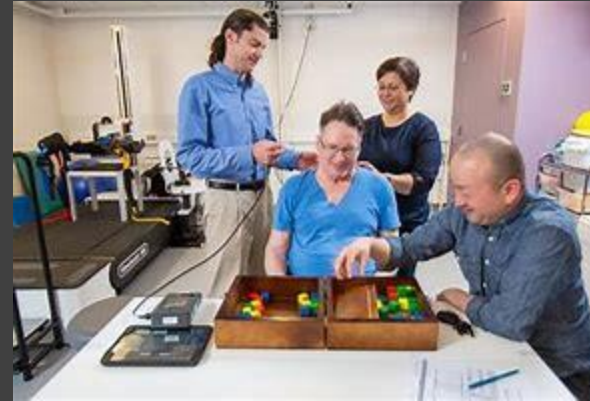# Neurotechnologies and Agency:
Ethical Issues in Engineering the Brain

SARA GOERING

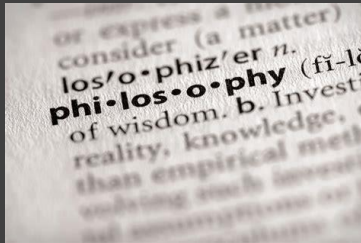DEPARTMENT OF PHILOSOPHY

CENTER FOR NEUROTECHNOLOGY

UNIVERSITY OF WASHINGTON, SEATTLE

MARCH 2021

My positioning

UW Department of Philosophy

Disability Studies Program

Bioethics, philosophy of disability, feminist philosophy

Ethics thrust leader for the UW Center for Neurotechnology (CNT)

Focus groups and interviews with end-users, disabled people, and other stakeholders; normative and conceptual work

NIH RF1: "Human Agency and Brain-Computer Interfaces: Understanding users' experiences and developing a tool for improved consent"
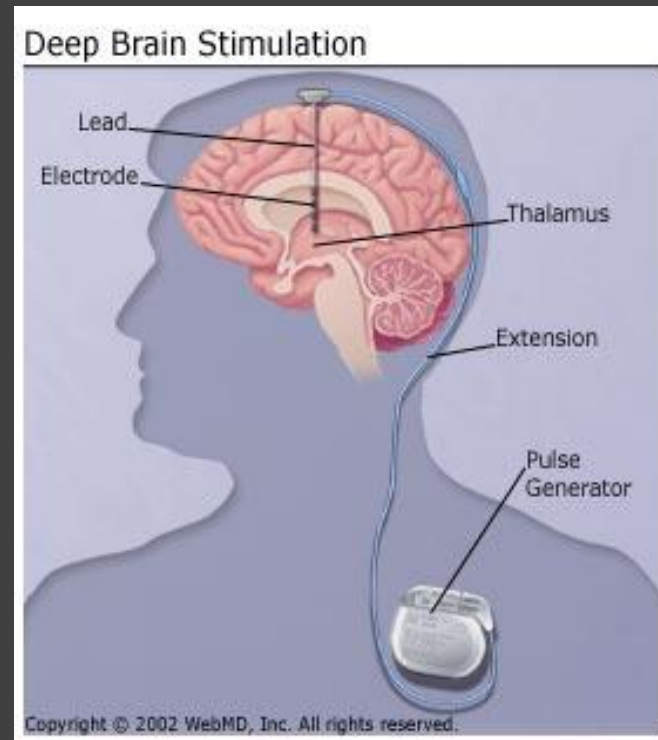
Co-PI with Eran Klein

What are neural technologies?

Brain computer interfaces

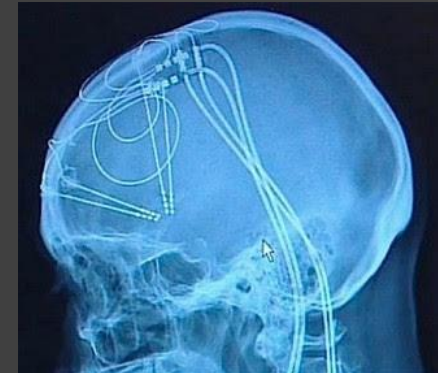Deep brain stimulators

What are the targets for use?

Motor disabilities (Parkinson's, essential tremor, spinal cord injury, stroke)

Communication disabilities; control over technology at distance

Psychiatric disabilities (treatment resistant depression, OCD, anorexia, etc.)

Why be concerned about neural technologies and agency?

They function within parts of our selves that seem central to our being *agents* with particular identities. Our brains serve as a kind of control center for our identities and initiation of action.

If they function well, they may restore an individual's sense of agency and identity, and operate relatively seamlessly and immediately. (Scheuermann; Hemmes in Fixed; Wolpaw story)

If they function poorly –sometimes as intended for the main target, but with side effects -- they may make us question our responsibility for action and change our understandings of who we are. ("Did my implant make me do that?")

Next generation devices

Seek to "close the loop" on neural interfacing:

recording from *and* stimulating the brain

providing sensory feedback

designed to be adaptive (with AI/machine learning)

may be *semi-autonomous* themselves

(Kellmeyer et al. 2016)

What is agency?

Initiating action

Acting intentionally

Doing vs. being acted upon

Feeling that we are "authors" in our lives (our actions flow from our beliefs and intentions)
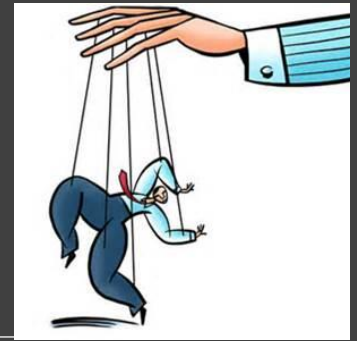
But also:

Need not be *conscious* initiation (e.g., habitual actions)

Need not be fully intentional (e.g., implicit bias)

Need not be free from outside influences (nudging vs. coercion)

Sense of agency?

A subjective awareness that "I am the one who is causing or generating an action" (Gallagher 2000, 15).

A problem – we can falsely have a sense of agency (Vlek et al. 2014; Haselager 2013; Wegner 2003, 2004).

◦ Illusion of control (we think we are acting when we have not acted; e.g., misleading visual cues)

◦ Automatism (a belief that we are *not* acting when in fact we are; e.g., some studies of facilitated communication that show guidance when the facilitator thought they did nothing).

The potential problem

"Since motor activity is basically controlled by electrical signals generated by computer algorithms, sometimes it may become difficult to decide **who the real agent is**." (Clausen 2008)

"[T]he device introduces **a third party** that comes between what was a direct connection between his brain and mind. This appears to diminish the causal role that he and the conscious mental states with which he identifies play in his actions. **The implanted device may threaten his conviction that he can control how he thinks and acts."** (Lipsmann and Glannon 2013)

We should be concerned about neurotechnology "insofar as it is **a threat to agency** – the ability to make informed and rational choices – as **when a person's actions do not flow from her intentions or beliefs but rather are the result of direct brain manipulation."** (Baylis 2013).

Empirical evidence that agency may be an issue

Early DBS studies on people with Parkinson's disease:

*"**I don't feel like myself anymore**"*

"I haven't found myself again after the operation"

"I feel like an electronic doll"

*"**I feel like a robot**"*

(Schüpbach et al. 2006)

Other reports of side effects such as increased impulsivity, hypersexuality, mania, depression, speech dysfunction, memory issues, etc.

Ambiguous or confused agency

DBS studies for psychiatric disabilities (e.g., depression and OCD)

"..there are parts of this where you just wonder how much is you anymore, ...'How much of it is my thought pattern? How would I deal with this if I didn't have the stimulation system?' You kind of feel artificial."

"I've begun to wonder what's me and what's the depression, and what's the stimulator. I mean, for example, I can be fine, and then all of a sudden... and, and I might realize it later, I do this, something that's socially or, or interpersonally, just not... just not right...I mean, there are three things – there's me, as I was, or think I was; and there's the depression, and then there's, uh, depression AND the device...and, it, it blurs to the point where I'm not sure... frankly, who I am."

"So, there's people in my family that (sighs)... sometimes question, you know, how much of it is me anymore and how much of me is, you know, bein' programmed. So that's been a, that's a hard thing to deal with sometimes, but, you know, like I said, I'll take it over what I had."

(Klein et al. 2016)

BCI and agency
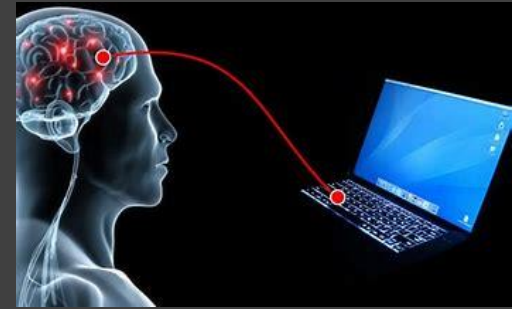
SUCCESS: BCI users who see the device's outcome match their intended action (e.g., touch the target, trace the path) feel responsible for that outcome. "I did that!" or "Over time, it's almost like a muscle in my mind and in my ability to push the cursor."

FAILURE: BCI users whose intentions are not enacted may be more likely to believe the *device* failed in some way ("errors occurred" vs. "I made a mistake").

UNCERTAINTY: BCI users who experience *difficulties* moving as intended are more likely to feel ambiguous about their responsibility:

> "I guess I didn't concentrate hard enough. But it also may be that the measuring is not optimal."
> "Did the device fail to recognize my mental imagery?"
> "Sometimes computers are dumb at figuring out what your brains are thinking….There are still so many points of failure that are possible in these systems."   (Kögel et al. 2020)
> "Is the device not 'tuned' well? -- It's pulling down and left"
> "It's a back and forth. Sometimes the computer is with me, is in sync with me and I don't have to give it a second thought. I just do it. Other times, the computer is not in sync with me and we have to retrain in order to get it as close to being in sync as possible. When it's not in sync with me, that's why I feel like I'm struggling, it's frustrating. I just want to move." (unpublished data)

BCI and agency

So far, there are fewer empirical studies of BCI users.

But in reading out from the user's brain to control an external device:

- The user's intention may be misidentified
- The device may fail mechanically
- With "smart" BCIs, an intended action may be "auto-corrected" (as with auto-spelling), with consequent errors
- Concerns about subconscious motives/intentions sparking activity
- Closed-loop DBS can also seem to "auto-correct" in its own way, by returning user to a set baseline (funeral scenario from Klein et al. 2016)
- Concerns about security/hacking

These concerns should give us pause in respect to how devices can be used to improve agency; however, we may also need to rethink what we *mean* by agency.

Relational agency
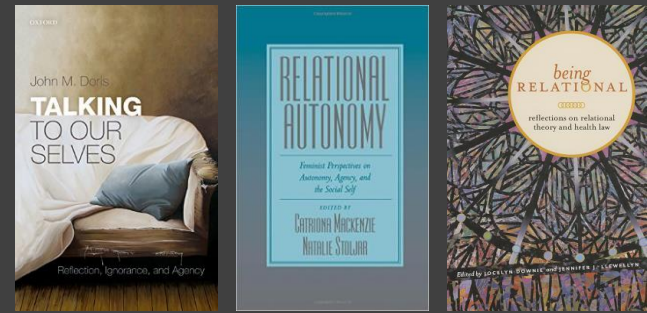
Maria Lugones (2003) articulates an "I<> we" behind agential action. Acting requires backup/intelligibility: "Intending may 'feel' as arising in a subject, but surely the production of intentions is itself a haphazard and dispersed social production. Subjects participate in intending, but intentions acquire life to the extent that they exist between subjects. The trick of individual intentionality lies in making one believe that if one intends into the hegemonic system's vein, one is the author of those intentions and of actions invested with their point."

Alisa Bierria (2014) – my action is always, in some way "socially authored"; "recognition of our intention becomes a part of the social choreography of agency" (131)

John Doris (2015) calls this "collaborativist agency" because we rely on others to help us exercise our agency. "…[T]he exercise of individual agency is a substantially interpersonal phenomenon." (103)

"For agency, the key word is *participant*. While it's true that people are subject to social influences that constrain their self-understandings, they are not passive entities in these negotiations….This is not to deny that inequalities of power may impede the exercise of agency; that's one thing that makes inequalities of power bad. But in other instances, the opposite is true: relationships help people express their values in their lives, as they do in the right sort of friendships, romantic involvements, and institutions. Agency requires not *freedom* from influence, but *mutual influence."* (Doris 148)

The relational agency of disabled people

This relationality of agency is well-known to disabled people, many of whom already use devices and/or caregivers to help enact their wills on the world.

> A person who relies on a caregiver for feeding does not fail to be an agent because she cannot lift a spoon to her mouth independently.

> She expresses her agency through choosing a caregiver and directing their action.

> "Independent living for disabled people means … having choices about who helps you and the ways they help. It is not necessarily about doing things for yourself, it is about having control over your day-to-day life" (independentliving.org).

> The person *directs* the action by making choices, even if the intention is put into effect with assistance from the caregiver.

Disability experience calls our attention to the social and environmental factors that influence how agency gets enacted (Timpe 2019).

Neural devices as relational others?

What happens if we frame neural devices as relational others? (Goering et al. 2017)

This might be a kind of co-agency.

Compare to wheelchair and wheelchair user (Belser 2016)

"Wheelchair Dancer often uses the language of "We/I" to underscore the partnership she experiences with her chair, phrasing that highlights the interactivity between wheeler and wheels … The complex pronoun "We/I" is emblematic of disabled relationality, a form of being-together in which the chair becomes part of the self while simultaneously existing as an independent other." (13)

Or consider a BCI user who sees herself in a "collegial" relationship with the BCI; it is more of a "friendly helper than just a technical thing." (Kögel et al. 2020)

Another BCI user likens working with the device to others: "That frustration you get when your instructor tells you that you have to do a group project and then you get that one person in your group who just doesn't pull his weight. … Well, that's kind of that frustration. Like I'm working here, and that computer is screwing up today."

Just as we collaborate with others to enact our agency, so too might we collaborate with *devices* to achieve our goals. The aim would be not full control by the individual, but a kind of co-authorship, or active participation in creating action.

Design for maintaining agency

Local control over neural technology by end-user (off or ramp-down switch; individual modulation within set parameters; preferences regarding auto-correct; possibility for "read-out" of device activity).

These may provide the user with a greater degree of control over the device, and more awareness of its influences, as well as the possibility of resistance.

As relational agents, we rely on others to help us identify when we are behaving differently. (Compare to a controlling intimate other; Goering 2014)

The "check" on whether our actions are "mutually influenced" (in Doris's term) in the right balance is likely tied back to the relational identities we negotiate.

Importance of feedback from end users and research participants

Different closed loop models; possibility of keeping the user "in the loop" (e.g., Goering et al. 2017, Kellmeyer et al. 2016, Brown et al. 2016) vs. having an autonomous device looping inside the person.

Need to know more from users whether having such control is preferable, acceptable, or worth the tradeoffs. This may vary across neural prosthetic modalities (e.g., motor vs. psych).

Our NIH project involves interviewing BCI users, over an extended period, about their experience using neural devices.
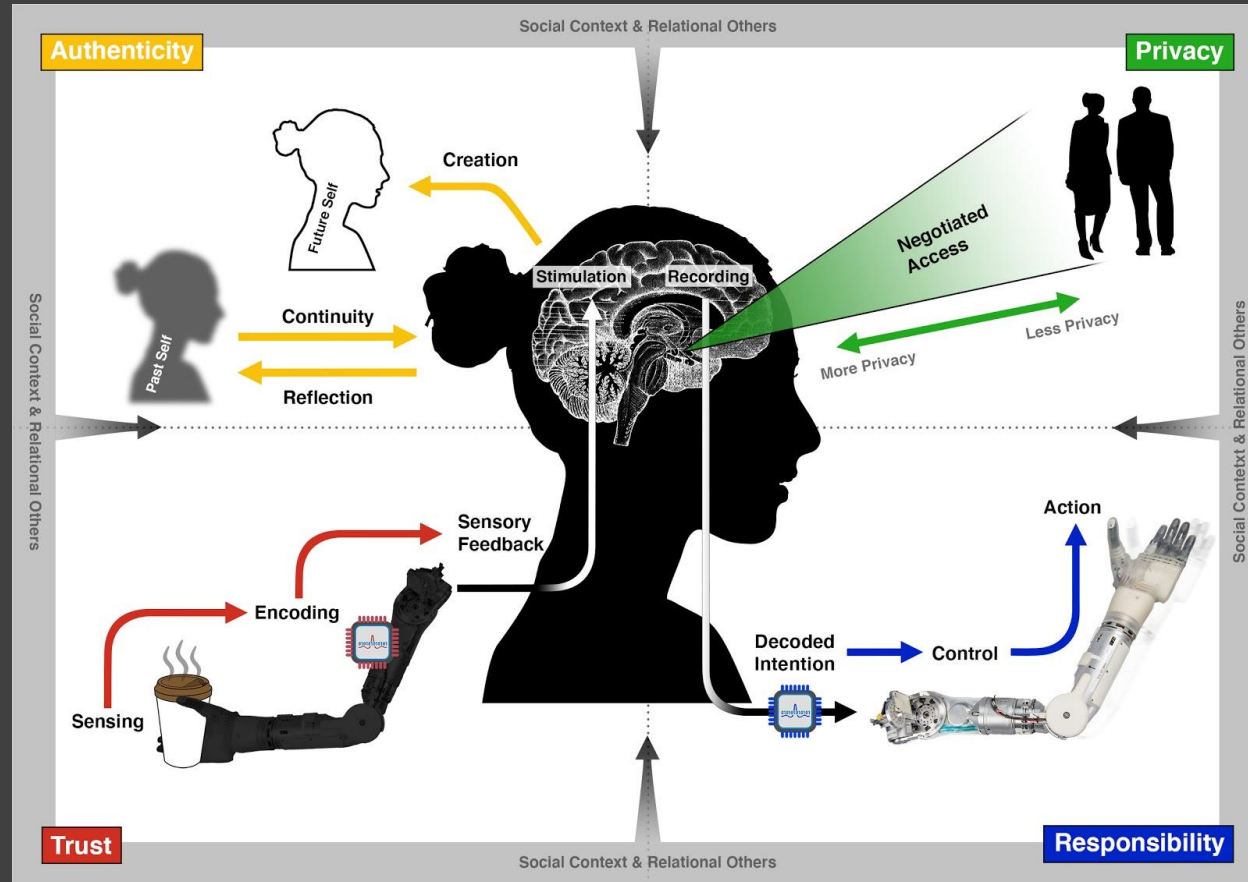
Conceptualizing (relational) agency

Four dimensions of agency:

Authenticity (Does this *feel* like me?)

Privacy (Who has access to information about me? How can I negotiate my own boundaries?)

Responsibility (Did *I* do that? Am I accountable?)

Trust (Did that really just happen? Can I trust myself to know if it did?)

Concluding thoughts

Neural devices offer opportunities to improve and extend agency.

Some empirical evidence suggests that even when the devices successfully address symptoms of concern, they create new uncertainties about agency and concerns about manipulation.

A relational framework for agency highlights how much we rely on others for agential assistance.

Relying or being dependent on others to enact our agency is not somehow a diminishment, but the normal state of affairs.

Device design – and relationship building – should focus on kinds of influence, ways to recognize that influence, and capacities for resistance.

The agency we seek to protect and secure is not a solo phenomenon, but a deeply relational one.

The promise—and threat—of neurotechnologies can help us recognize that reality.

References

Baylis, F. (2013) "I am who I am" *Neuroethics* 6: 513-526.

Belser, J. (2016) "Vital Wheels" *Hypatia* 31(1): 5-21.

Bierria, A. (2014) "Missing in Action" *Hypatia* 29(1): 129-145.

Brown, T. et al. (2016) "Controlling our Brains" *BCI* 3(4): 165-170.

Clausen, J. (2008) "Moving Minds" *Biotechnology Journal* 3: 1493-501.

Doris, J. (2015) *Talking to Our Selves* Oxford University Press.

Gallagher, S. (2000) "Philosophical Conceptions of the Self: Implications for Cognitive Science" *Trends in Cognitive Science* 4(1): 14-21.

Goering, S. et al. (2017) "Staying in the Loop" *AJOB Neuroscience* 8(2): 59-70.

Goering, S. (2014) "Is it Still Me? DBS, Agency, and the Extended, Relational Me" *AJOB Neuroscience* 5(4): 50-51.

Haselager, P. (2013) "Did I do that? Brain-computer Interfacing and the Sense of Agency" *Minds and Machines* 23: 405-418.

Kellmeyer, P. et al. (2016) "Effects of Closed Loop Control Medical Devices on Autonomy and Accountability of Persons and Systems" *Cambridge Quarterly of Healthcare Ethics* 25(4): 623-33.

Klein, E. et al. (2016) "Brain-computer interface-based control of brain closed-loop stimulation" *BCI* 3(3): 140-148.

Klein, E. et al. (2015) "Engineering the Brain" *Hastings Center Report* 45(6): 26-35.

Kögel, J. et al. (2020) "What is it Like to Use a BCI?" *BMC Medical Ethics* 21(2); doi: 10.1186/s12910-019-0442-2

Lipsmann, N. and Glannon, W. (2013) "Brain, Mind and Machine" *Bioethics* 27(9): 465-70.

Lugones, Maria. (2003) "Tactical Strategies of the Streetwalker/ Estrategias Tácticas de la Callejera." *Pilgrimages/Peregrinajes: Theorizing Coalition against Multiple Oppressions*. Lanham, Md.: Rowman & Littlefield.

Schupbach, X et al. (2006) "Neurosurgery for PD: A distressed mind in a repaired body?" *Neurology* 66(12): 1811-1816.

Timpe, K. (2019) "Moral Ecology, Disabilities and Human Agency" *Res Philosophica* 96(1): 17-41.

Vlek et al. (2014) "BCI and a User's Judgment of Agency" in *BCI in Their Ethical, Social and Cultural Context* (eds. Grubler and Hildt), Dortrecht: Springer.

Wegner, D.M. et al. (2003) "Clever Hands…" *Journal of Personality & Social Psychology* 85: 5-19.

Wegner, D.M. et al. (2004) "Vicarious Agency…" *Journal of Personality & Social Psychology* 86: 838-848.

Yuste, R. et al. (2017) "Four Ethical Priorities for Neurotechnologies and AI" *Nature* 551(7679): 159-163.

Questions?

Feel free to contact me:

sgoering@uw.edu

Thanks to my wonderful neuroethics research team:

Ishan Dasgupta, Eran Klein, Joe Stramondo, Erika Versalovic, Tim Brown, Andreas Schönau, Marion Boulicault, Natalia Montes, Paul Tubig, and Darcy McCusker.