CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Brain-State Transitions, Responsibility, and Personal Identity

Stephen Rainey[1,2,3]* ⓘ, Karmele Olaciregui Dague[4] ⓘ and Roger Crisp[1,2,5]

[1]Oxford Uehiro Centre, University of Oxford, Oxford, UK, [2]Faculty of Philosophy, University of Oxford, Oxford, UK, [3]TUDelft, Delft, The Netherlands, [4]Department of Epileptology, Bonn University Hospital, Bonn, Germany and [5]St Anne's College, Oxford, UK
*Corresponding author. Email: s.rainey@tudelft.nl

Abstract
This article examines the emerging possibility of "brain-state transitioning," in which one brain state is prompted through manipulating the dynamics of the active brain. The technique, still in its infancy, is intended to provide the basis for novel treatments for brain-based disorders. Although a detailed literature exists covering topics around brain-machine interfaces, where targets of brain-based activity include artificial limbs, hardware, and software, there is less concentration on the brain itself as a target for instrumental intervention. This article examines some of the science behind brain-state transitioning, before extending beyond current possibilities in order to explore philosophical and ethical questions about how transitions could be seen to impact on assessment of responsibility and personal identity. It concludes with some thoughts on how best to pursue this nascent approach while accounting for the philosophical and ethical issues.

Keywords: neuroethics; neurotechnology; responsibility; personal identity; philosophy; brain-state transition

Introduction

There are clear cases of benefits to be derived from neurotechnology. Brain-controlled devices like artificial limbs, for example, allow those who might otherwise be hampered by physical disability to regain mobility and independence. Even still, philosophical questions arise concerning issues of responsibility for technology-mediated actions.[1,2] These typically center on how we ought to evaluate levels of control over prosthetic limbs and similar devices. These kinds of issues reappear when we consider neurotechnologies aimed at addressing neurological or psychiatric conditions. Again, benefits are clear where neurotechnology can relieve the burdens of, for example, severe epilepsy in halting seizures or restoring consciousness,[3] or alleviating depression.[4] Questions of responsibility and control may take on another dimension in such cases, as the object of control is no longer an "external" action, but the activity of the brain itself.[5] The brain, as essential to all we think, and do, presents an especially sensitive focal point for intervention, and, given the close relationship between the brain and selfhood or identity, possibilities of self-change of a different order to those arising from, say, artificial limbs.

Through understanding brain dynamics, *control* of brain activity seems increasingly plausible. This control might go beyond even that required for treating epilepsy or depression, into areas of prompting brain-state transitions. A medium-term future of instrumentalized brains could open the possibility of

brain control in quite fine-grained ways for therapeutic contexts—as of using the brain as a control panel to bring about desired mental, bodily, or dispositional traits.[6]

To explore this, we will first need to see what kinds of technologies would allow brain-state transitioning. The technology is in its infancy. Nevertheless, for a relevant critical appraisal of questions around brain control and its consequences, it is important to get a handle on the actual technology and its potential future capabilities, to permit constructive speculation. Having got to grips with this, we will pursue philosophical and ethical questions surrounding brain-state transitioning, personal responsibility, and personal identity.

Technologies for Brain-State Transitions

What should we think about cases where technology-mediated action is not the use of a limb to move an object, or neural-signal controlled steering of a wheelchair, but control over brain activity itself? The action undertaken in such a case might be the manipulation of brain activity according to a brain-state-transitioning paradigm, such as that explored in silico by Kringelbach and Deco.[7,8] Their ambition is to provide a means with which to intervene on brain activity in order to transition from a given to a target brain state, and with that to produce a cognitive or behavioral effect.

The definition of brain states is contested among neuroscientists[9,10,11,12,13] (and philosophers)[14] especially as it relates to their experimental findings such as neuroimaging or neurophysiology.

Brain states can be partially experimentally defined by their associated findings, which depend on the metric being applied. We can illustrate this concept using the sleep stages, as defined by their EEG findings. There are several ways to determine the stage of sleep, and we may not require any devices to tell that someone is asleep. Nevertheless, when defining sleep brain states as N1, N2, N3, and rapid eye movement (REM), we are incorporating more than the presence or absence of eye movements to the definition, namely the presence or absence of certain EEG markers such as sleep spindles or K-complexes. The same can be said of the application of EEG to define brain states in anesthesia such as burst suppression. Similarly, recent results[15] suggest that, at the microscopic and macroscopic level, unconscious brain states can be fundamentally characterized by synchronous activity,[16] and conscious states by asynchronous activity.[17,18] The appearance of certain findings is statistically tied to a certain brain state, although the finding and the brain state are not identical.

A good working definition of brain states would ideally incorporate the causal processes involved, and the complexity of the interplay between these processes[19] given that even the most disparate brain states, from coma to mania, essentially arise from them. In this vein, Goldman et al. suggest the possibility that brain states derive from interactions of varying dynamical complexity between neuronal populations,[20] similar to chemical states of matter:

> Much as different states of matter like solids, liquids, and gases emerge from interactions between populations of molecules, different brain states may emerge from the interactions between populations of neurons. (…) In this sense, macroscopically observed high synchrony, low complexity brain signals recorded from unconscious states may be accounted for by an increased coupling in the system's components, behaving more like a solid (…). In contrast, conscious brain states may be described as higher complexity (…), perhaps liquid-like.

Although this spectrum is helpful in clarifying the crucial differences between broad states, such as consciousness and unconsciousness, the transitions between these are fluid, spreading dynamical changes that give rise to more complex[21] intermediate states. Finding a definition and an experimental model with sufficient complexity while still being flexible in its application is a challenge.

When trying to define, and possibly influence, increasingly complex brain states, metrics become conversely more diverse and complex. Whole-brain models (WBMs) open a wide range of possibilities in

this respect. WBMs are an array of equations that characterize dynamics between populations of neurons and aim to link them to anatomical structures.

Unsurprisingly, the definitions of brain states that arise in this arena are also dependent on the applied method, and vary among authors.[22,23] Moreover, there remains a significant disconnect between metric-dependent definition of brain states and intuitive/subjective experience of brain states (especially for mood, etc.), that is, the experiential content of a mental state. Although it is relatively straightforward to distinguish sleep from wakefulness, or a seizure from a nonseizure state, other states may prove more elusive. Distinctions among "happy" states, for instance, will abound while at the same time evading definition. So too among other familiar but difficult to pin down mental states, like grief, melancholy, focus, flow, calm, and so on. Outside of the scope of discussion here, but of vital interest, is the connection among brain states, mental states, and language: words for mental states need not map onto any scheme of brain-state findings we can possibly discover.[24,25]

Kringelbach and Deco present "…a mechanistic framework for characterizing brain states in terms of the underlying causal mechanisms and dynamical complexity,"[26] which in practice means an ensemble of neuroimaging, spatiotemporal dynamic, empirical behavioral, EEG, MEG, fMRI, and PET data. This ensemble presents a picture of brain states that includes their dynamic relations to one another. This heterogenous approach allows WBMs in silico that aim to strike a balance between complexity and realism to provide a detailed, yet still computationally workable, model of the living human brain in vivo. One aim of this approach is to permit instrumental intervention on the brain, in order to prompt transitions from one brain state to another:

> "(A) The goal of whole-brain modeling is to predict potential perturbations that can change the dynamical landscape of a source brain state such that the brain will self-organize into a desired target brain state."[27]

If perturbations can be predicted, modeled in silico, then they might also be created as stimuli that can serve to bring about self-organization into a target brain state. One such transition discussed in the literature is that between waking and sleep states. Deco et al.[28] used stimulation of WBMs in silico to force transitions between model sleep and wakefulness states, with some success in prompting changes from one to the other (less with wakefulness to sleep). This kind of technology is in its infancy, with human trials not yet begun. But basic scientific payoffs are hotly anticipated, as are therapeutic rewards:

> The approach could be used as a principled way to rebalance human brain activity in health and disease… by causally changing the cloud of metastable substates from that found in disease to the health in order to promote a profound reconfiguration of the dynamical landscape necessary for recovery, rather than having to identify the original point of insult and its repair.[29]

The possibility of brain-state transitioning raises questions across a variety of scientific grounds, but also prompts considerations from conceptual and ethical perspectives. For the purposes of this article, we extend the principle of brain-state transitioning into the future, albeit guided by the possibilities latent within Kringelbach and Deco's account. We wish to retain proximity to realistic scientific possibility, but also to anticipate questions and puzzles prior to their emergence. Moreover, "near-future" scenarios can offer vivid means of getting to grips with otherwise tricky scientific material.

With a technology of brain-state transitioning, a user could control the states of their brain in order to evolve desirable outcomes like more recuperative rest, boosted prosocial dispositions, motivation to work, or elevated mood. The brain in such a scenario becomes both the origin of, and the target for, increased control. Similar questions about responsibility and control will arise here as with brain-computer interface (BCI)-driven devices but compounded by a reflexive note. We will be able to formulate some puzzles that will emerge, in which the reflexive note makes clarity difficult to achieve in analyzing brain-state transitions.

Addressing questions arising here not only holds value in terms of how best to understand empirical investigation into emerging neuroscientific technology, but also raises far-reaching conceptual questions. These include the interrelations among brain, mind, and action as well as the meaning and role of intention in action. Part of the point is that brain states are underdetermined by their findings, and mental states are underdetermined by the subjective reports that describe them. So, with brain-state transitions, there is a double underdetermination problem: one dimension is scientific, the other experiential. It becomes difficult to figure out what to aim at as a "target brain state" into which to transition, because the "state" retains elements of obscurity in both scientific and experiential terms. This will have ramifications in terms of how brain-state transitions can be accounted for. In scientific terms, this might be in terms of their findings or in terms of behavior associated with dispositions correlated with those brain states. In experiential terms, accounting for brain-state transitions might come in terms of first-person reports, or interpersonal accounts from other people familiar with the person undergoing transitions. This is important as it will have a bearing on evaluations of success in brain-state transitions. To explore further, we will first go beyond the current horizon of possibility, in order to vivify these and related questions through examples.

Future Iterations of Brain-State Transition Technology

The potential for whole-brain computational modeling opens the possibility for reimagining the brain as a means for taking control of our cognition and behavioral dispositions and with that ourselves. The examples from above provide a way into thinking about these remote possibilities. They are "remote" in the sense of being beyond the horizon in fact. Nevertheless, already there are neurotechnology companies offering consumer devices whose promises include that users may "connect your mind to the digital world and turn science fiction to reality."[30] Knowing what the brain is doing, via visualizing neurofeedback, might help us try to change our behavior and seek different feedback. Other devices might be such that their activity "…trains the brain to maximize attention, focus and alertness, putting the … user in the right mindset to accomplish tasks and perform at a higher level."[31] Even if the technology is not yet there in terms of deliverability,[32] the ambition and trajectory are clear.

Let us imagine a couple of speculative examples, based on a futuristic version of the technology here described. We will return to these throughout the article, in order to illustrate points and to enliven the questions that arise:

1) Ada cannot sleep, and so adjusts her neuromodulation device to the "relax" setting, dampening brain activity such that she begins to drift off at last, transitioning from a stressful to an unconscious state.
2) Bob is nervous before an important job interview, and so adjusts his neuromodulation device to a "go-getter" setting to promote neural activity correlated with confidence and mild risk-taking.
3) Carrie is extremely sad about the death of her dear old aunt, and so adjusts her neuromodulation device to the "cheerful" setting, stimulating her neural activity to an ensemble of states correlated with an upbeat disposition.

What would we think of these kinds of self-interventions? The cases appear to increase in complexity from the relatively innocuous alleviation of insomnia, through the enhancement-implications of the nervous interviewee, to the erasure of a natural emotional response. In each case, some sense of agency is boosted: Ada can sleep better and presumably reap the rewards of good rest. Bob can perform to a standard of which he is capable, but not necessarily on cue. Carrie can go about her day without distractions from grief. But we can probably discern issues attaching to the implications for what this type of control over one's dispositions means. In each case, there may be a failure to grasp root causes of issues, for example, and a rush to treat symptoms. Ought Ada to find effective, sustainable means of addressing anxieties robbing her of sleep? Would Bob be better off gaining confidence through therapy to unravel what is holding him back? Might grief be a human emotion one ought—morally, or for reasons of mental

health—to go through in the face of bereavement in some sense beyond being able to get through a day's tasks undistracted?

Open questions such as these become all the more acute if we take more elaborate imagined cases in which control over brain states is taken to greater lengths.

4) Diana admires Eddie's sense of style and taste. She uses her neuromodulation device so as to promote states that prompt her to like the kinds of aesthetics Eddie likes. She now goes about her day enjoying the kinds of things Eddie does.

5) Fatima has certain dispositions that make it hard for others to like her and can make little academic progress with the intellectual capacities she has. She engages in a major modification process of both her emotional and behavioral dispositions, and her capacities for acquiring knowledge. These enable to enter a challenging career, where she excels. Those who know her speak of her as a completely different person.

In these cases, what is aimed at is not just an instrumental gain, an agency boost based on ameliorating some perceived obstacle. In these cases, Diana and Fatima each seek to alter *themselves* in some substantial sense. Although cases 1–3 suggested control over one's brain to realize better chances at enacting one's own desires, cases 4 and 5 suggest control over the parameters constituting the self, to create specific kinds of beliefs and desires. These aim at self-perfection and consist in self-alteration through intervention on the brain as an instrument for novel self-production. Again, there are open questions: if Eddie and Fatima can recognize the way they want to be, but cannot realize it (or cannot realize it quickly enough), why should they not seek self-perfection through instrumental means? Conversely, they might sell themselves short by using shortcuts such that they end up with a kind of Potemkin personality, lacking the depth of that wrought through dedicated pursuit of one's ends.

On the one hand, the agency-boosting potential in control over brain states is apparent, both in terms of instrumental aims and wider personal self-perfection. But on the other, we may have pause to think about what is going on with examples like those above, where the potential is acted on in some diurnal contexts. We explore this further with a discussion of action, including technology-mediated action.

Brain-Controlled Devices in General

We will assume here a minimalistic account of action, based on beliefs and desires. There are more complex accounts available, but for the purposes of this analysis, beliefs and desires will suffice. Essentially, sorting the physical world into relevant types for use in an explanation of action, or for a scientific account of the mind, requires mental concepts like "belief" and "desire." This can be framed in terms of a "difference-making" model of causality. The use of mental concepts is what *makes the difference* in a causal intervention instigated by a person. This can be illustrated with Gary, who wants a cup of coffee.

6) Gary wants coffee. He believes the mug in front of him contains coffee. He reaches with his arm, grasps the cup, draws it to his mouth, and takes a sip.

Looked at as a series of movements alone, Gary's reaching is just so much limb velocity. But if we are interested in what Gary is doing, why he decides to drink the coffee rather than the water in front of him, there are questions. Presumably, Gary drinks the coffee rather than the water, because he *wants* the coffee. It must be accounted for how two brain states, one desiring coffee and another desiring water, might produce different explanations for overt behavior. There is a difference between Gary's reaching for coffee, reaching for water, or flailing an arm in an objectless reaching motion. To explain the action, it is necessary to use mental concepts like "desire." Although a causal account is correct in highlighting the causal structure of bodily movement, from brain through limbs, to observable macrolevel behavior, it is incomplete in terms of explaining the *action*.

Put the other way around, if we want to change Gary's behavior and have him drink the water, we would typically intervene at the level of desires, not causality. Convincing Gary that the water is better for him might change his mind and have him reach for the water. Physically diverting his reach toward the water might succeed only in baffling or frustrating him. The relevant factors in agency are those relating to mental concepts. Gary reaches for the coffee, *because he wants coffee* and *believes* the cup to contain it.

Although the physical facts of the brain are not the same as facts about the mind, the physical facts of the brain matter for the mind. Donald Davidson relates the mental to the physical in terms of supervenience:

> …mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect.[33]

These ideas arise acutely in cases of technology-mediated actions, where technologies are used to sense the states of brains and translate the characteristics of brain activity into outward device actions. BCIs are increasingly available in rehabilitative medicine, where prosthetic limbs, brain-controlled software programs, or speech synthesizers can be used to replace or restore body function lost to illness. In understanding ascriptions of action to the users of BCIs, it is essential to relate the actions done via the BCI device to the reasons the users can, or could possibly, produce for that action. This brings into practical use the "difference-making" account of causality.

7) Hilary wants to hail a taxi. She believes the cab with a light on means it is available, and that raising her arm indicates to the driver she would like to be picked up. She reaches her BCI-controlled prosthetic arm upward in a busy street.

Were Hilary to raise her arm in this exact fashion, whether to point to a sign, to signal to a friend her presence, or to lash out at an enemy, the motor neuron activity of her brain would exhibit very comparable activity. But if while trying to hail the taxi Hilary accidentally struck a passer-by, she would apologize in terms of her reasons for raising her arm, not her motor neurons. Her *action* is hailing the taxi, and the accident ought to be seen in those terms. It is the taxi-hailing that makes the difference between her hailing a taxi and punching a passer-by.

This last point serves to highlight an important feature of action as opposed to mere bodily movement—the possibility for responsibility-taking. One can be praised, blamed, or otherwise held accountable for one's actions in ways not available for mere behavior. If Gary is mistaken about the coffee cup, he can accept that mistake and explain it somehow. Hilary's explanation of her mistaken striking of the passer-by does a similar job of responsibility-taking.

How, then, ought we to think of the earlier examples, 1–5? Can the action taken to control the brain and to produce desirable outcomes be understood in the same ways as conventional action? Such an action is appreciably nonnatural, at least in the sense that it is mediated by some intervening device. There is a lively literature on BCI-mediated action, responsibility, and control. Here, we want to go deeper into the complexities of a case in which a limb and action performed with it are not the target. This goes deeper into the complexities, as the relations among brain states, mental states, and responsibility-taking are probed further. To preempt some questions: if an agent's reasons explain her mental states such that they are the difference-makers in explaining actions, and these depend on the brain's physical states, what of situations in which an agent wants to alter their own brain states in order to bring about desired mental states?

"Paradox of Self-Control"

Efforts at changing one's mood can be unreliable. Perhaps melancholy can be dispelled most of the time by taking a walk. But this may not always be the case. In setting out on a walk to clear one's mind, there

may be a fully formed intention, and a strong desire to lift one's spirits, that nonetheless are thwarted. On another occasion, cheerfulness may return without much consideration. Is there something worthwhile in the uncertainty of striving that makes its success valuable? There may be something informative, disclosive of wider self-knowledge in striving and failing. Compare with the case of pharmacological intervention to change something like a mood disorder. A relevant difference between the brain-state transition case and the pharmacological is one of time and the related room for deliberation. Pharma intervention takes a while, and can accommodate a user's reflection leading to alteration of dose, and so forth. Brain-state transition maybe not: it could be immediate, like the deep brain stimulation treatments for Parkinson's tremors. This could leave no room to consider its personal effects and impacts on others.

Brain-state transitioning might be more like drinking to forget than psychological therapy. Maybe there should be consideration given to be variation in appropriateness for applications, for example, when target brain states can be well defined or clinically good ends discerned. Transition technology for post-traumatic stress disorder (PTSD) might then be seen as a worthwhile cause, less so for acute depression following something like a death in the family or loss of a job. The former constitutes a recognizably therapeutic context, whereas the latter may fall more into something unfortunate that is just part of life. If we think about the value afforded the PTSD sufferer whose symptoms are alleviated via transitioning, we see things like quality of life, and well-being, personal and social functioning, and the possibility for their restoration or rehabilitation. For the kinds of grief or depression associated with loss, however, we might think the value is purely in terms of short-term instrumental gain. The former case appears therapeutic, and the latter may appear more of an avoidance tactic or something that will in time pass anyway.

Is "wanting to be happy" and inducing that happiness an action in the sense that Gary's coffee drinking, and Hilary's taxi-hailing? The difference between random arm-waving and those instances is explained in terms of the reasons each had for making the specific movement—the beliefs each held about the world around them, and the desires they had about interacting with that world. In the brain-state transition case, things are not so clear cut. There is no correlate for the "specific movement," as there is with reaching for a cup or raising an arm to hail a taxi (cf. Figure 1).
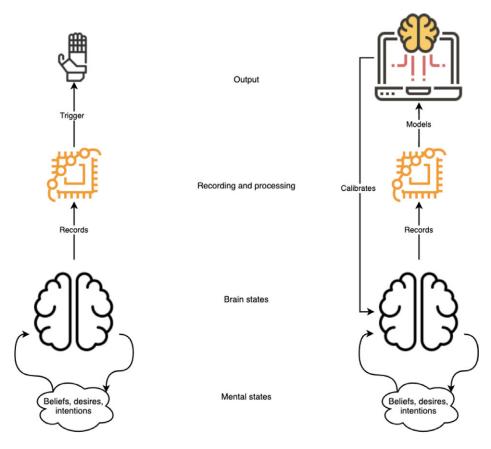
In those cases, (1) a desire is realized by a brain state, and (2) satisfied by onward motor activity, that is (3) understood best using intentional descriptions that capture the "difference-making" characteristic of the action.

Brain-state transitioning begins with (1) but moves then to (2*) algorithm-led brain stimulation realizing a brain state conducive to an onward disposition/dynamically preclusive of specific clouds of metastable substates such that (1) is satisfied directly. The difference-making might be located in the desires at (1) or in the stimulatory action at (2*). If the difference-making is located at (2*) it is not intentional action as normally understood.

Put differently, Diana's desire to be more like Eddie is satisfied in being more like Eddie, not by doing any specific action. The realizer for this Eddie-likeness is dynamic activity of the brain-state transition device, recording, stimulating, and calibrating Diana's brain so as to attract her to things Eddie is disposed to like (cf. Figure 1 again). This realization is therefore best understood in terms of the functioning of the transition device. It is not best understood in terms of intentional descriptions. The difference-maker is the device. The device-maintained brain state, conducive to the production of Eddie-like dispositions, seems more like a technologically mediated novel self than it does an intentionally controlled version of Diana. The device can be intentionally used, but from the moment it is in use, it seems intentional control becomes a murky notion in virtue of the ensemble of dispositions generated.

Something like a paradox of self-control appears to be a possibility here: in trying to gain control over the brain in order to realize target brain states, and with those, desirable dispositions and behaviors, there emerges the risk that one ends up with less self-control. This may produce a gap in responsibility, and questions about identity.

Where Hilary's intended taxi-hailing went wrong, she could explain what happened in terms of her unrealized desire or mistaken belief, if, for example, a sudden cramp meant she could not raise her arm,

Figure 1. A distinction between brain-computer interface-mediated limb movement and brain-state transition technology. In the first case, like in example 7 (above), brain states trigger a device and ensuing action is explicable in terms of a user's reasons. In the second case, brain states themselves are recorded, modeled, and then fed back to condition further brain states with the intention of modifying mental states. How an agent's reasons ought to be included in this scheme is in need of exploration.

or she realized the car ahead was not a taxi at all. If instead her dispositions had been altered according to a brain-state transition device, it seems worth asking if she could similarly explain. In order to do so, Hilary would be expected to dispel her own disposition to realize the error. But given the device has placed her in a target brain state, by hypothesis conducive to correlated mental states, it seems less plausible she could evaluate her state of mind as she could a bodily action. Although there is nothing to suggest she will not remember how she came into that state of mind, or that she would not be aware of how she intentionally came to be in it through her own action, there remain issues of risk and how informed Hilary might really be about achieving her target brain state. The brain state will constrain onward thought and action in ways it is intended to, but that could nonetheless be seen as quite unusual.

Imagine Diana wishes to emulate Eddie's coolness under pressure, but her device nudges her more toward nonchalance. Diana could not be expected to realize the mistake, because she would be in a state of nonchalance, and so the difference from her posttransition perspective may seem simply unimportant. Although she is responsible for flipping the switch, so to speak, and choosing to use the brain-state transition device, everything thereafter is out of her hands. In flipping the switch, the user of brain-state transition technology would risk becoming a significantly different person, as opposed to a more controlled version of themselves.

There may be something to be gained here by thinking about Aristotle on responsibility in Book 3 of the Nicomachean Ethics.[34] Imagine someone like Fatima, but who has made themselves much more clever and vicious, thinking that this will provide them with more power and money. Let us further

imagine the posttransition self ends up not being powerful and rich, but reviled as being too coarse, arrogant, and venal. This person might say of their new outlook: *my dispositions are now so ingrained that I cannot be held responsible for what they produce. I am what my transition made me, I'm not to blame!* This kind of response might be so if one were to make a really extreme change.

This sounds about as plausible as someone who says: "I know I smashed up the bar. But I was very drunk." Here, Aristotle would say, "You knew this *might* happen, but you went ahead. You took a risk, and so you are responsible for what happened." One of the problems, perhaps, with early transitions is that people may not know very much about how they are likely to be affected, and in these cases, we might not want to hold them responsible for actions which they really could not have foreseen.

Related to this is the question of whether massive momentary transitions might actually change personal identity, rather than produce a "new person" in the sense of a *very different kind* of person. They would not on something like a Cartesian view—if I have an Ego or soul, it will be there after the transition. But with reductionist views, things are different. Most, after Derek Parfit,[35] think that continuity and connectedness really matter, so major and sudden transitions would be a kind of suicide. And if that is true, a transitioned person's very being is an instrumental end of an act committed by a perpetrator now dead. Fatima has created Fatima*, a version of herself she hoped to be, but in the process, Fatima has disappeared. Unlike slower kinds of personal change, like those brought about through habit formation or pharmacological intervention, there may be no room in an immediate and suddenly induced change for reflection or critical appraisal of what is changing. Perhaps what is at stake here is the idea of personal change as a project, versus the same as an action. In the latter case, without opportunities for reflection, how might such an action be judged successful or not? In the absence of Fatima, there seems to be no appropriate perspective from which to assess the success of the newly minted Fatima*. Fatima* is the realization of Fatima's beliefs and desires that prompted the whole endeavor, but the point of view form which Fatima* can be judged a success is missing.

The Aristotelian point might serve to suggest responsibility for brain-state transitions can be assigned on the basis of known risk-taking. The Parfitian reflection on Fatima* might pull in the other direction: the possibility for change in identity may mean the loss of the responsible party through the generation of a perspective disjointed from any prior point of view held by the person in question.

Upshots

Although the speculative vignettes used here have highlighted circumstances quite far from technical possibility, they nonetheless serve to raise important themes for brain-state transitioning in principle. This kind of transitioning has the potential to create questions about responsibility-taking, and to produce impacts on personal identity. In terms of responsibility-taking, transitioning may be difficult to explain as an action taken by a device user at a time, as there may be unexpected subsequent transition-prompted behaviors or dispositions. These behaviors and dispositions attach to the user in an unusual way, one unlike more run of the mill intentional action motivated by beliefs and desires. In terms of impacts on personal identity, radical transitions may lead to ensembles of actions, beliefs, and desires that from at least some perspectives amount to those of novel persons. This possibility resides in the chance of significant discontinuity between psychological states. Finally, these two conceptual issues can merge in the question of who may be responsible for a radical transition if, as discussed, it amounts to some form of suicide or self-erasure.

From these reflections, we might extract three main messages for brain-state transitioning as a future practice:

1) Subjects must be properly informed before causing a transition, that is, before being enabled to do so.
    a) A therapeutic context is essential, but so too is the cultivation of an appreciation of wider issues in responsibility-taking and personal identity. This is important at least in accounting for the differences among a physician's control over another's brain; control over one's own brain from one time to the next; and a change of one's brain state.

2) Any progress here must be very gradual—technologies for transitions ought to be embedded in wider clinical and multistakeholder contexts to ensure the value of the endeavor.
   a)   The content of transition states ought to be the object of careful reflection, for clinicians and transitioners, and others likely to be affected. This is important at least in terms of understanding issues of control, risk, and the promises of brain-state transitions as prompting changes in mental states, dispositions, personalities, or other potential characterizations.
3) Specific transitions ought to be gradual, for similar reasons.
   a)   We should try to avoid massive, momentary transitions, especially to avoid problems for evaluation and assessment of the transition itself. This is important at least in terms of how scientific accounting for brain states can be resolved with experiential reports of mental states and dispositions, and from which kind of perspective transition success can best be evaluated.

Manipulating the brain through prompting brain-state transitions might seem to permit a new kind of self-making. Perhaps one could boost practical agency by gaining or refining capacities, as with rehabilitative medicine. But, by the same token, if the brain is used to promote and inhibit states according to our choices, it might cause these states in ways that preclude onward reflection and evaluation. We might find dimensions of choice tapering off, as one state tends toward promoting the next, and the next, and so a net *loss* of agency might be the end result. A productive way to galvanize the potential for positive advances with brain-state transition technologies would come from careful progress, bearing these very general considerations in mind.

Notes

1. Rainey S, Maslen H, Savulescu J. When thinking is doing: Responsibility for BCI-mediated action. *AJOB Neuroscience* 2020;**11**:46–58.
2. Matthias A. The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology* 2004;**6**:175–83.
3. Kundishora, A. J, Gummadavelli, A, Ma, C, Liu, M, McCafferty, C, Schiff, N. D, *et al.* Restoring conscious arousal during focal limbic seizures with deep brain stimulation. *Cerebral Cortex* 2017;**27**:1964–75.
4. Perez-Caballero L, Soto-Montenegro ML, Hidalgo-Figueroa M, Mico JA, Desco M, Berrocoso E Deep brain stimulation electrode insertion and depression: Patterns of activity and modulation by analgesics. *Brain Stimulation* 2018;**11**:1348–55.
5. Gilbert F, O'Brien T, Cook M. The effects of closed-loop brain implants on autonomy and deliberation: What are the risks of being kept in the loop? *Cambridge Quarterly of Healthcare Ethics* 2018;**27**:316–25.
6. Campbell J. II—Control variables and mental causation. Proceedings of the Aristotelian Society (*Hardback*) 2010;**110**:15–30.
7. Kringelbach ML, Cruzat J, Cabral J, Knudsen GM, Carhart-Harris R, Whybrow PC. *et al.* Dynamic coupling of whole-brain neuronal and neurotransmitter systems. *Proceedings of the National Academy of Sciences of the United States of America* 2020;**117**:9566–76.
8. Kringelbach ML, Deco G. Brain states and transitions: Insights from computational neuroscience. *Cell Reports* 2020;**32**(10):108128.
9. Carhart-Harris RL, Muthukumaraswamy S, Roseman L, Kaelen M, Droog W, Murphy K. *et al.* Neural correlates of the LSD experience revealed by multimodal neuroimaging. *Proceedings of the National Academy of Sciences of the United States of America* 2016;**113**:4853–8.
10. Goldman JS, Tort-Colet N, Di Volo M, Susin E, Bouté J, Dali M *et al.* Bridging single neuron dynamics to global brain states. *Frontiers in Systems Neuroscience* 2019;**13**:75.
11. Tononi G, Sporns O, Edelman GM. A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences of the United States of America* 1994;**91**:5033–7.

12. Deco G, Tononi G, Boly M, Kringelbach ML. Rethinking segregation and integration: Contributions of whole-brain modelling. *Nature Reviews Neuroscience* 2015;**16**:430–9.

13. McCormick DA, Nestvogel DB, He BJ. Neuromodulation of brain state and behavior. *Annual Review of Neuroscience* 2020;**43**:391–415.

14. Brown R. What is a brain state? *Philosophical Psychology* 2006;**19**:729–42.

15. Brown EN, Lydic R, Schiff ND. General anesthesia, sleep, and coma. *The New England Journal of Medicine* 2010;**363**:2638–50.

16. Henry JC. Electroencephalography: Basic principles, clinical applications, and related fields, fifth edition. *Neurology* 2006;**67**:2092.

17. Boly M, Phillips C, Tshibanda L, Vanhaudenhuyse A, Schabus M, Dang-Vu TT *et al.* Intrinsic brain activity in altered states of consciousness. *Annals of the New York Academy of Sciences* 2008;**1129**:119–29.

18. Sanchez-Vives MV, McCormick DA. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nature Neuroscience* 2000;**3**:1027–34.

19. Escrichs, A, Sanz Perl, Y, Uribe, C, Camara, E, Türker, B, Pyatigorskaya, N, et al. Unifying turbulent dynamics framework distinguishes different brain states; 2021; available at http://biorxiv.org/lookup/doi/10.1101/2021.10.14.464380. doi:10.1101/2021.10.14.464380.

20. See note 10, Goldman et al. 2019, at 1.

21. Huber R, Felice Ghilardi M, Massimini M, Tononi G. Local sleep and learning. *Nature* 2004;**430**:78–81.

22. Barttfelda, P, Uhriga, L, Sitta, J. D., Sigmane, M, Jarrayaa, B, Dehaene, S. Signature of consciousness in the dynamics of resting-state brain activity. *Proceedings of the National Academy of Sciences of the United States of America* 2015;**112**:887–92.

23. Tagliazucchi E, Crossley N, Bullmore ET, Laufs H. Deep sleep divides the cortex into opposite modes of anatomical-functional coupling. *Brain Structure and Function* 2016;**221**:4221–34.

24. Gauker C. *Thinking Out Loud: An Essay on the Relation Between Thought and Language*. Princeton University Press, Princeton, NJ; 1994.

25. Churchland PS. *Neurophilosophy Toward a Unified Science of the Mind Brain*. MIT Press; Cambridge, MA; 1989.

26. See note 8, Kringelbach, Deco 2020, at 1.

27. See note 8, Kringelbach, Deco 2020, at 5.

28. Deco, G, Cruzata, J, Cabralf, J, Tagliazucchii, E, Laufs, H, Logothetis, N. K. *et al.* Awakening: Predicting external stimulation to force transitions between different brain states. *Proceedings of the National Academy of Sciences of the United States of America* 2019;**116**:18088–97.

29. See note 28, Deco et al. 2019, at 18094.

30. INSIGHT—5 Channel EEG Brainwear®. EMOTIV; available at https://www.emotiv.com/insight/.

31. tDCS Device for Improving Focus—LIFTiD Neurostimulation; available at https://www.getliftid.com/.

32. Wexler A, Thibault R. Mind-reading or misleading? Assessing direct-to-consumer electroencephalography (EEG) devices marketed for wellness and their ethical and regulatory implications. *Journal of Cognitive Enhancement* 2019;**3**:131–7. doi:10.1007/s41465-018-0091-2.

33. Davidson D. *Essays on Actions and Events*. Clarendon Press and Oxford University Press, Oxford, UK; 2001, at 214.

34. Aristotle, Barnes J, Tredennick H. *The Nicomachean Ethics*. Penguin Classics, London, UK; 2004.

35. Parfit D. Personal identity. *The Philosophical Review* 1971;**80**:3–27.