideration of the effects of neurotechnologies on both authenticity and other conditions for autonomy, such as competence conditions.

The first concern was that the personal change achieved using neurotechnologies cannot be fully owned by a person, in contrast to personal change achieved through other means such as education or psychotherapy. This concern implies that the kinds of self-work deemed superior are those involving introspective self-discovery, such that self-change is authentic when it results from correct self-understanding. This assumes an essentialised picture of the self as ‘awaiting discovery’, and overlooks the possibility that self-change might occur when one interacts with others differently or undertakes new activities, and this generates new self-knowledge. The latter kind of process is what seems to occur with neurotechnological interventions. The intervention may alter how people behave or interact with others first (e.g. by altering their emotional responses), and in doing so support the development of better self-knowledge. For instance, a neurotechnology might ameliorate a depressed mood that tended to result in overly negative self-interpretations. Kramer notes that in his experience “many patients, including some who have never had a diagnosable mental illness, are better able to explore both their past and their current circumstances while they are taking Prozac” (278). Neurotechnologically-achieved change may support autonomy in this way, though there might indeed be other cases where neurotechnologically-achieved change does not support autonomy.

Our discussion further implies, however, that it is the effects of any changes, and how those changes impact on the person, that would be at stake in making judgements about these matters – not the method by which the changes are achieved. On the view we have developed, a reason’s authenticity is not determined by its causal history. An authentic reason might be arrived at through introspective reflection, but can also arise through social and relational influences (and various combinations of these in processes of endorsement and acknowledgement). As such, reasons we have as a result of some process that includes a neurotechnological intervention may be either authentic or inauthentic (or have some degree of authenticity). Authenticity, when plausibly construed, is not a helpful concept in making this determination.

Second, the concern about quietism was that neurotechnologies might reduce or remove negative feelings that are appropriate responses to social oppression, dysfunctional relationships, or other situations that one *should* not accept. This concern, if understood as a concern about authenticity, again relies on an essentialising conception of authenticity, since it implies that social oppression or dysfunctional relationships are problems because they interfere with being true to a pre-existing authentic self that is the source of the negative emotions. Understanding this concern as, instead, about autonomy provides an alternative interpretation, which better captures its real force. People would indeed be worse off if neurotechnological interventions removed negative feelings that signalled that there was something wrong. But this is not because the neurotechnologies would thereby prevent people from recognising some such ‘essentialist’ self. Rather, it would be because the neurotechnologies would result in people being less able to control their own lives, by influencing them to accept, rather than resist, social oppression or dysfunctional relationships that interfere with their autonomy.

As well as better capturing what motivates this concern, understanding it in terms of autonomy enables consideration of how neurotechnologies might impact on other conditions for autonomy, in addition to authenticity conditions. By altering feelings, neurotechnologies might impact authenticity conditions: perhaps the loss of negative feeling would change what reasons a person acknowledges to be their own, or the reasons from which they feel non-alienated. But they are also likely to impact on other conditions for autonomy. Both physical and psychological conditions treated with psychopharmaceuticals and implantable brain devices – Parkinson’s, depression, dystonia, obsessive compulsive disorder – can impact on the competence conditions for self-governance. Some of these conditions impact on physical capacities, and others reduce capacities for self-control; neurotechnologies may thus support autonomy by supporting autonomy-relevant competences. They might also affect other capacities that (on various theories) support or are required for autonomy. For instance, conditions such as depression might reduce capacities for self-knowledge, or have a negative impact on self-attitudes like self-trust or self-esteem (Mackenzie 2008). All of these attributes may be relevant in resisting social oppression or dysfunctional relationships. An overall judgement about whether or not a neurotechnology is likely to induce quietism would need to take these non-authenticity conditions for autonomy into account.

Third, the concern about conformism was that neurotechnologies might increase social pressure for individuals to align their personalities to those personality types that are most highly socially rewarded. This might undermine individuality, and thus authenticity. Again, if understood as a worry about authenticity, there seem to be essentialist assumptions at work, since the concern positions ‘authentic’ individuality in opposition to a homogenous, socially-approved personality type. The underlying assumption is that social pressures are a pernicious influence on individual choices that prevent people from being their real selves. Again, regarding this concern to be about autonomy provides an alternative interpretation, which avoids essentialising notions of the self, and better captures the real force of the worry. We would be worse off if neurotechnologies increased conformism, but this would be because conformist pressures could threaten people’s autonomy. Again, conformist pressures might impact authenticity conditions for autonomy by making it more difficult to acknowledge reasons that are less socially acceptable as one’s own, or by influencing people to accept reasons as valid even though, on historically sensitive reflection, they might have evaluated the reason differently. But these effects would be concerning because they reduce people’s capabilities for self-governance, not because they prevent people from being some ‘real self’.

Conformist pressure might also impact on other conditions for autonomy. Increased conformist pressure might, like reduced negative emotions, reduce capacities for self-knowledge or impact on self-attitudes such as self-trust. Further, increases in social conformity might constrict the social imaginary, thereby restricting individuals’ imaginative repertoires (see e.g., Mackenzie 2000). However, neurotechnologies might conversely have positive influences in relation to conformism – such as increasing assertiveness or a sense of self-trust that supports individuals’ willingness to challenge conformist pressure. We do not seek to resolve whether or not it is likely that neurotechnologies will increase conformism: our point is that autonomy is a better frame for such evaluation.

The fourth concern was that of loss of respect for the given ­– the idea that there is value in the way things are, or in being able to accept things as they are. This concern seems to rely on an even more strongly essentialist picture than those above. It implies that one’s self is ‘given’ and therefore must be accepted. This concern is difficult to reinterpret if one eschews an essentialist self. A more plausible aspect of the concern, perhaps, is the idea there could be value in the *ability* to accept the way things are. Capacities for resilience, or stoicism in the face of difficulty may well be highly useful in life, for example. But again, this would be because they may help people to gain more control over their lives, to be more autonomous, not because doing so allows people to align themselves with some given object of reverence. Our view of the authenticity conditions for autonomy recognises that we are sometimes relatively limited in our capacity to control who we are or what we are like, and includes the need to accept and acknowledge fixed aspects of ourselves. But acknowledgement is not always acceptance of the given, or reverence for it: sometimes it involves repudiation. While we do need to acknowledge that some aspects of the self are relatively fixed – and that the perspective from which we make decisions about self-change relies on our pasts and on social influences – there seems no reason to conclude that acceptance of what is given is the best response, or that what is unchosen should be revered.

In general then, the concerns, when framed as relating to authenticity, tend toward essentialising conceptions of the self, leading those warning against the use of the neurotechnologies to position technology as a source of inauthentic reasons, while those in favour position them as a way of becoming more authentic. This helps explain why authenticity is not a helpful way to frame concerns about neurotechnologies: it can be conceptualised as either compatible with or opposed to technology, but both critics and proponents rely on an implausible, essentialist understanding of the concept of authenticity. A debate that turns on attempting to understand the relation between an essentialist ‘true self’ and a neurotechnology is unlikely to be fruitful, or to be resolved. Further, we have shown that what plausible content there is to these concerns is better articulated in terms of autonomy. Framing the debate in terms of autonomy enables consideration of how neurotechnologically achieved change might impact on the authenticity conditions for autonomy, but also draws attention to the potential impacts on other conditions for autonomy, including competence conditions. This is important given that, in many cases, clinical gains from neurotechnologies are likely to support important autonomy competences, such as physical capacities and self-control, and these may significantly outweigh concerns about authenticity in decision-making about neurotechnological interventions.

# 5. Conclusion

We have argued that the focus on authenticity in relation to ethical concerns about neurotechnologies is unhelpful. When these ethical concerns are understood in terms of authenticity, they tend to imply an implausible, essentialist view of the self. Our examination of the concept of authenticity identified a range of ways authenticity may be conceptualised, and we argued that a plausible conception should avoid both overly essentialist or existentialist extremes, and views that position the self as opposed to socialisation. We proposed a conception of authenticity that avoids these problematic views: an agent’s reasons are more authentic when the agent would endorse them upon (historically sensitive) reflection, or when she acknowledges and takes responsibility for them. The view implies that authenticity is of value because it can increase autonomy, a more expansive concept. Applying this view to the ethical concerns about neurotechologies, we showed that insofar as these are plausible ethical concerns, they are better articulated as worries about autonomy.

# References

Baylis F. 2013. “I am who I am”: On perceived threats to personal identity from deep brain stimulation. *Neuroethics* 6:513-526.

Benson P. 2005. Taking ownership: Authority and voice in autonomous agency. In *Autonomy and the Challenges to Liberalism: New Essays*, ed. J Christman, J Anderson, 101-126. Cambridge: Cambridge University Press.

Bolt I, Schermer M. 2009. Psychopharmaceutical enhancers: Enhancing identity? *Neuroethics* 2:103-11.

Christman J. 1991. Autonomy and personal history. *Canadian Journal of Philosophy* 21(1):1-24.

Christman J. 2009. *The Politics of Persons: Individual Autonomy and Socio-Historical Selves*. Cambridge: Cambridge University Press.

DeGrazia D. 2005. *Human Identity and Bioethics*. Cambridge: Cambridge University Press.

De Haan S, Reitveld 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):e0175748.

Elliot C. 2003. *Better Than Well*. Cambridge, MA: Harvard University Press.

Elliot C. 2011. Enhancement technologies and the modern self. *Journal of Medicine and Philosophy* 36:364-74.

Frankfurt H. 1971. Freedom of the will and the concept of a person. *Journal of Philosophy* 68(1):5-20.

Friedman M. 1986. Autonomy and the split-level self. *Southern Journal of Philosophy* 24(1):19-35.

Gilbert F. 2013. Deep brain stimulation for treatment-resistant depression: postoperative feelings of self-estrangement, suicide attempt and impulsive-aggressive behaviours. *Neuroethics* 6:473-481.

Gilbert F, Goddard E, Viana JNM, Carter A, Horne M. 2017. I miss being me: Phenomenological effects of deep brain stimulation. *AJOB Neuroscience* 8(2):96-109.

Gilbert F, Viana JMN, Ineichen C. 2018. Deflating the “DBS causes personality changes” bubble. *Neuroethics* (online first) doi:10.1007/s12152-018-9373-8.

Goddard E. 2017. Deep brain stimulation through the “lens of agency”: Clarifying threats to personal identity from neurological intervention. *Neuroethics* 10:325-335.

Goering S, Klein E, Dougherty D, Widge AS. 2017. Staying in the loop: Relational agency and identity in next-generation DBS for psychiatry. *AJOB Neuroscience* 8(2):59-70.

Guignon C. 2004. *On Being Authentic*. New York: Routledge.

Guignon C. 2008. Authenticity. *Philosophy Compass* 3/2:277-290.

Heyes C. 2006. Foucault goes to weight-watchers. *Hypatia* 21(2):126-49.

Hoffman GA. 2013. Treating yourself as an object: Self-objectification and the ethical dimensions of antidepressant use. *Neuroethics* 6:165-178.

Johansson V, Garwicz M, Kanje M, Shouenborg J, Tingstrom A, Gorman U. 2011. Authenticity, depression, and deep brain stimulation. *Frontiers in Integrative Neuroscience* 5(21) :1-3.

Johansson V, Garwicz M, Kanje M, Halldenius L, Shouenborg J. 2014. Thinking ahead on deep brain stimulation: an analysis of the ethical implications of a developing technology. *AJOB Neuroscience* 5(1):24-33.

Juengst E, Mosely D. 2015. Human enhancement. In *Stanford Encyclopedia of Philosophy* (Spring 2016 edition), ed E Zalta. Available at <https://plato.stanford.edu/entries/enhancement/>, retrieved 15 November 2018.

Kraemer F. 2011. Authenticity anyone? The enhancement of emotions via neuro-psychopharmacology. *Neuroethics* 4:51-64.

Kraemer F. 2013. Me, myself, and my brain implant: Deep brain stimulation raises questions of personal authenticity and alienation. *Neuroethics* 6(3):483-97.

Kramer P. 1993. *Listening to Prozac*. New York: Viking.

Larkin M, Griffiths MD. 2002. Experience of addiction and recovery: The case for subjective accounts. *Addiction Research and Theory* 10(3):281-311.

Levy N. 2011. Enhancing authenticity. *Journal of Applied Philosophy* 28(3):308-318.

Mackenzie C. 2000. Imagining oneself otherwise. In *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, ed. C Mackenzie, N Stoljar, 124-150. New York: Oxford University Press.

Mackenzie C. 2002. Critical reflection, self-knowledge, and the emotions. *Philosophical Explorations* 5(3):18-206.

Mackenzie C. 2008. Relational autonomy, normative authority and perfectionism. Journal of Social Philosophy 39(4):512-33.

Mackenzie C. 2014. Three dimensions of autonomy: A relational analysis. In *Autonomy, Oppression, and Gender*, ed. A Veltman, M Piper, 15-41. New York: Oxford University Press.

Mackenzie C, Stoljar N. 2000. Introduction: Autonomy refigured. In *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, ed. C Mackenzie, N Stoljar, 3-31. New York: Oxford University Press.

Mackenzie C, Poltera. J. 2010. Narrative integration, fragmented selves and autonomy. *Hypatia* 25(1): 31–54.

Mackenzie C, Walker MJ. 2015. Neurotechnologies, personal identity, and the ethics of authenticity. In *Springer Handbook of Neuroethics*, ed. J Clausen, N Levy, 373-92. Dordrecht: Springer.

Meyers D. 1989. Self, society, and personal choice. New York: Columbia University Press.

Oshana M. 2005. Autonomy and self-identity. In *Autonomy and the Challenges to Liberalism: New Essays*, ed. J Christman, J Anderson, 77-97. Cambridge: Cambridge University Press.

Oshana M. 2006. *Personal Autonomy in Society*. Aldershot: Ashgate.

Oshana M. 2007. Autonomy and the question of authenticity. *Social Theory and Practice* 33(3):1-19.

Parens E. 2005. Authenticity and ambivalence: Toward understanding the enhancement debate. *Hastings Center Report* 35(3):34-41.

President’s Council on Bioethics. 2003. *Beyond Therapy: Biotechnology and the Pursuit of Happiness*. Available at <https://biotech.law.lsu.edu/research/pbc/reports/beyondtherapy/beyond_therapy_final_report_pcbe.pdf>, retrieved 15 November 2018.

Rousseau JJ. 1775/1973. Discourse on the Origins of Inequality. In *The Social Contract and Other Discourses*. New York: Everyman.

Schechtman M. 2004. Self-expression and self-control. *Ratio* 17:409-27.

Schechtman M. 2010. Philosophical reflections on narrative and deep brain stimulation. *Journal of Clinical Ethics* 21(2):133-139.

Schermer M. 2013. Health, happiness and human enhancement – dealing with unexpected effects of deep brain stimulation. *Neuroethics* 6:435-445.

Schupbach M, Gargiulo M, Welter ML, Mallet L, Behar C, Houeto JL, Maltete D, Mesnage V, Agid Y 2006. Neurosurgery in Parkinson disease: A distressed mind in a repaired body? *Neurology* 66: 1811–1816.

Stoljar N. 2018. Answerability: A Condition of Autonomy or Moral Responsibility (or Both)? In *Social Dimensions of Moral Responsibility*, eds. K. Hutchison, C. Mackenzie, M. Oshana, 231-252, New York: Oxford University Press.

Sveneaus F. 2009. The ethics of self-change: Becoming oneself by way of antidepressants or psychotherapy? *Medicine, Health Care and Philosophy* 12:169-178.

Taylor C. 1989. *Sources of the Self*. Cambridge, MA: Harvard University Press.

Taylor C. 1991. *The Ethics of Authenticity*. Cambridge, MA: Harvard University Press.

Taylor C. 1994. The politics of recognition. In *Multiculturalism and the Politics of Recognition,* ed A Gutmann, 25-73. New Jersey: Princeton University Press.

Trilling 1971. *Sincerity and Authenticity*. Cambridge, MA: Harvard University Press.

Varga S, Guignon C. 2014. Authenticity. In *Stanford Encyclopedia of Philosophy* (Fall 2017 edition) ed. E Zalta. Available at <https://plato.stanford.edu/entries/authenticity/>, retrieved 16 November 2018.

Velleman JD. 2002a. Identification and Identity. In *Contours of Agency: Essays on Themes from Harry Frankfurt*, ed. S Buss, H Overton, 91–123. Cambridge, MA: MIT Press.

Velleman JD. 2002b. Motivation by ideal. *Philosophical Explorations* 5(2):89-103.

Walker MJ. 2014. Powerlessness and responsibility in Twelve Step narratives. In *Sobering Wisdom: Philosophical Explorations of Twelve Step programs*, ed. N Plants and J Miller, 30-41. Charlottesville: University of Virginia Press.

Walker MJ. 2018. Two senses of narrative unification. *Philosophical Explorations* (online first 18 May). https://doi.org/10.1080/13869795.2018.1474943.

Westlund A. 2009. Rethinking relational autonomy. *Hypatia* 24(4):26-49.

Witt K, Kuhn J, Timmerman L, Zurowski M, Woopen C. 2013. Deep brain stimulation and the search for identity. *Neuroethics* 6:499-511.

1. Much of the literature on deep brain stimulation (DBS) analyses these kinds of statements in terms of identity, variously understood. For reasons articulated previously by, e.g., Baylis (2013), Mackenzie and Walker (2015), Goddard (2017) we begin instead from authenticity. [↑](#endnote-ref-1)
2. Autonomy is also sometimes defined in this way, as we discuss below, reflecting the view that authenticity is one condition for autonomy. One of our aims in this paper is to clarify the relationship between the concepts of authenticity and autonomy. [↑](#endnote-ref-2)
3. A similar argument is made by Baylis (2013) in slightly different terms. Baylis counters claims that DBS alters personal identity, and suggests that arguments about the ethical implications of neurotechnology should focus instead on agency. Goddard (2017) similarly proposes shifting focus from identity to autonomous agency, but she also discusses authenticity. [↑](#endnote-ref-3)
4. The concept of authenticity, accordingly, also plays a role in the extensive literature debating cognitive and other sorts of enhancement, which we leave aside here (for an overview see Juengst and Mosely 2015, section 4). [↑](#endnote-ref-4)
5. As noted above, debate about DBS has been partly conducted in terms of identity. As critics have noted, this may have generated confusion given variation in how identity is understood (Baylis 2013; Mackenzie and Walker 2015; Goddard 2017). For an overview of concepts used in debates about DBS see Gilbert et al (2017). [↑](#endnote-ref-5)
6. Christman distinguishes between competence and authenticity conditions for autonomy (2009, 134). Competence conditions include meeting some threshold for minimal rationality, having self-control, and lacking pathology. Authenticity conditions refer to acting on reasons that are really one’s own. [↑](#endnote-ref-6)
7. Others reject authenticity as a condition for autonomy, suggesting alternative conditions, such as ownership (e.g. Benson 2005), accountability or responsibility for self (Westlund 2009), while Meyers (1989) understands authenticity in terms of autonomy competence. [↑](#endnote-ref-7)
8. This kind of view is sometimes developed in recognition of the implausibility of regarding the self as fixed and for discovery. While Sartre’s view is an extreme version, more moderate versions have been developed by a number of thinkers, from Hegel through to Foucault and Rorty (Trilling 1971, Ch.2, Heyes 2006; Varga and Guignon 2014). These views emphasise processes of self-constitution that are consistent with recognising the self’s relatively fixed or unchosen aspects, rather than complete self-creation. Some of these views are broadly consistent with that we propose below, however we focus here on the extreme version as the contrast will be instructive. [↑](#endnote-ref-8)
9. Stoljar (2018, 239) argues that the conflation of autonomy and authenticity can be traced back to Frankfurt’s influential (1971) paper “Freedom of the Will and the Concept of a Person”. [↑](#endnote-ref-9)