Staying in the Loop: Relational Agency and Identity in Next-Generation DBS for Psychiatry

Sara Goering, Eran Klein, Darin D. Dougherty & Alik S. Widge

To cite this article: Sara Goering, Eran Klein, Darin D. Dougherty & Alik S. Widge (2017) Staying in the Loop: Relational Agency and Identity in Next-Generation DBS for Psychiatry, AJOB Neuroscience, 8:2, 59-70, DOI: 10.1080/21507740.2017.1320320

To link to this article: http://dx.doi.org/10.1080/21507740.2017.1320320

Published online: 16 Jun 2017.
In this article, we explore how deep brain stimulation (DBS) devices designed to “close the loop”—to automatically adjust stimulation levels based on computational algorithms—may risk taking the individual agent “out of the loop” of control in areas where (at least apparent) conscious control is a hallmark of our agency. This is of particular concern in the area of psychiatric disorders, where closed-loop DBS is attracting increasing attention as a therapy. Using a relational model of identity and agency, we consider whether DBS designed for psychiatric regulation may require special attention to agency. To do this, we draw on philosophical work on relational identity and agency, connecting it with reports from people using first-generation DBS devices for depression and obsessive-compulsive disorder. We suggest a way to extend a notion of relational agency to encompass neural devices.

Keywords: brain–computer interfaces, deep brain stimulation, depression, mental health, neuromodulation, personal identity

Deep brain stimulation (DBS) has been investigated for treatment of neuropsychiatric conditions like depression and obsessive-compulsive disorder (OCD) with mixed success (Greenberg et al. 2010; Lozano et al. 2012; Holtheimer et al. 2012; Schlaepfer et al. 2013; Morishita et al. 2014; Dougherty et al. 2015; Widge 2015a; Bergfeld et al. 2016; Widge et al. 2016c). The first generation of DBS for motor and psychiatric conditions has been “open-loop,” with DBS devices providing a constant level of stimulation. This requires battery recharging or replacement at fairly short intervals. Next-generation DBS aims to be bidirectional, allowing a single device to both record brain signals and stimulate the brain. This might mean that sensors placed within the brain could detect brain activity patterns related to psychiatric symptoms and use this information to adjust DBS stimulation. This would “close the loop” on neural processing with the device, creating a closed feedback loop. Purportedly, this would be similar to how human brains and bodies typically interact by receiving and sending signals without the use of implanted electrodes.

Closing the loop with a DBS device has a variety of potential benefits, including preserving battery charge (the system is only “on” when it needs to be, vs. stimulating when it is not needed), adjusting to match evolving symptoms, and making the system operate “seamlessly” so that the individual user need not attend to or be fully aware of its operation (in the way that diabetic individuals who use automated insulin pumps don’t need to consciously think about monitoring their insulin levels or perhaps in the way wheelchair users sometimes seem to extend their body schema to incorporate their chairs). Ideally, the smoothly functioning device would fade into the background and allow the individual to carry on with her activities.

Because we tend to identify fairly closely with our brains—more so than we identify with our endocrine systems or even with our hearts, though we clearly rely on each in significant ways—closing the loop on neural devices for treatment of psychiatric conditions also creates
concerns for our understanding of agency, or what we might think of as the individual's ability to deliberately act. An insulin pump that operates automatically is a helpful tool, not at all a threat to our agency. Similarly, a cardiac pacemaker—working well—allows an individual to carry on with the activities of life, without concern. The same is not true of neural devices that involve cognition and mood. An elevated mood or an averted thought pattern (e.g., obsession) may color an individual's perception of a situation or tilt an individual toward making one decision rather than another. As such, neural devices are intertwined with agency in complex and varied ways, and perhaps even more so when functioning automatically, outside of conscious awareness. If we do not know how the functioning of a device is influencing decisions—but we presume it is influencing them in some way—then agency is, at minimum, at issue. Insulin pumps and cardiac pacemakers can operate largely free of an understanding of agency but neural devices cannot.

In this article, we first describe how the next generation of DBS devices for treatment of psychiatric conditions may incorporate closed-loop control schemes. We then discuss how neural devices, including DBS for psychiatry, have raised concerns about identity and agency. We illustrate some of the concerns with agency by discussing results of a qualitative study with individuals with DBS systems for depression and obsessive-compulsive disorder. We then argue that the addition of closed-loop technology will magnify concerns about agency. We suggest that a relational model of identity and agency—and viewing neural devices as embodying forms of relational agency—can help address these concerns about agency.

SCIENCE OF NEXT GENERATION DBS: FROM OPEN-LOOP TO CLOSED-LOOP CONTROL

DBS was “invented” when an astute clinician noted that high-frequency electrical stimulation mimicked the effect previously achieved by lesioning the same brain area (Benabid et al. 1991). This led to rapid adoption in the treatment of Parkinson's disease (PD) and other movement disorders, essentially replacing the thalamotomy and pallidotomy as the dominant surgical intervention for tremor. The general thinking (now somewhat disfavored) was that DBS created a “virtual lesion” of the target tissue, one that could be reshaped or removed by adjusting the stimulation settings. Since internal capsule lesions (i.e., anterior capsulotomy) were known to have a beneficial effect for intractable obsessive-compulsive disorder (OCD), DBS was also attempted in the anterior limb of the internal capsule to treat intractable OCD. Nuttin and colleagues published the first successful case series in OCD (Nuttin et al. 1999), followed by open-label experiments in multiple centers worldwide (Greenberg et al. 2010) and one randomized trial (Goodman et al. 2010). Those investigators observed that DBS in the internal capsule caused mood improvement before it led to OCD improvement. That led to trials of the internal capsule target in treatment-resistant depression (TRD), where an initial open-label study was positive (Malone et al. 2009), the United States-based randomized trial (RECLAIM) was negative (Dougherty et al. 2015), and an independent European randomized trial was also positive (Bergfeld et al. 2016). Similarly, at an alternate DBS target for TRD, the subgenual cingulate gyrus, Mayberg and colleagues found a strong initial signal (Mayberg et al. 2005), and had encouraging results in open-label trials (Kennedy et al. 2011). Unfortunately, that clinical benefit was not reliably replicated in a formal randomized trial (BROADEN) (Morishita et al. 2014; Widge et al. 2016c). The Mayberg group continues to study the subgenual target and predicts that better outcomes may be possible through patient-specific neuroimaging (Riva-Posse et al. 2014). Within these mixed clinical trial results at both targets, there is evidence for efficacy beyond a mere placebo effect. If patients' DBS systems are surreptitiously or inadvertently deactivated, their symptoms return, and remain until the DBS is turned back on. This effect was the basis of a novel trial design that led to the positive European trial of internal capsule DBS (Bergfeld et al. 2016) for TRD. There is also evidence of substantial risk. Ten to 20% of patients in DBS studies commit suicide, a rate far higher than in the general population (Malone et al. 2009; Greenberg et al. 2010; Kennedy et al. 2011; Holtzheimer et al. 2012; Dougherty et al. 2015). The internal capsule target causes hypomania (a disabling syndrome of impulsivity and impaired decision making) in up to half of patients (Widge et al. 2015b). As a result, psychiatric DBS remains firmly in the realm of experimental/investigational therapies, with the only marketing approval being a humanitarian device exemption permitting its use for the very rare treatment-resistant OCD population.

The effects of DBS on the brain are difficult to study. Functional magnetic resonance imaging (fMRI), the most common human neuroscience technique, is usually not safe with brain implants in place due to the potential for MRI-induced heating and tissue damage. Electroencephalography (EEG) is safe to use, but studies to date have had conflicted results about whether there is a meaningful biomarker of DBS response (Bahramisharif et al. 2016; Widge et al. 2016c; Broadway et al. 2012). EEG is also limited to signals from the outer cerebral cortex, whereas DBS is, by definition, most strongly affecting deep brain nuclei far from the cortex. These limitations led to the development of neurostimulation platforms that can directly sense activity from invasive brain electrodes (Lo and Widge 2017). Neupace, a company targeting epilepsy, makes a device with both stimulating and long-term recording capability, though recording comes at the cost of limited ability to stimulate without draining the battery. Medtronic, a leading DBS manufacturer, has developed the Activa PC+S DBS System, which has yielded new mechanistic insights in Parkinson's disease (PD). Both systems are in use for preliminary studies of the neurophysiology of mental illness in humans. Large-animal trials are underway on
newer devices that will expand recordings to over 300 channels sensed from multiple sites in the human brain, a roughly 30-fold increase over existing DBS (Wheeler et al. 2015; Bjune et al. 2016).

As these emerging data combine with a growing body of psychiatric neuroscience data, they offer the possibility of closed (feedback)-loop therapy: sensing the brain’s electrical activity, classifying it as closer to a “healthy” or “symptomatic” state, and adjusting the location or intensity of stimulation in response. Closed-loop DBS has its roots in the field of brain–computer interfacing (BCI): algorithms that attempt to “decode” the brain and identify a patient’s intentions for control of prosthetic/assistive devices (Hochberg et al. 2012; Collinger et al. 2013; Widge et al. 2014). Such algorithms have been deployed in epilepsy and tremor control, with encouraging initial results (Berney et al. 2015; Little et al. 2013; Malekmohammadi et al. 2016). There is strong patient interest in closed-loop psychiatric therapies. In a recent qualitative study, a majority of patients who had received psychiatric DBS identified the trial-and-error nature of DBS adjustment as a major source of dissatisfaction (Klein et al. 2016). The main barrier is the identification of electrical signals that reliably correlate to symptoms and can guide closed-loop control. As recently reviewed elsewhere (Widge et al. 2016a; Widge et al. 2017), one major difficulty is that psychiatric diagnoses have high internal heterogeneity compared to the early neurological success stories. New efforts from the U.S. National Institute of Mental Health to redefine nosology may help substantially in resolving this issue (Widge et al. 2016a; Widge et al. 2017). It may also be possible to give the patient a degree of direct mental control over the neurostimulator, essentially adapting the BCI concept into a psychiatric space (Widge et al. 2014). It is not clear, however, that patients would find this desirable; in our study, many feared that this could be an inappropriate level of control (Klein et al. 2016).

IDENTITY AND AGENCY CALLED INTO QUESTION WITH OPEN-LOOP NEURAL DEVICES

DBS raises questions about human identity and agency. Open-loop DBS, for instance, has in some cases successfully treated problematic symptoms (e.g., tremor) but simultaneously created personality changes and/or affected an agent’s control over behaviors (e.g., leading to impulsivity, hypersexuality, mania, and gambling) (Glannon 2009; Gisquet 2008; Agid et al. 2006; Widge et al. 2013; Widge et al. 2015b). Schupbach et al. report that DBS recipients experience a kind of alienation or estrangement following surgery and DBS use, for example, “I feel like a robot” or “I feel like an electric doll” (Schupbach et al. 2006). Such changes do not occur in all cases and may be similar to past concerns about personality changes encountered in patients receiving neuropharmacologic treatment (Kramer 1993; Grosset et al. 2006). Some debate exists over whether they occur because of the stimulation or because of the implant itself (Johansson et al. 2014), because of user interpretations based on individual frameworks for understanding the self–brain relationship (Mecacci and Haselager 2014), or even due to difficulty adapting to a newly functioning body (Gilbert 2012). Regardless of their exact cause, such changes can make the individual feel somewhat alienated from her former self (Kraemer 2013), and raise concerns about threats to her narrative identity (Schmer 2010; Schechtman 2009). Learning to live with the neural device may be more complicated than initially envisioned.

Two points deserve consideration here. First, are these devices significantly different from pharmaceutical interventions that also work via their effects on the brain and nervous system, and create changes in personality and behavior, requiring some adjustment or adaptation? Anti-depressants, for instance, have been reported to change parts of some individual’s personality and behaviors (Kramer 1993). Extensive ethical debate in relation to pharmaceutical treatment and enhancement (e.g., Parens 1998; Elliott 2004; Greely et al. 2008; President’s Bioethics Commission 2015) has explored how pharmaceuticals may shift a user’s identity in ways that are desirable or aid in conformity, but may not feel authentic to the individual user (Bolt and Schmer 2009) or appropriate to cultural critics (e.g., Elliott 2004). Some of the same concerns arise in the context of neurotechnologies, but the devices are typically more immediate and precise in their control (e.g., stimulation levels can be adjusted with immediate impact on the individual’s felt mood) (Synofzik et al. 2012), and they often are designed to operate without the patient’s direct awareness. Taking a pill every day for depression is a reminder that one’s mood is dependent on one’s brain chemistry. If problematic effects occur, one can also stop taking the medicine. With a neural device—especially an implanted one—the stimulation may be delivered automatically, with little direct control by the patient. Importantly, the hassle of taking pills is also an opportunity for deciding not to take pills if the individual doesn’t feel right. The convenience of automaticity may thereby be in tension with this opportunity for an individual to exert control and express a sense of self.

A second question is whether these kinds of changes to an individual’s identity should really be understood as threats to identity, or simply alterations to identity that may be beneficial or detrimental (Baylis 2013; Schmer 2009). If we have a vision of personal identity as something that is relatively fixed and stable over time—as we sometimes appear to believe, as when we laud the personality of “good old reliable Granny” who never ceases to be happy or able to see the positive side of things—we might think that any pharmaceutical or neural device that changes personality or behavior is threatening to the individual’s identity. Granny just wouldn’t be herself without her sunny disposition. But perhaps that vision of an essential self is itself suspect. We may have some relatively stable features, but is our identity threatened if any of those features change? If Granny becomes less reliably inclined...
to optimism (perhaps she reels after the loss of a dear old friend, or becomes increasingly disenchanted with the difficulty of navigating a technologically sophisticated world and seems downright grumpy with its intrusions in her life), we still identify her as Granny; she's just Granny with some new attributes. She’s changed, but not so much that her identity is at risk. After all, the rest of us do some of the work of holding Granny in her identity (Lindemann 2014).

This kind of dynamic and relational view of self and identity can account for the ways that we change when, for instance, we gradually lose track of memories and relationships that once defined us, or suddenly become impaired after an accident and learn to adjust to disabled life, or, more positively, gain a new colleague who helps to reinvigorate or redefine our research priorities. In this view, we don’t have essential traits (without which we would no longer be) but rather we are the dynamic products defined by internal claims about how we understand ourselves and external allowances about who we can be. Our identities are negotiated and in flux.

Baylis notes, “Relational identity is a dialectical process aimed at achieving equilibrium—some kind of temporary and temporizing balance between self-ascription and ascription by others (others who are part of one’s familial, social, cultural, and political clusters of meaning and belonging)” (Baylis 2013, 518). Equilibrium here is not constant, but ephemeral; it has to be renegotiated in light of new “events, experiences and perceptions” (Baylis 2013, 519). New experiences might include loss of loved ones, the start of new personal relationships, the onset of disease or a new impairment, or, for our purposes here, the implantation of neural devices. Any change in one’s embodiment, relationships, or circumstances may then occasion a potential shift in identity, but such shifts are often simply part of life. We define and redefine ourselves through our experiences, sometimes struggling to be who we want to be, and at other times accepting who we are, or are allowed to be (Lindemann 2014; Parens 2014).

Given this dynamic understanding of identity, how should we think about personality and behavioral changes that sometimes accompany the use of DBS? Clausen (2008) notes the difficulty in ascertaining what would indeed constitute a threat to identity:

Identifying precise criteria for distinguishing acceptable shifts in personality from ethically problematic alterations of a person's identity remains extremely challenging. This is true because personality and personal identity are not fixed entities that can be examined and preserved. Rather than constituting a solid ground or a fixed structure, personality and personal identity are dynamic constructs that need to be actively established over and over again. (1497)

Still, perhaps we could point to the cause of the changes as morally significant. If one acquires new personality traits or behaviors over time, as a "consequence of natural personal development" (Schechtman 2009)—as might be the case when someone gets more cantankerous with age, or decides not to continue athletic pursuits given an aging body—we might think that is just part of dynamic identity. When the changes are somewhat rapidly and artificially induced by implanted electrodes (or for that matter, by medications or non-implantable brain stimulation, e.g., transcranial magnetic stimulation), we may have cause for concern. These changes can feel externally imposed when they demand a more rapid pace of integration than familiar means of volitionally induced change (e.g., psychotherapy).

Baylis (2013), however, rightly notes that rapid and artificially induced changes need not be a problem for identity so long as the individual requests and endorses the consequent change. You can go to AA (Alcoholics Anonymous) meetings and, over time, proclaim yourself an ex-alcoholic; most people won’t see this change as a threat to your identity. Alternatively, perhaps one day you can consent to the implantation of a DBS that will take away your alcoholic urges, and you will declare yourself an ex-alcoholic. Should we be worried about the latter option, simply because it changes your behavior more quickly, and not clearly as a consequence of “natural personal development”? Both changes alter your behavior, and you may endorse both changes, even though only one of them has a more immediate effect that seems, at first, to take you out of the loop. Still, if you request the DBS, then you are choosing to manage your condition—you’re using a tool available to you—and you are still generally in charge of how things go for you.

If the DBS does not help you to achieve the goal you intended, or if it makes you feel unlike yourself (as some patient reports suggest), would we thereby be warranted in saying it threatens your identity? Again, Baylis (2013) seems to suggest that we would not. It alters your identity, but alterations to your negotiated identity happen throughout life—you get a new job, you are in an accident that leads to an impairment, you lose visual acuity with age, and so on—and many of them are not in your control. Your life doesn’t always go as planned, so you have to renegotiate your identity in light of changes to your embodiment, your relationships, or your social context. Unwanted changes, then, simply force us to reimagine or shift our narratives; they are parts of our identities, rather than threats to our identities. If this reading of Baylis is fair, the choice to get a DBS to try to achieve a goal—for example, getting rid of tremor, even if that effort in turn makes your personality change somewhat—is but one, albeit new, element that becomes a part of your negotiated identity.

Still, something seems amiss here if the neural device itself causes changes to your identity that you do not endorse or desire. You presumably try to avoid events that would alter your identity in ways you do not prefer. You do your best to avoid accidents that could be identity-altering by wearing a helmet or seatbelt; you don’t apply for jobs that you think will be bad for your healthy relationships. If you knew that a DBS device might significantly
alter your personality, such that you might not really feel like yourself anymore once you had it, you would presumably be reluctant to consent to the surgery, or at least have serious reason to carefully weigh the benefits against this cost. Even acknowledging Baylis’s point about how identities are social and negotiated and not entirely in our control, we can preserve space for critiques of neural devices that alter identity in ways that individuals would not prefer, and thus constitute potential threats to identity.

Consider this case, presented by Lipsman and Glannon (2013):

A 70-year-old retired English professor who underwent bilateral DBS implantation for Parkinson’s Disease returns to the clinic with his daughter. She has grown increasingly concerned over her father’s behaviour since his surgery. He has become increasingly impulsive and emotionally labile, sometimes fluctuating from morose and depressed to euphoric. This is not in keeping with his previous mild-mannered, quiet personality. “It’s as if he’s someone else entirely,” she reports, “he’s there, and he’s our dad, but not like we’ve always known him.” When the patient was asked if he had noticed any differences, he answered “Nope … but I do feel more energetic these days.” He is otherwise neurologically intact, with his tremor and rigidity markedly improved. (466)

The man doesn’t see a change in himself, but his loved ones do, and they despair over his new, more difficult personality. Although he does not recognize the change, his family’s concerns signal a problem, and possibly one that he would have preferred to avoid. Had he been informed prior to surgery that his personality might become more difficult, more impulsive, would he have chosen to trade off his existing personality for the benefit of losing the tremor? This case demonstrates both the relevance of relational identities (for noticing when we are changing), and also the threats from undesired identity changes, even when no one questions that his basic identity—“he’s our dad”—is still intact.

To be fair, Baylis argues that one kind of threat to identity that we ought to take seriously is when the agency (not just the personality or behaviors) of the individual is shifted. She expresses a concern about DBS “insofar as it is a threat to agency—the ability to make informed and rational choices—as when a person’s actions do not flow from her intentions or beliefs but rather are the result of direct brain manipulation” (Baylis 2013, 524). Perhaps, though, personality, behaviors, and agency cannot be so precisely distinguished in practice, and thus the concerns expressed by individuals with neural devices may be, in this sense, rightly understood as worries about identity.

AGENCY AND OPEN-LOOP DBS IN PSYCHIATRY

Our group conducted a qualitative study of 14 participants in trials of open-loop DBS for treatment-resistant depression and OCD (Klein et al. 2016). Methods employed in the study have been previously described (Klein et al. 2016). Participants in a focus group and individual interviews were generally positive about their experience with DBS, but some expressed worries related to agency. In response to a question about how closed-loop systems might maintain a good feeling for users, regardless of their circumstances (e.g., being at a funeral), one participant reported that she had this very experience with her open-loop DBS, having gone to a funeral right after having had her settings programmed.

I just … couldn’t cry! … I just was surprised by my kind of reaction to the whole kind of process. And that kind of bothered me, you know? I don’t KNOW what it is—that’s the whole thing—I don’t KNOW, kind of what it is … because you DON’T know what it is—is IS it the device? Is it all the pills I’m on? Is it the circumstances I’m living in? And I can’t tell the difference … Problems concentrating or whatever, is it just … which IS it? Should I, do I need to go in and have, like a tweak? … I’ve had a setting change, I don’t know, a month or more ago, and it’s like, is THAT what’s causing it? I’ve had sleep issues, I’ve been on stuff, I know all of the to-do’s, but is it that I’m like anxious or nervous or worried? Or is it the device is maybe charged up too much? I can’t tell.

Another participant recalled his experience with DBS and wondered whether he or the device was responsible for what he was—or wasn’t—feeling.

I’ve begun to wonder what’s me and what’s the depression, and what’s the stimulator. I mean, for example, I can be fine, and then all of a sudden … and, and I might realize it later, I do something socially or interpersonally, just not right. I’ll say something that is insensitive or just misread a person entirely, say something that either makes ME look like a fool, or, hurts them, or, something along that line. I can’t really tell the difference. There are three things—there’s me, as I was, or think I was, and there’s the depression, and then there’s depression AND the device and, it, it blurs to the point where I’m not sure, frankly, who I am.

This individual describes a sense of not being sure about his own control over his actions. The “all of a sudden” language suggests something that happens to him as compared to an action. Although his body moves in the absence of obvious intentions compelling the movement, it doesn’t seem right to claim this as a mere happening (in the absence of other explanations of the action), so he reports confusion about whether he is, indeed, the de facto originator of the action. Similar concerns have been noted elsewhere. A participant in Schupbach et al. (2006) describes feeling as if under “remote control.” Kraemer (2013) has suggested that such feelings can lead to a sense of alienation. Put simply, basic open-loop DBS in psychiatry can raise important concerns about agency.

1. This raises a further question about whether feelings of alienation due to the experience of uncertain agency are harmful. Elsewhere, Kraemer argues that feelings of alienation are harmful because they signal inauthenticity (Kraemer 2011). Gilbert (2015a; 2015b), on the other hand, locates the harm of alienation in the felt powerlessness that can accompany a lost (or threatened) sense of agency. We owe this helpful point to a reviewer.
To better understand what to make of these agency concerns, let’s look briefly at what we mean by action and by agency in general. Commonsense thinking about action suggests that when we perform actions, we presume that we are in control of what our bodies do, even if the eventual consequences of our actions are, at least in part, out of our control. You can reach for a cup and pick it up—the act of an agent—and typically you know what the result will be. If the wind picks up suddenly and blows the cup away, or if someone bumps the table, moving it aside, your hand may not reach the cup. Nonetheless, you have acted as an agent, attempting to reach a cup. Of course, you might also act as an agent, reaching for the cup unconsciously, for instance, if you are unaware of your habit of clearing the table while talking after dinner. You wouldn’t say that you consciously intended to take the cup in that case, but you nonetheless did reach for it, and it’s clear that you acted. If we put strings around your arm and hand and pull on them to try to bring your hand to the cup, though, you are not an agent who reaches for a cup; you are simply a pawn in our cup-tapping ploy. In the former cases, you act (first with intention, then not intentionally); in the final case, something happens to your body. The vast literature on action theory is filled with nuanced analyses of action. For our purposes, it makes sense simply to distinguish between actions you take, and things that merely happen to your body. Even so, on the action side, we can recognize differing levels of action (unconscious, purposeful or goal-directed, intentional, and autonomous) (Wilson and Shpall 2016).

In respect to agency, we can think of it as the capacity for or the exercise of the ability to act or refrain from acting. Entities have agency when they interact with others (agents or objects), and typically, agency is restricted to beings that demonstrate some intentionality (Schlosser 2015). Thus, prototypical cases of agency involve someone intending an action and doing it. As with the previous example, though, we might say that your clearing the cup from the table while talking after dinner consists of your exercise of agency even if you do not consciously intend (are unaware of your acting) to clear the table. You still undertake a goal-directed action.

Kellmeyer and colleagues helpfully simplify what we mean by an autonomous agent, suggesting that one must “(i) interact with objects or other agents; (ii) possess reliable heuristic and decision-making capacity; (iii) be the de facto originator of particular actions; and (iv) act in accord with her (its) beliefs” (Kellmeyer et al. 2016, 624). They recognize that not all agents will fully meet condition (iii), because, for instance, not everyone can initiate actions on their own (e.g., someone in a minimally conscious state may rely on a neural stimulating device to initiate or achieve an action) (see Fins 2015). Further simplifying, they suggest that “personal autonomy arises from the subject’s experience of congruence of motive and action, which gives rise to the feeling of individual agency” (Kellmeyer et al. 2016, 624).

Based on this understanding of agency, we can see how open-loop neural devices, such as traditional DBS, might complicate our sense of agency. If you get an implanted device and then experience changes in how you relate to others or the environment that seem out of character, an incongruence between motive and action may emerge. This is the phenomenon that some participants in our study seem to be describing.

VOLITIONAL AND CLOSED-LOOP CONTROL AND AGENCY

As described in the preceding, brain–computer interface technology is now being combined with DBS to treat neurological and psychiatric conditions. BCI systems—whether wearable or implantable—record and decode electrical brain activity and translate it into signals that can then be used to control a variety of targets: a robotic arm, a computer cursor, a prosthetic arm, a wheelchair, and so on. Current research efforts are targeted toward using information extracted from implanted electrodes to modulate stimulation through DBS-like electrodes. While DBS has been explored for a variety of conditions, including movement disorders, epilepsy, obesity, Tourette’s syndrome, addiction, and anorexia nervosa, interest has also turned to treatment of psychiatric symptoms, such as those found in depression and obsessive compulsive disorder. The promise of combining DBS with BCI is to create closed-loop systems: where information is gleaned from the brain to allow more effective and timely intervening in the brain (Klein and Nam 2016). In light of failed trails of open-loop DBS for depression, there is a growing belief that adding closed-loop technology to DBS represents a promising path forward (Widge and Dougherty 2015; Widge et al. 2017). One rationale for closed-loop technology is that it could enhance a sense of agency (Goering 2015; Gilbert and Cook 2015; Glannon and Ineichen 2016).

The promise of agency enhancement can be seen most easily in closed-loop BCI for motor control. If the device allows a person to control a robotic arm with her motor intentions, and then returns sensory feedback (e.g., from a glove covered with sensors) to the user’s brain, we can imagine the user’s overall experience being more complete. The return of sensory feedback to the brain would allow the individual to orient the arm with more robust input, and would give the user the sensation of feeling whatever the arm touches. Coordinating all these signals might be difficult, and given the ways that our brains typically function, neural engineers are interested in using devices with machine learning to allow for finer control and smoother integration.

As an example, we might think of how a baby learns to move her arms and to grasp an object with her hand. It takes practice. Orientation, depth perception, grip, and so on all require significant trial and error. A person using a BCI has to learn how to imagine the trigger for the movement, to concentrate to make the signal clear, and then

---

46 ajob Neuroscience
The idea that a closed-loop system could generally support agency is also a rationale supporting use of BCI seizure-control systems. Gilbert (2015) and Hoppe et al. (2015) both suggest that patient preferences often depend on how much the different systems interfere with or support the user’s daily activities (whether they enhance or detract from the user’s ease). Perhaps users would be willing to sacrifice some level of local control over the device in favor of having greater capacity to do what they want more broadly. Additional empirical work in this area will be important. Understanding what would count as reasonable trade-offs for the user must be part of the design process, and it might seem that a system that offers options to the user would be preferable. But the upshot is that on balance such closed-loop systems could enhance an individual’s agency.

However, there are ways in which closed-loop systems would seem to complicate, if not undermine, agency (Klein et al. 2015). Here again, we might imagine that the user who reaches for a cup, intending to crush it for effect (rather than pick it up for use), might find herself fighting the device, which interprets her intention as reach and grasp to pick up. Perhaps this is merely a technicality—an engineering specification that could be improved over time, such that finer degrees of control and more nuanced identification of intentions would be possible. But even if they are much improved, the agent using the device may still sometimes 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.

The prospect of undermined agency would seem to be even more complicated in applying closed-loop technology to DBS for psychiatry, where a disconnect can open between feelings and responsiveness to circumstances. Returning to the example of attending a loved one’s funeral, the individual’s capacity to regulate her mood—perhaps not a fully conscious capacity—for the occasion might seem off. Interestingly, depression itself seems not to allow this normal range of emotional responsiveness. Is closing the loop on a device to treat depression a way of changing the baseline (positively), but not the capacity for responsiveness? One might well prefer the more positive frame of mind if forced to choose, but recognize the loss as significant nonetheless, and potentially agency-undermining. One participant in our study noted, “If that’s the situation, this [hypothetical] person that you’re describing, that she goes to a funeral and she doesn’t have the same sadness that she’d [normally] notice for her grandma, then, if that’s a side effect, I would think that, you know, people would be kind of (chuckles lightly) be okay with that side effect.” Still, even if you don’t feel sad at a funeral of a loved one, will you always know to attribute that to your DBS, or might you start to question your feelings for the deceased, and the significance of your relationships? Iterative cycles of feedback outside conscious awareness may lead to a drift in responsiveness to the environment or to other people that goes unnoticed by me until an extreme circumstance (e.g., a funeral) makes it evident. Hence, Kraemer’s concern with alienation resurfaces.

One approach to this agency problem might be to opt instead for volitional control systems or closed-loop systems with discrete opportunities for volitional input. Such forms of volitional input would go beyond giving individuals the ability to shut off a device in an emergency or choose between a few preset stimulation settings, as some DBS devices currently allow (Widge et al. 2014). As Kellmeyer and colleagues note in discussing fully closed-loop BCI systems (with no volitional control) for seizure detection and treatment, this option “has the advantage of convenience—the machine operates independently to reduce the risks of seizure—but the disadvantage of leaving the patient out of decisions regarding whether to act in response to an increase in seizure risk” (Kellmeyer et al. 2016, 626). They suggest that the alternative—leaving the subject in the loop—might be accomplished by an external device (e.g., a traffic light-like warning system, perhaps worn on a watch or other external device) that could signal the approach of a seizure, to allow the individual to decide how to respond (to stop riding a bike, take medicine or not, etc.). An analogous example from work in motor control is a closed-loop DBS system developed for essential tremor, in which neural patterns portending tremor onset are detected and in turn trigger increased stimulation to abort tremor (Brown et al. 2016). One can imagine the agent staying “in the loop” by being notified via watch or other device when the stimulation would begin or be increased, and/or by being able to reject the stimulation, whether by pressing a button or by removing the watch. Stimulation, in these instances, could be offered to the user as an option to help control the tremors, and the user could decide case-by-case whether she prefers to make use of the assistance.

Still, it is clear that except in limited contexts, the “completely closed loop” versus “partially closed loop” dichotomy is a false choice. A closed-loop system almost always involves some volitional input. A closed-loop system is influenced by acts of will, whether by engaging in physical activity (e.g., running), imbibing a glass of wine, indulging in a romantic movie, reminiscing about one’s youth, or attending a funeral. Our volitional input into closed-loop systems may lack precision (how will a good run affect your mood today?), but, similarly, it is doubtful that effects of device-mediated volitional control will have precisely intended effects on closed-loop systems.
either. On the other side, in the normal course of daily life our cognition, affect, and mobility are governed, in large part, by biologically ingrained closed-loop systems. Our self-conscious acts of will—whether handing a glass of wine to a dinner guest or jumping into a conversation—are always conducted against a background of subconscious cognitive, affective, and bodily processes governed by these closed-loop systems. The complex interplay of closed-loop and volitional control is evident in the limited control that individuals with BCI can currently exert over robotic arms (Wodlinger et al. 2015). Adding degrees of freedom and easier and more fluid control of robotic movement will require, at least in part, a better scientific understanding of where volitional control leaves off and closed-loop control begins (and vice versa) (Widge et al. 2010).

The provision of volitional regulator switches for closed-loop systems does not solve the more fundamental problem that patients may not always recognize an agency loss as such. The individual who over time drifts into compulsive gambling (and identifies with this avocation) or into hypomania may not feel that she has a loss of agency at all, but is simply acting as she wants. She may mistakenly believe that she has agency, or she—and those around her—may simply be uncertain. As we have already encountered, Baylis (2013) gives voice to a species of this concern when she notes that neural devices are a threat to agency when they involve a kind of “brain manipulation.” Changes to personality not easily articulated by the individual (i.e., the story the individual tells won’t easily accommodate them) and induced in a way that bypasses agency are thus problematic. They represent a substantive threat to the individual. But what is problematic about an appeal to brain manipulation is that it lacks identifiable boundaries. Since we lack brain proprioception (Haselager 2013), how are we to know when brain manipulation is at play and not a natural personality drift due to a closed-loop system? The appeal to “brain manipulation” renames the problem of agency but offers little help in identifying or addressing it.

**RELATIONAL AGENCY AND AT-RISK AGENCY**

If we think about the agent in a relational way, we may have a way to address some of the concern surrounding agency and closed-loop technology. Acting autonomously may mean that you develop your motives through dialogue with others, that your motives are not entirely your own, and that even your actions may be shared in some sense. In a simple case, if you aim to clean the gutters out, and someone holds the ladder for you, you are supported in your agency related to gutter cleaning. In still more directly collective agency, say a shared project of going for a walk together, a loved one may not simply support you, but also share a collective intention that will be thwarted if one of you fails in your action (Gilbert 1990). Linda Barclay (2000) suggests that in relational agency, we often have a “shared task” that we do “in concert and conversation with others” (Barclay 2000, 68). Given relational agency, an individual’s agency (connecting her motives and actions) can be supported through and entwined with a variety of external others. This is illustrated by another participant in our study who describes ways in which her agency is bound up with her family. She says: “So, there’s people in my family that (sighs) . . . sometimes question, you know, how much of it is me anymore and how much of me is, you know, being programmed. So that’s been a, that’s a hard thing to deal with sometimes, but, you know, like I said, I’ll take it over what I had.” Here the individual highlights the fact that her family members question whether she is in charge of her actions, or whether she is “being programmed” by the device to act. What is notable in this quote is the way in which others are involved in helping us identify and evaluate the effect that implanted devices can have on agency.

If we understand ourselves as relational agents, who rely in significant ways on others to help form and hold us in our identities, and to help us exercise our agency and autonomy, we can see identifying and evaluating instances of “brain manipulation” as merely part of holding us in our identities. We expect our loved ones to point out when we are at risk of manipulation, as we may not recognize manipulation until we are in too deep. This is familiar enough in our everyday relationships. If your new love interest turns out to be controlling and isolating, but you can’t see it well—he’s charismatic and focused on you, and you see his actions only as evidence of his love—you may need the other people who know you well to help make you aware of how this new relationship is changing you for the worse. They can help you reclaim your agency in the face of this controlling other. You can certainly push back against a friend’s admonition and reject the label of manipulation—you may indeed know what you are getting yourself into—but the friend’s intervention is valuable for making you take this extra step. The extra set of eyes and willingness to exercise tough love do not usurp your agency, or guarantee it. At best, and of critical importance, it supports and nurtures your agency. Put another way, a relational account of agency allows us to deal with “brain manipulation”-type concerns by providing a check on instances where agency is at risk.

A more fundamental question may be whether we want to frame the concern about agency in terms of “brain manipulation” at all. There is a way in which the device is just doing what we want it to do. As discussed earlier, we want the device to become unobtrusive, to seamlessly integrate into our daily lives. In part this is because there are cognitive, social, and emotional burdens that we experience when tools fail to recede into the background of our practice. If you are given a new technology to make your life easier—from an electric mixer to a smart prosthetic—but getting it to work...
takes great effort, and the learning curve is steep, then you are more likely to resort to simpler tools that are reliable and easy to use. Further, these are devices that are freely chosen by patients for the potential benefits that they purport to offer, some of which are benefits, as we have discussed, that redound to agency itself. By characterizing these devices as inherently involving forms of brain manipulation (or similar), we may undermine earnest patient efforts to incorporate these devices into their “body schemas” (Heersmink 2013) and into “structures of decision-making and acting” (Clausen 2008). The language of manipulation signifies devices as objects to be feared and kept at a healthy distance.

Rejecting the label of brain manipulation does not obviate the underlying concern about agency, but it may leave open a door to another way to address this. If we take a more neutral stance on neural devices, we see that it is possible—and maybe advantageous—to extend a relational account of agency to include devices themselves. Neural devices can be tools that support agency, not wholly unlike how friends, family, and others support agency. As a closed-loop device integrates more fully into daily life, we can imagine such a device stepping in to support the user in achieving her aims. This isn’t to say that the device is a friend, but rather that we all rely on other people and things (including devices like cell phones, computers, etc.) to support our activities and actions. Here again, we retain the capacity to push back against these others who support us—we can refuse to listen to friends, argue with family members, and silence our buzzing smartphones. Some might argue that the devices themselves could have a kind of agency. Our point is only that the devices can serve to support an individual’s agency in significant ways, not wholly unlike the kinds of support the individual receives from other agents.

Neural prosthetic devices have the potential to affect our identities, perhaps supporting us so that we can be who we want to be, but perhaps also changing us in ways we wouldn’t want, even to the point of challenging our identity and/or agency. In this they are not entirely different from other people. We may not always realize when we change—it may take a friend or family member to notice that we have become despondent, upset, or anxious. Relational agency relies on certain features of connection and intertwined action that are most conspicuous in our relationships with friends and family. But just because we notice them in relation to people, and in fact see them most clearly in respect to people, does not mean that we can’t imagine other ways of supporting agency more broadly (i.e., in the construction of adaptive neural devices). If we view neural devices as part of our relational agency, then we will still want the user (perhaps in conjunction with her family and friends) to have the capacity to maintain authority over the device, and perhaps—when necessary—to reclaim her authority from the device.

So too, with neural devices, we will need to rely on friends and family to help us identify how the devices may be changing us for the worse rather than supporting us in ways that enhance our agency (Goering 2014). The Asimov laws of robotics may apply here: A robot shouldn’t control a human being, but rather must obey the orders given it by human beings, except where such orders would injure or cause a human to come to harm. The human being who should be giving the orders to a neural device ought to be the individual in whom it is implanted, unless she has reason to think that unwise. As with robots, neural devices can help to aid and support us, but should do so only in ways that allow us to maintain our relational agency.

CONCLUSION

Next-generation neural devices that incorporate closed-loop control schemes will challenge existing notions and norms of agency. Emerging neural technology may be threatening to agency in some ways—we may feel at times “out of the loop” and as if we have lost moment-to-moment control of our emotions or bodily movements. Yet what we have argued here is that just as our friends may be supporting or controlling of our agency, so too can neural devices be enhancing or threatening to our agency. We can use our understanding of the ways in which our everyday agency is relational to better recognize when a device is supportive, and when it is undermining (or perhaps a bit of each). The example of closed-loop DBS for depression and obsessive-compulsive disorder illustrates this complexity. Closed-loop neural devices—in psychiatry and beyond—may be developed to be the kinds of devices that support our agency, much as we rely on our friends to support our agency.

FUNDING

This study was supported by a grant from the National Science Foundation.

REFERENCES


2. One worry to consider here is that our relationships with other people may change, given the presence of a neural prosthetic device. For instance, some participants in Klein and colleagues’ (2016) focus group noted that having the implant affected how their family members perceived them and their concerns: “I think there’s this expectation which I find with my family, that, if you’re not feeling well, ‘go flip a switch and you’ll be better’ or ‘Go get your settings adjusted!’ That’s [my family’s] answer to everything!”


