PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Opinion piece

Cite this article: Roskies AL. 2015 Agency and intervention. *Phil. Trans. R. Soc. B* **370**: 20140215.
http://dx.doi.org/10.1098/rstb.2014.0215

Accepted: 3 March 2015

One contribution of 15 to a theme issue 'Controlling brain activity to alter perception, behaviour and society'.

Subject Areas:
neuroscience

Keywords:
deep brain stimulation, agency, intervention, ethics, rationality, folk psychology

Author for correspondence:
Adina L. Roskies
e-mail: adina.roskies@dartmouth.edu

Agency and intervention

Adina L. Roskies

Department of Philosophy, Dartmouth College, Thornton Hall, Hanover, NH 03755, USA

Novel ways to intervene on brain function raise questions about agency and responsibility. Here, I discuss whether direct brain interventions, and in particular, deep brain stimulation, pose a threat to agency in individual cases, or to our general conceptualization of what it is to be a responsible agent. While I do not currently see evidence that these interventions constitute a global challenge to our concept of agency, they do have the potential to diminish agency in individuals. I consider whether the lack of evidence for a global challenge ratifies our folk conceptions, or is a necessary consequence of them. In closing, I propose that our theoretical understanding of agency and our therapeutic approaches could be improved with a more nuanced, multidimensional view of agency.

1. Introduction

The past several decades have witnessed the development of several novel ways to intervene in brain function. Indirect manipulations, such as those that affect brain function systemically with psychopharmacological substances, have been around for quite some time, and our ability to intervene and control illness pharmacologically has continued to improve. More recently, scientists and medical practitioners have developed ways to directly influence brain function through focal electrical interventions, and some new tools show promise in treating diseases that have thus far been resistant to treatment by other means.[1] Although direct brain interventions are not new, the toolbox we have now is significantly more sophisticated than the one we had just a decade or two ago, and it promises to become ever more powerful. As with any clinical therapeutic method, the promise of these interventions to treat disease must be weighed against the risks and the costs. In this paper, I aim to provide a framework for thinking about the impact these interventions may have on the agency of a patient, though I will not do the weighing. This paper is primarily theoretical, raising potential issues and calling for additional information and collaboration among professionals who may not have had occasion to work together prior to the development of these techniques for use in humans.

What or who are agents? Most basically, an agent is something that acts on the world. A naive perspective on agency might view agency as a binary property: either something is an agent or is not. Perhaps Descartes could be accused of having such a view, since he thought of non-human animals as soulless automatons, and thus as non-agents. Few today embrace this view and are likely to opt for a more nuanced view in which agency can vary on a spectrum: simple organisms that are more than just stimulus–response machines would have some agency, in that they can act decoupled from direct stimulation; we may attribute still more agency to animals that can form plans or act on the basis of memories, and more still to higher animals with complex social interactions, though they will fall short of being morally accountable for their actions. Humans have more still, for they are self-reflective, and we can ask questions about autonomy and authenticity, and hold them responsible for what they do. Among humans, levels of agency may still vary: intuitively, adults are more agentive than infants because they possess capacities that infants do not; those not in the grip of addictions or delusions may be more fully agentive than those with psychiatric disease or drug habits. However, even viewing agency on a spectrum is, I think, too simplistic. I will suggest that a more complex, multidimensional view of agency is called for, and that adequate consideration of effects of both disease and treatment of many neuropsychiatric illnesses require a more finely articulated notion of

THE ROYAL SOCIETY PUBLISHING

agency. In this paper, I will focus on agency in its most full realization: an agent is an autonomous actor persisting through time, who can be held responsible (causally, and potentially legally and morally) for his/her actions. I deviate somewhat from common usage here, in that many creatures that we commonly think of as full agents, such as dogs, are not, on this view.

How might brain interventions affect agency? How might they affect the way we conceptualize agency? I begin with a brief survey of some of the brain interventions currently available, and a description of those on the near horizon. For the majority of the paper, I will focus on one widely used technique for treating psychiatric and neurological illnesses: deep brain stimulation (DBS). I then discuss the impact that DBS can have on agency, and consider two different ways in which it might be understood to threaten agency: by threatening our conception of agency, or by threatening agency in individual cases. These are related, since, I will argue, properly understanding the way in which agency is affected in individual cases might, at least in theory, require a reworking of our general conception of agency. My conclusion is provisionally a deflationary one: I do not now see brain intervention techniques seriously threatening our ordinary conceptions of agency. I do recognize that these techniques have the ability, when applied in individual cases, both to restore or augment agency and to diminish or compromise it, and to do so in a variety of ways. I conclude by recommending the development of a new taxonomy of agency and suggest some ways to proceed.

2. Current interventions and effects; future interventions

In this paper, I do not mean to give an exhaustive account of current interventions, but aim only to give a sense of some of the methods that have shown considerable promise. I begin with transcranial magnetic stimulation (TMS), a non-invasive technique that induces or inhibits neural activity and is frequently used in the laboratory for research purposes. The other techniques I mention are invasive, limiting their use to clinical settings.

(a) Transcranial magnetic stimulation

Transcranial magnetic stimulation is a technique that can non-invasively activate or disrupt local brain networks. An electromagnet is placed close to the scalp, and a pulsed magnetic field is delivered, inducing electrical activity in underlying neurons in the cerebral cortex. TMS can activate or inhibit a region of cortex depending on the intensity, frequency of pulses and number of pulses. TMS can non-invasively provide evidence for the causal involvement of a brain area in a particular task, by creating a temporary 'virtual lesion' in normal subjects. One recent study used TMS to show that people's judgements of culpability depend upon activity in the rTPJ (right temperoparietal junction) [1]. Repetitive TMS (rTMS) can have longer-lasting effects than single pulses. TMS is a spatially crude technique, allowing only rough estimates of where the maximal effect of stimulation will be located in the underlying cortex. It tends to be used more for research than for clinical purposes.

(b) Electrocorticography

Unlike TMS, electrocorticography (ECog) is invasive. It involves placing a grid of electrodes on the brain surface, and thus can only be used in the context of brain surgery. Although ECog is often used to record electrical signals from the brain surface, for example to measure ictal episodes in epilepsy, the electrodes can also be used to stimulate underlying tissue. In a recent study, ECog was used to show that stimulation of a face-sensitive cortical area, the fusiform face area (FFA), led to a visual morphing or distortion of perceived faces, but not other objects [2]. The use of ECog has provided insight into structure and function of multiple brain areas, but it is used only in clinical settings.

(c) Optogenetic techniques

Although not currently used in humans, optogenetic techniques are poised to greatly enhance our ability to treat human diseases by allowing us to precisely modulate the activity of specific cell types. Optogenetics targets specific cell populations, using retroviruses to genetically manipulate these cells to express light-sensitive proteins. Once these proteins reside in the cell membrane, the cells can be activated and/or inhibited merely by exposing them to light of specific frequencies. This allows for unprecedented precision in spatial and temporal control over neural activity. Optogenetics have enabled researchers to control behaviours in animals from fruit flies and worms to rats. The technique is now being pioneered in monkeys, and when adapted for human clinical use it is likely to greatly improve the efficacy and specificity of the treatment of a number of neurological diseases [3,4].

(d) Deep brain stimulation

DBS is a highly invasive, but also highly effective clinical technique that is increasingly being used for treatment-resistant neurological and psychiatric diseases. In DBS, a stimulating electrode is surgically implanted deep in the brain. The electrode is connected to a power supply, and controls that are implanted under the skin in the chest area can be wirelessly activated or deactivated. When activated, electrical pulses are sent through the stimulating electrode. These pulses electrically modulate the activity of neurons in the tissue surrounding the electrode. Electrode placement and stimulation parameters vary for different illnesses.

Of all the techniques discussed here, DBS is the most highly developed for use in clinical settings. DBS was FDA approved for treatment for Parkinson's disease in 1997 and was granted a limited approval for obsessive compulsive disorder (OCD) in 2009 [5]. It is being used in clinical trials for, among other things, treatment-resistant depression (TRD), Tourette's syndrome, addiction, anorexia and chronic pain [6]. More than 100 000 patients have already been implanted with stimulating electrodes for the treatment of Parkinson's disease and dystonia alone (see http://professional.medtronic.com/pt/neuro/dbs-md/prod/index.htm#.VGpriqVOiRo).

Despite its demonstrated clinical efficacy, the mechanisms by which DBS works are not well understood [6,7]. Even relatively basic questions, such as whether affected neurons are activated or suppressed by stimulation, and which cell populations are affected, are currently unresolved [7,8]. Early models treated DBS as creating functional lesions, but these have been overturned in favour of models that implicate DBS in regulation of interacting dynamic oscillatory networks [8]. Exactly how this rhythmic activity affects neural systems remains to be elucidated. Scientists are likewise uncertain about whether or when long-term stimulation leads to changes in the underlying neural tissue, and some

are now trying to exploit principles of brain plasticity to use DBS to institute permanent changes in neural networks.

In general, DBS patients are treated symptomatically: the frequency and amplitude of the stimulating pulses are adjusted and the resulting behavioural effects are monitored, and parameters tweaked again. In many cases of neurological or psychological illness, we lack a clear understanding of what cellular or network properties result in disease symptoms, or what changes would be effective in alleviating them. In some cases, such as in Parkinson's disease, stimulation sites are chosen on the basis of work in animal models. However, with many psychological diseases good animal models are unavailable. Researchers have thus had to make educated guesses about intervention sites. In sum, DBS treatment is more empirically than theoretically grounded, and despite the sometimes-remarkable success of the treatment, scientists do not really understand its mechanism of action.[2]

Despite this, the efficacy of DBS in individual cases can be impressive. When DBS is effective in the treatment of, for example, Parkinson's disease, the results are often dramatic. A person contorted, unable to articulate speech fluently, or to walk or prevent the tremor in his limbs from dominating his movements, can be immediately transformed into a person who appears to be practically asymptomatic, who can speak unencumbered, walk steadily and manipulate objects with dexterity. Then, with the flick of a switch, when the stimulator is turned off, the tortured figure of the Parkinson's patient reappears. If the reader has not yet seen such a transformation, then I recommend viewing https://www.youtube.com/watch?v=uBh2LxTW0s0. Similar results have been seen with TRD. DBS clearly is life-changing for numerous people. However, although many patients experience significant relief from debilitating symptoms that are recalcitrant to other, less-invasive types of interventions, that relief sometimes comes at a cost: unwanted side effects can compromise a patient's quality of life in ways that seemingly undermine agency. Stimulation-induced side effects include sensory–motor effects such as paresthesia, dyskinesia, gait ataxia, slurred speech. Personality and mood changes also occur, as when stimulation leads to maladaptive behaviors such as pathological gambling, hypersexuality and mania. Typically, these side effects are acute and disappear when the stimulation ceases.

In the rest of this paper, I draw my examples from cases of DBS, though in principle my discussion applies to any brain intervention technique.

3. The folk conception of agency

The idea that medical interventions into people's brain function could have ethically troubling effects is not new. Surgical interventions to change people's affect and personality are always ethically fraught; early psychosurgery is a forceful illustration of this fact. In an early paper defining neuroethics, I postulated that new neuroscientific interventions could have the potential to change personality and even personal identity [11], and that this would raise ethical questions that other kinds of medical interventions would not. To date, the majority of philosophical work discussing DBS has been stimulated by the observation that the interventions can lead to changes in personality and mood, and it focuses on the worry that these treatments can disrupt or alter personal identity. Whether or not personal identity is threatened depends significantly on what one means by personal identity (e.g. [12,13]).[3] While few thinkers have suggested that DBS can actually change the numerical identity of the person by sufficiently disrupting psychological continuity (but see [14]), many have recognized another type of identity that may be at issue: the person's own self-identification or self-alienation. Schechtman [15] explored the question of when various actions and mental states are properly attributable to a person, and various constructs, such as 'narrative identity', and Witt *et al.*'s [16] 'Individual identity', are meant to track whether or not a person sees herself as the same self over time, feels alienated from her pre-treatment self or feels herself to be her own authentic self, as well as whether we should characterize that person as being one and the same self. There is much more room for interventions to have an impact on narrative or individual identity than on the question of numerical identity, as the clinical case literature fairly explicitly demonstrates that this kind of alienation is not uncommon [17]. Even if questions of identity are not fundamental to the concept of agency, they are clearly relevant to our understanding of what it is to be an agent of the sort we are, insofar as our self-conception affects and is affected by our actions. But since these issues have received thorough treatment elsewhere (e.g. [5,12,13,18]), I focus instead on other, perhaps more foundational, aspects of agency.

The question of free will or self-authorship is distinct from the question of identity. An agent is the author of her actions when she has the capacity to make self-generated and free decisions, and to act according to reasons that she herself endorses. Thus, we might ask of brain interventions: Do these treatments interfere with an agent's ability to act voluntarily or freely? Do they undermine agency by forcing or compelling decisions that normally are autonomous? Do they threaten responsibility? Research into the nature of decision and voluntary action is ongoing in neuroscience, and these are topics of central interest to philosophers of mind and ethicists. Despite this, the question of whether the volitional aspect of agency is threatened by brain interventions has been less frequently addressed in the neuroethics literature, and it is to this question that I now turn.

Agency is perhaps best conceptualized in terms of a constellation of capacities that full agency presupposes. This 'capacitarian' view of agency is broadly congruent with capacitarian views of free will, and moral and legal responsibility. Briefly, the idea is that to be an agent, one must possess a set of capacities that are the same or closely related to the capacities that make possible voluntary action, self-control and moral responsibility.

Capacitarians about free will are compatibilists that hold that what it is to have free will is to have a set of capacities sufficient for rational agency. Capacitarians about moral and legal responsibility similarly hold that the criteria for being held morally or legally responsible can be described in terms of capacities of the agents. Although the capacities and criteria that each view targets may be slightly different, there is broad overlap between them, and between the most fundamental capacities for being a full agent.[4] Vincent nicely characterizes this commonsensical take on the concept when she writes, 'In lay contexts responsibility is often thought to require such things as the ability to perceive the world without delusion, to think clearly and rationally, to guide our actions by the light of our judgments, and to resist acting on mere impulse'. [19, p. 22].

This characterization can be augmented and operationalized by asking a series of questions:

— Does the person form an intention to act?
— If so, is the intention to act formed in the normal way by the agent or brutely caused by the intervention?
— Does the agent act for reasons?
— Are the beliefs contributing to the reasons rationally formed?
— Are the desires ones that the agent endorses or identifies with?
— Does the agent retain the ability to rationally assess his intentions?
— Can the agent inhibit his responses?
— Does the agent experience his action as voluntary?

Positive answers to these questions are indicative of capacities necessary for full-fledged agency. To the extent that these answers are qualified or negative, we can say that agency is diminished. To preview what is to come, if we can more finely articulate this operationalization, and better capture the ways in which these capacities can be impaired, we will achieve a better understanding of the many ways in which agency can be threatened, whether by disease or by therapeutic intervention.

4. Interventions and agency

Since agency is closely linked to our conception of free will it follows that interventions into the mechanisms that mediate the internal commerce between a person's perceptions and his subsequent decisions and actions may deprive that person of or undermine his agency. The philosophical literature is familiar with the problem of intervention with respect to the problems of free will and responsibility, most notably in Frankfurt's scenario of the neuroscientist that ensures that a person decide to act in a particular way by predicting and influencing his brain activity [20–22].

An old philosophical chestnut posits an (for colour, evil) neuroscientist, who has the ability to read someone's neural data and intervene in their brain events (call that person Harry) just in case he is about to decide to take an action that the neuroscientist disprefers. In these so-called Frankfurt cases, one's intuition is supposed to be that if the intervention occurs, Harry is not responsible for his action, because he is not really a free agent: he is forced to act as he does by the intervention [21]. However, one is also supposed to have the intuition that where the neuroscientist does not intervene, because Harry chooses to do what the neuroscientist wants him to do, Harry is the agent and is responsible for his action, even if he could not have done other than he did. One might be tempted to conclude from this example that it is the fact of the intervention that deprives a person of free will and absolves him of responsibility. That would be a mistake.[5]

Imagine, for example, that the neuroscientist is not evil, and merely wants to ensure that Harry is able to carry out his decision. Unfortunately for Harry, the neural circuit that would allow him to implement his willing is damaged: there is a disconnection that prevents his will from issuing in action. Now the helpful neuroscientist intervenes in the following manner: if he detects a neural signal at the input end of the disconnection site, he is to send an electrical pulse to the output side of the disconnection, effectively reuniting the normal circuit so that Harry can exert his agency. The intuition I have here is that this intervention is benign—it restores agency and does not reduce Harry's responsibility. If so, then not all interventions are threats to agency.[6] We know that DBS interventions can often restore agential properties eroded by disease, although as some case studies relate, they may at the same time compromise other aspects of agency. Rather than focusing on whether there is or is not a brain intervention, I suggest that we should assess threats to agency by the effects an intervention has on agential capacities.[7] The way in which we assess responsibility in disease is the same way in which we assess it in cases of interventions.

I am not disputing that brain interventions can threaten agency. But they do not do so merely because they are interventions. Whether they are a threat, and if so, how and why, are matters that depend on the effects of the intervention: how it alters the capacities of the person.

5. Two global skeptical challenges to agency

Before discussing how DBS interventions influence individual patients' capacities, I want to raise two ways in which one might think they pose a general threat to our concept of agency. The first I address in order to promptly dismiss it. It is conceivable that people will view the efficacy of DBS as evidence for all people's inability to act voluntarily, because it shows that they are mere physical objects. This reaction implicitly assumes the following kind of argument:

(1) DBS demonstrates that electrical interventions on the brain change agentive behaviour;
(2) electrical interventions are physical interventions;
(3) thus, DBS demonstrates that agentive behaviour is merely a physical phenomenon;
(4) physicality is inconsistent with agency; and
(5) therefore, agency is an illusion.

This argument rests on a dualistic view of agency that is surely mistaken. Moreover, the argument is not specific to DBS, but can be raised for almost any medical intervention on human behaviour. Suffice it to say that DBS may be a dramatic illustration of our physicality, but it does not therefore automatically undermine agency. The kinds of arguments I am interested in are specific to the type of interventions we see in DBS and are local challenges of the sort I will shortly describe.

A second global challenge is more difficult to counter. Might the unforeseen side effects of DBS undermine our commonsense theory of human agency? I will call this the challenge from eliminativism, for, just as eliminative materialism predicts that folk psychology will be supplanted by a more powerful neuroscience [25], I suspect that our ordinary folk psychological constructs regarding agency may be challenged by a careful analysis of effects of DBS. According to our folk psychology, human agents are partially rational actors who form intentions to act in order to fulfil their goals and desires on the basis of practical reasoning constrained by their beliefs about how the world is, their values, and their expectations of the consequences of their actions. Because the brain is a dynamical, highly complex conglomeration of neural networks, and the DBS electrodes are typically situated in the midst of highly interconnected subcortical regions and tend to stimulate a whole region of tissue, it is possible that our static and rather rough concepts

will not easily describe the kinds of effects seen in DBS. Unexpected effects or constellations of effects of DBS on patients' mental states could, in principle, take on patterns that fail to easily fit without commonsense taxonomy for understanding human action, and therefore call it into question as an adequate picture of agency. Whether DBS and other brain interventions will constitute an eliminativist challenge to agency is, at least in part, an empirical question that must be answered by looking at the effects of DBS on the entire spectrum of ways in which we conceive of agency. That is, examine the nature of *local* challenges.

6. Local challenges

Does DBS result in patterns of effects that seem to deny easy accommodation in our commonsense psychological framework? Examination of case studies should help us answer this question. Of particular interest are case studies in which patients undergo marked changes in their mood, commitments or proclivities while undergoing stimulation. Here, I have not undertaken an exhaustive search of the literature, but instead illustrate my point with a few case studies that stand out as relevant to this question. A thorough investigation of this issue would require a comprehensive survey of the cases available, as well as a forward-looking project to more thoroughly document the effects of DBS in the future, as I will discuss below.

(a) Beliefs

There are indications that DBS stimulation can affect the occurrent beliefs of patients. Klaming and Haselager describe a patient treated with DBS for his uncontrollable tics caused by Tourette's syndrome. They related that after prolonged treatment, when the stimulation amplitude was increased, the patient's behaviour changed dramatically. They describe him as entering 'an alternate identity state'. As they relate, 'Increasing the amplitude of stimulation resulted in the patient 'anxiously crouching in a corner, covering his face with his hands. He spoke with a childish high-pitched voice and repeatedly insisted that he was not to blame. Sentences were brief and grammatically incorrect. If approached by one of us, he fiercely kicked his feet because he feared being thrown in the basement' [26, p. 545]. When the amplitude of the stimulation was decreased, the patient's responses became adequate again and he was unable to recall what exactly had happened, although he could report to have been overwhelmed by bad childhood memories. . .' ([26] as reported in Klaming [14]). Although Klaming and Haselager describe this as a case of dissociation or change in the patient's identity, one might also consistently describe it as the establishment of irrational beliefs by electrical stimulation. This patient erroneously believed he was a child again, and that the doctors were going to hurt him and throw him in the basement. The beliefs were irrational in that they were not sensitive to evidence and were caused in an aberrant manner, not based on evidence.[8]

(b) Desires and motivation

Motivational states are the folk psychological states most often reported as being influenced by DBS. Because stimulation sites for many Parkinson's, Tourette's and OCD patients are in deep brain nuclei, located in or very near central parts of the reward system, it is not unusual for stimulation to result in the formation of pathological desires, probably owing to tissue conduction of the electrical currents.[9] It is not uncommon for addictions to gambling or pornography or for hypersexuality to develop in people formerly uninterested in these things, or those who may have had only mild interest that was easily sublimated while in an unstimulated state. More particular alterations in desires have also been reported, such as in the case of 'Mr. B', the 60-year-old OCD patient who developed a strong preference for the music of Johnny Cash with stimulation of the nucleus accumbens. Under stimulation, the patient wanted to listen exclusively to Johnny Cash music, despite the fact that before DBS the man had wide and eclectic musical tastes [27]. Moreover, his exclusive attraction to Cash's music was abolished when the stimulators ceased to function.[10]

Less focused motivational effects are also frequently reported with DBS [17]. For instance, a professional woman who had been energetic and dogged in fighting symptoms of her Parkinson's disease for years was successfully treated by DBS in that her motor symptoms were ameliorated. However, once she no longer identified herself as battling the disease, she experienced a lack of focus, motivation and energy. According to Kraemer, 'After 18 months of stimulation, she was no longer able to work, had a loss of inspiration and a taste for her work and for life in general. . .Her family no longer interested her, she was easily exhausted, and had a loss of vitality (in the absence of a depressive syndrome), which led her to interrupt all professional activity' [18, pp. 488–489].

(c) Reasons-responsiveness/rationality

One of the most widely cited cases in the literature involves a 62-year-old Dutch patient, who developed an uncontrolled mania while the stimulator was on [28]. As Glannon relates the case: 'A mood stabiliser failed to control his symptoms, which included megalomania and chaotic behaviour that resulted in serious financial debts. He became mentally incompetent. Adjustment of the stimulator resolved the mania and restored his cognitive capacity for insight and rational judgment. Yet this resulted in a return of his motor symptoms, which were so severe that the patient became bedridden. This left the patient and his healthcare providers with a choice between two mutually exclusive options: to admit the patient to a nursing home because of a serious physical disability, despite intact cognitive and affective capacities; or to admit the patient to a chronic psychiatric ward because of a manic state, despite restoration of good motor function' [29, p. 290]. The mania caused by the stimulation deprived the man of rationality, and thus undermined his rational agency in a straightforward way. Indeed, Mathews, who mentions this case, characterizes it as a choice between mental competence with motor deficits, and incompetence and restoration of motor control [5].

Thus, case studies provide evidence for effects upon beliefs, desires and motivation, and upon rational capacities. Other case studies demonstrate effects upon personality and sense of agency, as well as upon self-identification. Thus, upon the cursory consideration undertaken here, we can offer a provisional answer to the eliminativist challenge: although there are local challenges to agency in individual cases (in that while DBS may augment or improve agency

in some ways, it may also be detrimental in others), none of the cases here considered is difficult to reconcile with the folk taxonomy we use to discuss agency. In fact, so far as the case reports go, side effects of DBS seem to be described as altering a person's desires, leading to unfounded or irrational beliefs, or an inability to rationally assess or adjust his or her mental state. In many ways, it seems that DBS side effects can be described as causing a brittleness or inflexibility in folk-psychologically framed cognitive or emotional states that under normal conditions would be more flexible, malleable and easier to integrate into the rest of the psychological make-up of the person.

If this is correct, then the prima facie result is that the eliminativist challenge fails. However, there is a deeper methodological question that we cannot now answer concerning the framework within which we conceptualize agency. For although perusal of the available literature did not reveal any clear conflicts with the folk psychological taxonomy of relevant agential mental states and components, it is worth pausing to consider whether that lack of conflict reflects the objective validity of our conceptualization of agency, or whether it rather is a self-ratifying consequence of our pervasive yet perhaps erroneous perspective on the agential world. Is the lack of conflict with our folk psychological notions evidence of the correctness of these notions, or is it a straightforward consequence of the way we, including our clinicians, select and interpret information to report? Since the clinical observations upon which my provisional conclusion is based are themselves filtered through the conceptual schemes of the clinicians involved and are described to the reader in a familiar vocabulary, perhaps the sheer pervasiveness of folk psychology, not its validity, explains my negative conclusion. Thus, the question remains: do the observations neatly fit our prescientific folk concepts because these concepts represent an accurate model of psychological reality, or do our folk psychological concepts shape the observations in such a way that we are necessarily blind to subtle but real distinctions that would challenge our folk theory? This is a fascinating but exceedingly difficult question to resolve.[11] It is a further question whether, if it is a consequence of our folk notions, it is an inescapable one: could we develop or discover novel agential concepts that more closely carve nature at its joints? In the next section, I suggest ways in which we may approach these questions.

7. Moving forward: How to think about deep brain stimulation and agency

Both DBS and the diseases it is used to treat can threaten a person's capacities for agency. DBS can have deleterious effects on agency, but it can also restore agency that has been eroded by the insidious decline in dopaminergic cells in Parkinson's, or by the symptoms of other neurological and psychiatric diseases, such as TRD, OCD, or Tourette's. Thus, the simple question, 'Does intervention harm or help agency?' is too simplistic to capture the sometimes-conflicting considerations that must be weighed in assessing the value of an intervention. This realization calls into question both the naive binary view of agency and the less-naive but still simplistic view of agency as a unidimensional scalar quantity. Agency is multi-faceted, and thinking of it as a scalar quantity involves a simplification that threatens to hamper our best efforts to determine what course to take in practical situations, and to understand the theoretical basis for our deliberations more broadly. A more nuanced way of conceptualizing agency both philosophically and in the contexts of interventions needs to be developed.

I suggest that we instead envision agency as represented in a multidimensional space whose axes are yet to be determined. The previous section raised the question of whether there is some theoretically neutral way of determining what those axes ought to be. If the reader has some ideas of how to get traction on this vexing problem, I would welcome his or her input. But without that, I suggest trying to harness tools from statistics (such as multidimensional scaling or principal component analysis) to try to determine what dimensions are most descriptive of agency. This would involve coming up with a fine-grained list of possible aspects of agency from both intuition and clinical observation. A metric that takes into account variations on a multiplicity of dimensions should be developed. If we begin to chart normal individuals' locations in that provisional multidimensional space we may find that variation on some of these axes is correlated and redundant, whereas the dynamics along other axes can be better explained by postulating two rather than a single factor. Some initial candidate dimensions could be: motor control, inhibition of impulses, attention, self-identification and so on. If we can articulate such a space and a metric on it, we could identify points in space that describe the extent of a person's agentive capacities, with and without DBS. If DBS is effective, then on some dimensions a person undergoing DBS treatment will have values on some dimensions that exceed the values that describe them when untreated. However, unwanted side effects (at least those that affect agency) will often take the form of diminished values on other axes.

Moreover, the objective measurements of agentive properties on these dimensions are almost certainly not the ones that should govern treatment decisions. Much of the biomedical literature on autonomy and authenticity calls attention to the differing values people have and the importance of self-conception to mental health. Once we have delineated an agentive space, we are likely to find that certain people consider some dimensions more important than others. Think again of the Dutch patient who had to decide whether to live the rest of his days bedridden in a nursing home because of his motor incapacitation, or to instead undergo voluntary commitment to a psychiatric institution because of the unmanageable mania brought on by the DBS treatment that alleviated his motor impairments. This kind of dilemma is unfortunately one that ill people and the medical community must sometimes face. The Dutch patient described by Leentjens *et al.* [28] evidently valued his bodily autonomy and basic physical agentive capacities over his rational capacities and his (theoretical) liberty, but one can easily imagine another patient in the same position making the opposite choice.[12] One aspect of respecting the autonomy of patients is to allow their own perspectives about which dimensions trump others to govern, or at least to carry significant weight. Thus, the objective agency metric discussed above should be weighted in accordance with the patient's values, commitments and desires in order to help determine a course of action. We can conceptualize this as defining another multidimensional space that is a scaling of the first, which reflects not just the objective agentive

properties of a person under an available treatment, but also those properties as valued by the agent. Comparing the location of the patient in this space under various treatment conditions and without treatment may provide a better guide to practical matters than any we have yet.

What I have suggested is a theoretical goal of a new way for conceptualizing agency. The more data-driven we can be in developing such a framework, the better. Practically speaking, two innovations that would be useful for developing this kind of theoretical framework would be (i) an open anonymous database for DBS and other direct brain interventions and (ii) some tools for clinicians to use when evaluating their patients. The latter is important for accomplishing our theoretical goals of the former, since clinicians often fail to ask questions that are of central interest to philosophers and bioethicists. Thus, it would be worthwhile to consider ways in which we could guide them when dealing with patients. For example, the community could work to develop a questionnaire that probes aspects of agency that clinicians are not used to asking about. Or perhaps we could develop a battery of objective tests relevant to agency that clinicians could administer. The database should incorporate detailed descriptions of treatment effectiveness, and investigate and report in detail side effects and other impacts on a person's capacities and agentive properties. The database should also incorporate technical details of the treatment such as stimulation parameters, electrode placement and elements of the patient's diagnosis. The existence of such a database would make it easier to accomplish two worthwhile goals. The first would be to develop a data-driven analysis of the effects of DBS on agency, which would potentially enable us to detect patterns that might suggest that agency would be better conceptualized with an alternative framework than commonsense psychology supplies. The second is perhaps more important: a richly developed database could allow us to better understand the neuroscience of agency, to better understand the neural and behavioural effects of a technique that is largely empirical, and to understand the causes of (and potentially prevent the unwanted consequences of) adverse effects that sometimes occur.

Competing interests. I have no competing interests.

Funding. I received no funding for this study.

Endnotes

[1]It is difficult in fact to make a principled distinction between direct and indirect brain interventions. In some ways, pharmacological interventions are considerably more direct than, say, transcranial magnetic stimulation (TMS). So here I use 'direct' only to pick out a class of interventions that electrically stimulate or inhibit neural tissue.

[2]As our understanding of the mechanisms by which DBS works grows, our effectiveness in combatting the diseases is sure to improve, and the precision with which we make our interventions is also bound to improve. One can easily envision, for example, that the combination of DBS principles with optogenetics will lead to significant improvement in treatment outcomes [8]. Optogenetics is already being used to target the relevant cell populations in animal models of Parkinson's disease, and a light, rather than an electrode, is used to control neural firing [9,10]. With such technology, DBS stimulation targets just the neurons of interest, rather than all the cells in an area of tissue surrounding the electrode. It is likely that such targeted stimulation would significantly reduce unwanted side-effects.

[3]Here the literature is confused. Many authors seem to use the technical term personal identity, which in the philosophical literature signifies numerical identity (what it is to be the same person over time), when in fact they are discussing psychological identification: what it is to see oneself as the person over time, or to identify with other time slices of oneself.

[4]As an illustration of how capacitarian views may differ, consider the way in which judgements about agency and responsibility may differ in the case of a minor who coerces another minor to engage in sexual activities. We likely will judge that the minor acted freely: he possesses the basic capacity for rational agency. However, he may not be held fully legally responsible because he is underage. Justification for the age of majority is given in terms of the normal developmental trajectory of capacities for judgement and self-control. And we may or may not judge of him that he had all the capacities necessary to make him fully morally blameworthy, and be justified in our moral reprobation.

[5]And, to be clear, it is not a mistake philosophers have made.

[6]See also [23]. In actual fact, DBS does not appear to work according to simple disconnection principles, but that is not of primary importance here. A more detailed inquiry into whether different types of interventions pose different challenges to agency will be worth undertaking when we have a better understanding of how DBS and other interventions affect brain networks.

[7]This same sort of idea has been used to assess responsibility in cases in which there is a question of personality change or alterations in a person's narrative identity. For example, in a discussion of responsibility for actions in dissociative identity disorder, Kennett & Matthews [24] argue that responsibility is mitigated because 'the patient does not possess the relevant capacities of judgment and control' with respect to the actions committed in an alter state.

[8]I have not seen many case studies that report things that appear to be 'inserted beliefs' in DBS. This may be because deep brain stimulation sites are less apt to influence belief representations, which scientists think are stored in the cortex. However, because stimulation may affect fibres of passage and activate cells some distance from the stimulation site, I think the possibility cannot be discounted. This case may be an illustration of this.

[9]Parkinson's disease is often treated by stimulation in the subthalamic nucleus (STN); OCD often by stimulation in the nucleus accumbens.

[10]Interestingly, 'Mr. B. reported he felt very confident, calm and assertive and he started to call himself "Mr. B. II", the new and improved version of himself'. ([27], p. 152)

[11]For an attempt to go beyond the theory-ladenness of folk psychology in explaining delusions, see [30].

[12]This case is also interesting because there was a question about in which state the patient should make a decision about his future: his competent unstimulated state, or his manic stimulated state. In this case, the determination was made in the unstimulated (mentally competent) state, and it is interesting that the patient chose to remain in the psychotic state. As far as I can determine, he was not given the option of deciding in his manic state. One can however imagine situations in which the prudential choices a person makes might be state-dependent and unstable.

References

1. Young L, Camprodon JA, Hauser M, Pascual-Leone A, Saxe R. 2010 Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proc. Natl Acad. Sci. USA* **107**, 6753–6758. (doi:10.1073/pnas.0914826107)
2. Parvizi J, Jacques C, Foster BL, Withoft N, Rangarajan V, Weiner KS, Grill-Spector K. 2012 Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci.* **32**, 14 915–14 920. (doi:10.1523/JNEUROSCI.2609-12.2012)

3. Diester I, Kaufman MT, Mogri M, Pashaie R, Goo W, Yizhar O, Ramakrishnan C, Deisseroth K, Shenoy KV. 2011 An optogenetic toolbox designed for primates. *Nat. Neurosci.* **14**, 387–397. (doi:10.1038/nn.2749)
4. Tye KM, Deisseroth K. 2012 Optogenetic investigation of neural circuits underlying brain disease in animal models. *Nat. Rev. Neurosci.* **13**, 251–266. (doi:10.1038/nrn3171)
5. Mathews DJH. 2011 Deep brain stimulation, personal identity and policy. *Int. Rev. Psychiatry* **23**, 486–492. (doi:10.3109/09540261.2011.632624)
6. Holtzheimer PE, Mayberg HS. 2011 Deep brain stimulation for psychiatric disorders. *Annu. Rev. Neurosci.* **34**, 289–307. (doi:10.1146/annurev-neuro-061010-113638)
7. Kringelbach ML, Jenkinson N, Owen SLF, Aziz TZ. 2007 Translational principles of deep brain stimulation. *Nat. Rev. Neurosci.* **8**, 623–635. (doi:10.1038/nrn2196)
8. Little S, Brown P. 2014 Focusing brain therapeutic interventions in space and time for Parkinson's disease. *Curr. Biol.* **24**, R898–R909. (doi:10.1016/j.cub.2014.08.002)
9. Gradinaru V, Mogri M, Thompson KR, Henderson JM, Deisseroth K. 2009 Optical deconstruction of parkinsonian neural circuitry. *Science* **324**, 354–359. (doi:10.1126/science.1167093)
10. Kravitz AV, Freeze BS, Parker PRL, Kay K, Thwin MT, Deisseroth K, Kreitzer AC. 2010 Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nat. Neurosci.* **466**, 622–626. (doi:10.1038/nature09159)
11. Roskies AL. 2002 Neuroethics for the new millenium. *Neuron* **35**, 21–23. (doi:10.1016/S0896-6273(02)00763-8)
12. Baylis F. 2013 'I am who I am': on the perceived threats to personal identity from deep brain stimulation. *Neuroethics* **6**, 513–526. (doi:10.1007/s12152-011-9137-1)
13. Witt K, Kuhn J, Timmerman L, Zurowski M, Woopen C. 2013 Deep brain stimulation and the search for identity. *Neuroethics* **6**, 499–511. (doi:10.1007/s12152-011-9100-1)
14. Klaming L, Haselager P. 2013 Did my brain implant make me do it? Questions raised by DBS regarding psychological continuity, responsibility for action and mental competence. *Neuroethics* **6**, 527–539. (doi:10.1007/s12152-010-9093-1)
15. Schechtman M. 1996 *The constitution of selves*. Ithaca, NY: Cornell University Press.
16. Witt K, Kuhn J, Timmermann L, Zurowski M, Woopen C. 2013 Deep brain stimulation and the search for identity. *Neuroethics* **6**, 499–511. (doi:10.1007/s12152-011-9100-1)
17. Schüpbach M, Gargiulo M, Welter M, Mallet L, Béhar C, Houeto JL, Maltête D, Mesnage V, Agid Y. 2006 Neurosurgery in Parkinson disease: a distressed mind in a repaired body? *Neurology* **66**, 1811–1816. (doi:10.1212/01.wnl.0000234880.51322.16)
18. Kraemer F. 2013 Me, myself and my brain implant: deep brain stimulation raises questions of personal authenticity and alienation. *Neuroethics* **6**, 483–497. (doi:10.1007/s12152-011-9115-7)
19. Vincent NA. 2012 Neurolaw and direct brain interventions. *Law Philosophy* **8**, 43–50. (doi:10.1007/s11572-012-9164-y)
20. Fischer J, Ravizza M. 1998 *Responsibility and control: a theory of moral responsibility*. Cambridge, UK: Cambridge University Press.
21. Frankfurt H. 1969 Alternate possibilities and moral responsibility. *J. Philos.* **66**, 829–839. (doi:10.2307/2023833)
22. Frankfurt H. 1971 Freedom of the will and the concept of a person. *J. Philos.* **68**, 5–20. (doi:10.2307/2024717)
23. Levy N. 2008 Counterfactual intervention and agents' capacities. *J. Philos.* **105**, 223–239. (doi:10.2307/20620095)
24. Kennett J, Matthews S. 2002 Identity, control and responsibility: the case of dissociative identity disorder. *Philos. Psychol.* **15**, 509–526. (doi:10.1080/09515089.2002.10031978)
25. Churchland PM. 1981 Eliminative materialism and the propositional attitudes. *J. Philos.* **78**, 67–90. (doi:10.2307/2025900)
26. Goethals I, Jacobs F, Van der Linden C, Caemaert J, Audenaert K. 2008 Brain activation associated with deep brain stimulation causing dissociation in a patient with Tourette's Syndrome. *J. Trauma Dissoc.* **9**, 543–549. (doi:10.1080/15299730802226126)
27. Mantione M, Figee M, Denys D. 2014 A case of musical preference for Johnny Cash following deep brain stimulation of the nucleus accumbens. *Front. Behav. Neurosci.* **8**, 152. (doi:10.3389/fnbeh.2014.00152)
28. Leentjens AF, Visser-Vandewalle V, Temel Y, Verhey FR. 2004 Manipulation of mental competence: An ethical problem in case of electrical stimulation of the subthalamic nucleus for severe Parkinson's disease. *Nederlands Tijdschrift voor Geneeskunde* **148**, 1394–1398.
29. Glannon W. 2009 Stimulating brains, altering minds. *J. Med. Ethics* **35**, 289–292. (doi:10.1136/jme.2008.027789)
30. Gerrans P. 2014 *The measure of madness: philosophy of mind, cognitive neuroscience, and delusional thought*. Cambridge, MA: MIT Press.