JNphi

Herzog's Cartesian Theater: Skepticism, Self, and Stories in *Theater of Thought*

Peter Zuk

Abstract

This essay draws out some philosophical dimensions of Werner Herzog's recent film *Theater of Thought*, which interweaves an overview of contemporary neurotechnology development with consideration of classical problems in epistemology and the philosophy of mind. Over the course of the film, Herzog increasingly concerns himself with the challenge of philosophical skepticism. This preoccupation is framed as arising from his encounters, via certain of his scientist interviewees, with an outlook that I here term *neuroscientific postmodernism*. According to this outlook, modern neuroscience overturns our ordinary conception of ourselves as active, unified, conscious subjects of experience, and in so doing problematizes the application of the very concept of truth to the stories we tell about ourselves. As I interpret him, Herzog implicitly challenges this outlook on several fronts. His alternative emphasizes human spontaneity as it arises in everyday activities unmediated by digital technology, a reverential attitude toward consciousness and the mystery it presents, and a potential response to skepticism that invites comparison with those of Samuel Johnson, G.E. Moore, Ludwig Wittgenstein, and O.K. Bouwsma.

Key Words: neurotechnology, consciousness, illusionism, skepticism, transhumanism, Werner Herzog

DOI: 10.5281/zenodo.15860786

Introduction

In *Theater of Thought*, Werner Herzog supplies the viewer with a marvelous overview of contemporary neurotechnology development, covering a wide array of emerging technologies that includes wearable brain scanners, quantum computing, optogenetics, deep brain stimulation, augmented and virtual reality, brain-machine interfaces, mental state decoding, next-generation fiberoptics, and

Corresponding author: Peter Zuk
Address: University of Texas at Arlington, USA
e-mail ✉ peter.zuk@uta.edu

nanotechnology. Along the way, viewers are introduced to the neuroscience of embodied self-awareness, fear, olfaction, inner speech, and much else.

Of even greater significance than its pedagogical function are the various ways in which Herzog organizes the film and reacts to his subject matter during his own appearances within it. These elements give the film a genuinely philosophical character well worth exploring. I do not claim in the following to capture with any significant degree of reliability Herzog's actual intentions. Rather, I will simply state what seems to me a reasonable and cohesive interpretation of some key elements of the film while drawing out lessons that, intentionally conveyed or not, can be of use in considering the ethical, legal, social, and existential implications of the neurotechnologies and basic science that Herzog displays on screen.[2]

For Herzog, one of the most important aspects of neurotechnology is that it raises anew certain classic philosophical problems. Looming largest is the *epistemological* question of how to explain and justify both our self-knowledge and our knowledge of the world outside our own minds. This provides a throughline that connects several of Herzog's other major themes, which include the *relationship between brain, self*, and *narrative self-understanding*, how to interpret the idea of *technological progress*, and *wonder* at the profound mystery of consciousness. In what follows, I will consider each of these in their turn. First, however, I must say a word about Herzog's methodological approach.

Herzog's Methodology

Herzog has been known as a groundbreaking filmmaker for some time. One factor is surely his bold choice of subject matter, which in his documentary work alone includes topics as far-flung as Kuwaiti oil fields, volcanoes, capital punishment, Mikhail Gorbachev, the Internet, cave paintings, and a man who lived (and died) among grizzly bears. A second factor is the suffusion of Herzog's films with his own existential sensibilities. Both factors are on full display in *Theater of Thought*—the former in the sophistication of the technologies countenanced and the exotic nature of some of their applications, the latter through Herzog's own dialogue and directorial choices.[3]

Indeed, in many segments Herzog affects a posture that I can only liken to the *Socratic ignorance* professed by its namesake in many of

[2] As this film has not yet been widely released, quotations are drawn from my notes taken while viewing it rather than an official transcript. While I have endeavored as far as possible to capture each quote verbatim, I must ask both readers' and speakers' forgiveness for any inadvertent errors or paraphrases.

[3] I restrict my analysis to this particular film, leaving for another occasion the well worthwhile task of relating it to the rest of Herzog's body of work.

Plato's earlier dialogues.[4] As with Socrates, Herzog professes a lack of knowledge on fundamental human questions (here, the workings of the brain and mind) and seeks answers from those who do profess to know. This apparent naivete belies an intellectual depth from which proceeds sharply provocative questions and statements intended to probe what the experts really do know. Among the most interesting are:

> "Can you compute the human brain?" (to Darío Gil)
>
> "Telepathy is possible now." (to Jamie Daves)
>
> "Would Siri understand 'night' versus 'knight'? How stupid is Siri?" (to Tom Gruber)
>
> "How much saltwater is in your brain?" (to Ken Shepard)
>
> "But you are creating [neural] fiberoptics. Would you be daring enough to insert it into the brain of your little child?" (to Polina Anikeeva)

Herzog is forthright about not knowing quite what to make of the technologies to which his interviewees introduce him. Of computer scientist and electrical engineer Darío Gil's explanations of quantum computing, for instance, he says, "I admit that I literally understand nothing of this." Things have not improved much in this regard even by the end of the film, when he remarks that "my voyage into our mind ultimately leaves me more mystified." These professions notwithstanding, Herzog clearly has an incisive grasp of neurotechnology's potential applications and the consequences of these for human life. He is thus able to offer viewers an impressive array of critically important questions to ponder, even if he largely holds off on furnishing answers.

Herzog's explicit statements are coupled with a number of similarly provocative directorial choices, such as interviewing neuroscientist Jack Gallant in the dingy projector room of the Roxie Theater in San Francisco. There are also (judging by their reactions) several shots in which the camera lingers on interviewees for an uncomfortably long period of time after they have finished speaking, as well as an extended sequence in which materials scientist Polina Anikeeva walks through a series of corridors while eerie music plays. (There will be more to say about the potential significance of these later.) The overall effect is a scientific and philosophical tour de force. But what exactly might Herzog be communicating through all of this?

[4] e.g., "I am wiser than this man; it is likely that neither of us knows anything worthwhile, but he thinks he knows something when he does not, whereas when I do not know, neither do I think I know; so I am likely to be wiser than he to this small extent, that I do not think I know what I do not know" (Plato, *Apology* 21d).

Body, Mind, and Self

Unsurprisingly, a core theme of the film is the relationship between the brain, the mind, and the self. One increasingly influential way of characterizing this relationship has come to be called *neuroessentialism*: the claim that "our brains define who we are" such that "in investigating the brain, we investigate the self" (Roskies 2002). Rodriguez (2006) argues that the rise of neuroscience has led to the displacement of traditional folk psychological categories by reductive explanations of behavior and mental states in terms of the brain and its processes. According to this perspective, our understanding of the mind and the self must be understood in the terms of neuroscience, which has special status as a privileged level of explanation.

We encounter several instances of this broadly neuroessentialist thought-style in the film. Early on, neuroscientist Christof Koch gives a neurobiological definition of the concept of humanity: "What makes humans human" is the cortex. He then goes on to say:

> The central mystery of the ancient mind-body problem is: how does thought, how does conscious[ness], how does color and motion and pain and pleasure and love and hate, how does that emerge from this tissue? Because we know that if you stimulate this tissue, you can get all the thoughts and pain and love and hate, so how does it happen?

We also hear from neurobiologist Richard Axel, a great lover of music whose scientific work has focused on the neurobiology of olfaction. In response to Herzog's question about whether feelings are irreducible to something more fundamental, he says:

> I have a sense that feelings are far more complex. They are reducible. They are a consequence of the way your past has generated your present mind. And the way you respond to music and the way your respond to smell calls forth the experiences and feelings of your past. And so they are indeed reducible.[5]

There is at least a hint of lockstep causal determinism here, as if the content of the present mind is inescapably decided by the past regardless of its activity now. Most outspoken on this matter is Jack Gallant:

> Our conscious experience of the world and what we think is happening and what we think we're thinking about and what we think our motivations are and our judgments for our actions and our plans and goals, a lot of that is **a constructed narrative we consciously create to explain our daily behavior** (my bolding).

[5] It is worth noting, however, that Axel may not mean by this that they are reducible to *neurobiological* processes; the context leaves it unclear.

It appears that, for Gallant, much of our conscious experience is *post hoc* rationalization of behavior. Rather than guiding or accompanying behavior, conscious experience is here claimed to come in only after the fact, too late to the party to be hosting it. This, Gallant thinks, calls into question the accuracy of our conscious experiences and even the very application to them of the concept of truth. "The human brain really never contains any truth. It's always shades and nuances that are all reflections of the prior history of the person who experienced them," he continues. According to this outlook, which I will term *neuroscientific postmodernism*, our ordinary conception of ourselves as unified conscious subjects who think, feel, plan, and act is overturned by neuroscientific findings, and with this the applicability of the concept of truth itself to our conscious experience is called into question.[6]

Lyotard (1979, xxiv) (in)famously defines the postmodern as "incredulity toward metanarratives."[7] In the time since Lyotard wrote, it has become a platitude of postmodernity that each of us is henceforth free to author our own story rather than being forced to subsume ourselves under one or another grand legitimating metanarrative. Many would simultaneously regard postmodernity as synonymous with the "post-Truth" era (at least when truth is given its traditional capital "T"). We see in Gallant's remarks above that this idea possesses not only general cultural currency, but also a significant amount of scientific purchase. Neuroscientific postmodernism challenges what we might think of as the most fundamental metanarrative of all: the existence of an active, causally efficacious, conscious subject of experience.

Neuroscientific postmodernism thus has an overtly skeptical character toward the categories of everyday experience, making explicit that the "construction" of our own individual narratives entails their fictionality.[8] Earlier in the film, Herzog asks, "Do we invent our own stories? Do we make up our lives?" Neuroscientist Joseph LeDoux replies in the affirmative:

> What we are doing constantly is generating a mental model of the world, and so our minds are constantly narrating our lives. And that's the story we know about ourselves and we tell to ourselves and we tell to others. And that's who we are. You know, war heroes would go around the country after a war telling their story. And

[6] I propose this as a working definition only; a complete account of the phenomenon of neuroscientific postmodernism and its relation to other viewpoints in the same vicinity must await another occasion.

[7] Intriguingly, especially for our purposes, he then continues, "This incredulity is undoubtedly a product of progress in the sciences: but that progress in turn presupposes it" (Ibid.).

[8] In good old-fashioned regular postmodernism, by contrast, the fact of the fictionality of our narratives is often left implicit (perhaps due to its uncomfortably paradoxical character).

> each time the story got a little better every time it was told. So our narratives change every time we tell it.

The idea that we invent our own stories and change them in each telling carries the connotation that they are works of fiction, as Gallant similarly intimates. Neuroscientific postmodernism thus puts something of a damper on the liberatory outcome that was supposed to have been ushered in by good old-fashioned postmodernism. The neuroessentialist outlook with which we began the discussion now turns out to lead us down a road at the end of which lies its own subversion. Instead of our brains defining who we *are*, they instead come to define who we are *not*, since the self so defined is ultimately illusory.

This is at least one way of taking the various statements of Koch, Axel, Gallant, and LeDoux included in the film.[9] Does the film offer some alternative to this picture? Herzog speaks freely of the soul throughout it, suggesting that he believes the demise of the traditional notion of the self has been greatly exaggerated. While he does not explicitly formulate an alternative thesis, he at least gestures toward one. Reacting to Gallant's impressive neural decoding work, Herzog remarks, "Where our numbers and names and concepts are located can be mapped [in the brain]. But there is no map of our thoughts." Later, he asks neuroscientist Uri Hasson, "Could you in the future read my mind and read my new movie before I even film it?"[10] Herzog is apparently wondering whether there may come a day when there is no need for him to engage in the creative activity of filmmaking. In this imagined future, rather than exercising his own unpredictable spontaneity, the film he *would have* made can simply be decoded directly from his brain. Impressive as this hypothetical achievement might be, Herzog's implication seems to be that this gain in efficiency would be purchased at the loss of human agency and its accompanying significance.

If so, he is in good company. Arendt (1958) views spontaneity, understood as the ability to initiate something new, as a (perhaps even *the*) most fundamental element of our humanity. For Arendt, *acting*—in her special sense of the term which defines it in terms of initiating something new—and *speaking* in the public realm are the basis of a characteristically human life:

> A life without speech and without action… has ceased to be a human life because it is no longer lived among men.
>
> With word and deed we insert ourselves into the human world, and this insertion is like a second birth, in which

[9] Though I should hasten to emphasize that these are short soundbites and thus in themselves evince (at most) a general intellectual attitude or methodological outlook, rather than a fully worked-out theory on their speakers' part.

[10] Herzog also says something similar to Gallant ("You could decode my movie that's only in my head"), but this is presumably meant non-committally given the skepticism he soon expresses at the end of the segment (i.e., "[T]here is no map of our thoughts").

> we confirm and take upon ourselves the naked fact of our original physical appearance (Ibid., 176-77).[11]

For Arendt, it is speech and action (in her special sense) that disclose not merely *what* a person is (i.e., traits, abilities, and biographical and other information), but *who* (Ibid., 179). It is the latter in which our true individuality consists, but this cannot be captured propositionally in a collection of statements about us, no matter how detailed. It is instead "implicit in everything somebody says and does" (Ibid.). And it is just this public disclosure of the (strictly speaking) unutterable "who" in our words and deeds in which the *self* consists.

This self-disclosure is not the same as *authoring* our own life stories. On Arendt's view, our authoritative life story is the narrative that will (if we are lucky) be preserved in the public realm after we are gone.[12] While this narrative must have its basis in our actual words and deeds, how *exactly* we are remembered is not under our control. As she puts it, "Even though stories are the inevitable results of action, it is not the actor but the storyteller who perceives and 'makes' the story" (Ibid., 192). Importantly, this provides an intersubjective bulwark against the skeptical implications of neuroscientific postmodernism. Instead of an individually constructed fiction, the story of each of our lives is the *collective* memory of our individual spontaneity.

In this connection, one wonders (very tentatively) whether Herzog has something like Arendtian spontaneity in mind when he leaves the camera running to obtain uncomfortably long shots of certain interviewees (LeDoux, Shepard, and Goering) after they have finished speaking. This is a particularly poignant device in LeDoux's case, as he has just finished discussing the concept of fear. On one interpretation, Herzog might here be cajoling his interviewees into acting spontaneously after they have voiced their (presumably to some degree prepared) remarks. The results are quite interesting, and I will make no attempt to interpret the countenances of the various subjects with whom Herzog tries this tactic. Instead, I will simply register my suspicion that these attempts to capture candid moments are an implied pushback by Herzog on the neuroscientific postmodernism we hear voiced at certain points in the film. By inviting some of his subjects to go off-script in spontaneous Arendtian fashion, Herzog is inviting the disclosure of the personal *who* as opposed to the lockstep-determined *what* of the neuroscientific postmodern outlook.

[11] See also Arendt (1951, 438), according to which the elimination of spontaneity has the effect "of transforming the human personality into a mere thing." It should be noted that I am taking some liberty in relating Arendtian spontaneous action to Herzog's discussion of his filmmaking, since art is for Arendt the product not of action (in her special sense of the term) but of *work* (fabrication), which she distinguishes from action. See however Lewicki (2025)'s contention that the most consistent version of Arendt's view will include a category of art-action in addition to art-work.

[12] On this point, she is explicitly invoking the ancient Greek view of immortality through remembrance.

This is a good start in articulating a reasonable alternative to neuroscientific postmodernism. Something like an Arendtian intersubjective public realm is also relevant to the most prominent of Herzog's themes—philosophical skepticism—to which we can now turn.

Knowledge and Illusion

After encountering some of the expressions of neuroessentialism and neuroscientific postmodernism catalogued above (though not yet Gallant's), Herzog becomes noticeably preoccupied with the venerable epistemological questions of what we know and how we know it. After hearing about bioengineer Karl Dessiroth's work to control the mental states and behavior of mice using optogenetics—a powerful example of just how much sway targeted neuromodulation can possess over the mind—Herzog asks, "Could it be that we live in some kind of invented reality?" He asks a similar question of Anikeeva: "Could it be that in our thoughts, in our self, we are in some sort of theater of thought, in an imaginary world that's only existing in our brain?"

This provocative reference to the title of the film leads to an extended dialogue in which Herzog issues a skeptical challenge for Anikeeva to answer. "Is the building behind you real?" he asks. She responds:

> I think this building is real. I kept walking up and down those steps for the past decade. I feel the, sort of the stiffness of the stone. I hear the sound of my shoes clicking on that stone. I can get to my office through those stairs.

For Anikeeva, the evidence of her senses provides a literal *ground truth*, as it were. Perhaps anticipating familiar philosophical worries about the reliability of the senses, however, she continues:

> There are other people [who] seem to be perceiving the same reality. But I don't have a frame of comparison. I haven't had a chance to perceive an alternate reality, so for me this is the only one that exists.

Here we are offered an appeal to *intersubjectivity*: it is not just Anikeeva's senses, but also the senses of others, that testify to the reality of the world. There is perhaps an echo of Kant's critical philosophy here, to which we will return later.

Epistemological questions of this kind have gained a good deal of cultural traction in recent decades thanks to films like *Dark City* and *The Matrix* tetralogy, as well as philosophical treatments of their themes such as Bostrom (2003)'s *simulation hypothesis* (according to which it follows from certain plausible assumptions that we likely live in a digital computer simulation). The countenancing of these kinds of skeptical possibilities will of course forever be associated with René

Descartes. Indeed, Descartes receives an explicit nod from philosopher and bioethicist Sara Goering late in Herzog's film. And the very idea of a "theater of thought" immediately calls to mind a favored metaphor of some of Descartes' most important critics (namely, Gilbert Ryle and Daniel Dennett).

Recall then Descartes' method of hyperbolic doubt in the *Meditations*, according to which he (temporarily) rejects any of his beliefs that he can find grounds for doubting. When Descartes applies this method systematically, his belief in the reliability of his senses, of the sciences, and of mathematics, as well as the existence of his own body and of any external world whatsoever, each fall in their turn. But Descartes claims to find an Archimedean point, rock-solid and certain, that he can use as a foundation for a rational reconstruction of human knowledge. This certainty is the famous *cogito ergo sum* ("I think, therefore I am"). From there, Descartes argues for the existence of God and God's non-deceptive nature, on the basis of which he claims to win back reliable contact with the world beyond his mind.

It is difficult, however, to dispute Arendt (1958, ch. 6)'s contention that Descartes' formulation of the *problem* of philosophical skepticism has been far more influential than his proposed *solution*. Many contemporary philosophers will of course grant the wisdom of the profound first step of the Cartesian solution (the indubitable *cogito*). But far fewer endorse as sound the arguments for the existence of God that follow and on the basis of which Descartes claims to overcome skepticism. If the first step succeeds but not these latter ones, Descartes gives us back the conscious thinking subject but not a world outside it. Experience gives us the *appearance* of such a world—a "theater of thought" as we might say following Herzog. But the correspondence of those appearances with a reality beyond them will, if Descartes' arguments about God are not accepted, need to be secured on some other grounds if it is to be secured at all.

Not all contemporary philosophers are sold on even Descartes' basic presuppositions about the mind, however—especially when he is interpreted as claiming that we possess infallible or at least indubitable access to our own stream of consciousness.[13] They would thus deny that he has secured for us the existence of a unified, conscious subject of experience, let alone a world beyond. Some go so far as to deny that there even *is* such a thing as a unified stream of consciousness. Dennett (1991) famously voices his doubts as follows:

> There is no single, definitive "stream of consciousness," because there is no central Headquarters, no Cartesian Theater where "it all comes together" for the perusal of a Central Meaner [who could determine and report on its meaning]. Instead of such a single stream (however wide), there are multiple channels in which specialist

[13] Whether or not Descartes actually defends this doctrine is a matter for another time.

> circuits try, in parallel pandemoniums, to do their various things, creating Multiple Drafts as they go. Most of these fragmentary drafts of "narrative" play short-lived roles in the modulation of current activity but some get promoted to further functional roles… (Dennett 1991, 253-54).

On Dennett's view, a single stream of consciousness would require the existence of a "Cartesian Theater": an internal space where conscious experiences are presented, as though on a movie screen, to an internal "viewer."[14] But there is no such space in the brain, Dennett claims, and thus there cannot be one in the mind—unless we embrace Cartesian dualism about the mind-body relation, an option which is on Dennett's telling so beyond the bounds of philosophical respectability as to be no option at all.

On Dennett's alternative account, "multiple drafts" of a narrative or interpretation of stimuli from the environment compete to determine mental and behavioral response. We humans "are almost constantly engaged in presenting ourselves to others, and to ourselves, and hence *representing* ourselves — in language and gesture, external and internal" (Ibid., 417). In this regard we are unlike most other animals. "*Our* fundamental tactic of self-protection, self-control, and self-definition is not spinning webs or building dams, but telling stories, and more particularly concocting and controlling the story we tell others — and ourselves — about who we are," says Dennett (Ibid., 418). But the process of generating these stories is messy and fragmentary, with multiple stories competing within the brain to be the one told. The self turns out to be what Dennett calls a *center of narrative gravity*: the lead character in the stories that win out.

All of this, however, happens largely beyond our conscious control: "Our tales are spun, but for the most part we don't spin them; they spin us. Our human consciousness, and our narrative selfhood, is their product, not their source" (Ibid.). Moreover, centers of narrative gravity are, like the centers of gravity employed by physicists, theoretical fictions. The self is not real, or at least not "*really* real" (Ibid., 429). Dennett is on this basis plausibly construed an adherent of neuroscientific postmodernism, and indeed, one of its foremost proponents.[15] On his picture, the fact of consciousness itself will be only a narrative that we tell ourselves, and ultimately an illusion. So

[14] Dennett's teacher Ryle (1949, 207-08) uses the metaphor this way: "We do not, consequently, have to rig up one theatre, called 'the outside world', to house the common objects of anyone's observation, and another, called 'the mind', to house the objects of some monopoly observations."

[15] It is fair to suspect that, given Dennett's influence on cognitive science, his ideas may well have informed the views of some of the interviewees with whom Herzog speaks. But it should be noted that Dennett himself disavows any allegiance to postmodernism, at least if this is understood to imply unfettered relativism or the lack of a gap between appearance and reality (Dennett 2000). Whether this disavowal is in fact consistent with the views expressed in *Consciousness Explained*, however, is open to serious challenge.

much then for a "theater of thought," that is, for consciousness understood in anything like the usual way.

Unless, that is, we flip the script on Dennett's story by noticing that any attempt to tell it puts consciousness right back into the story. Consciousness as standardly understood is for Dennett an illusion. But even the *illusion* of consciousness would itself have to be an instance of consciousness, the very phenomenon being claimed illusory (Strawson 2018). Illusions are real to the extent that they are real *appearances* to an experiencing subject, even if what appears *in* them does not exist. But there is in consciousness no appearance-reality gap. To use one of Strawson's examples, seeming to be in pain is to be in pain. The intimacy with which we encounter our own consciousness brings, as Descartes saw, a special form of epistemic security (even if not infallibility). As Chalmers (1996, xii) says, "we are surer of the existence of consciousness than we are of anything else in the world." And to paraphrase how a friend once put it to me, "*Of course* we live in a Cartesian theater. The really interesting question is what's going on in here."[16]

And just what exactly *is* going on in here? The Cartesian Theater, we can notice, is a theater of a very strange kind, containing no seats, stairs, walls, fire exits, nor indeed any patrons or even a viewing screen. It is instead radically *immersive*: the "viewer" coincides, at least to a significant extent, with the experienced presentations being "viewed." While it is an interesting question whether the viewer consists *only* in the flowing-along sequence of presentations, we need not settle that here to affirm an intimate connection between them, such that Dennett's image of the Cartesian Theater as containing a separate *homunculus* who serves as the viewer is inapt. We can think instead of consciousness (and hence the viewer) as the film itself, or perhaps the film *plus* the screen upon which the various appearances are projected.[17]

Regardless of the precise metaphors we choose, as Herzog's literal film progresses we see him increasingly gripped by the idea of a theater of thought. It becomes for him the object of a kind of reverence, and this reverential status turns out to warrant some degree of caution in the face of technological change. We can now turn to the basis of this caution, after which we will return to the matter (so far left unsolved) of how the mind can be in reliable contact with the world.

[16] Carl Feierabend (in conversation).

[17] Lest we think the dependence on metaphor too great here, we can recall some lines from the final page of Dennett's book: "It's just a war of metaphors, you say — but metaphors are not 'just' metaphors; metaphors are the tools of thought. No one can think about consciousness without them, so it is important to equip yourself with the best set of tools available" (Dennett 1991, 455).

The Meaning of Progress

After listening to Gil's explanation of quantum computing, Herzog asks, "Would it be good that we can still go out and fish a trout?" Gil quickly agrees that this is "existentially important." Herzog repeatedly emphasizes the grounded activities of everyday life (e.g., Koch's boat-rowing, LeDoux's rock band, and Gil's fishing). The choice to interview non-scientist Philippe Petit on the topic of fear is also telling in this regard. Appearing just after LeDoux, a leading expert in the neuroscience of fear, Petit makes for a fascinating foil. What he lacks in scientific credentials Petit more than makes up for in his practical experience overcoming fear, having walked a tightrope between the two main towers of the World Trade Center in 1974, among other feats of daring.

These richly human (and in Petit's case, borderline superhuman) activities serve as a contrast to the neurotechnologically augmented modes of engaging with reality that several of the film's interviewees believe are on the horizon. Wearable neurotechnology is shown early in the film by entrepreneur Bryan Johnson. Johnson's device utilizes functional near-infrared spectroscopy (fNIRS) and is claimed to be capable of sophisticated mental state decoding (with neuroscientist, bioethicist, and chief scientific advisor of the film Rafael Yuste volunteering to demonstrate). Johnson has big plans for his device: "We will make it light enough so that you can wear it all day and low cost enough so that everyone will have one."

Herzog seems ambivalent about such prospects. After speaking with Dessiroth, he laments, "Some adolescents prefer a world of pure [digital] fantasy to the world just outside their doorstep." A similar prospect is raised later in the film by computer scientist Rajesh Rao's discussion of creating an augmented or fully virtual world, including artificial sensations, by stimulating the brainstem.[18] Following this segment, Herzog reminisces about Alexander Dovzhenko's 1930 film *Earth*, which depicts a dying man who intends to report, following his departure, on whether he has reached Heaven or Hell. Applied to our present context, the allegory might well be this: only once we have ushered in a neurotechnologically-driven future of the kind envisioned by Johnson or Rao will we truly be in a position to judge whether such a scenario is heavenly or hellish. This would seem to be quite the gamble when we could instead (unlike Dovzhenko's dying man) simply forego finding out altogether.

After this digression, Herzog's drive to the next filming location prompts some further reflections. The dull built environment surrounding a freeway overpass leads him to muse as follows:

> Could the First Nations, the natives of this land, have foreseen this? Where is the prairie, where did the buffalo

[18] Here it is difficult not to again recall the *Matrix* films, and with no small degree of apprehension.

> go? What kind of trees are these to the left and the right?
> The New Jersey turnpike allows no hint of the past.[19]

This suggests a subtlety worth noting: if by progress we mean acknowledgement of the past even while claiming to improve upon it, 21st century technocracy may not offer this progress to us. What we are increasingly offered might instead be construed as Herzog construes the New Jersey turnpike: a radical, technologically-mediated break with the past.

Nowhere is this tendency embodied more clearly than in the ongoing rise of *transhumanism* and concomitant movements' aim to not merely improve the human condition, but to transcend it altogether by modifying human nature in ways that will usher in a posthuman future (Bostrom 2005). Like the Tyrell Corporation in *Blade Runner*, the stated end goal of this movement is to produce beings who are "more human than human." While transhumanism receives no explicit acknowledgment in Herzog's film, it stands to reason that his pessimistic outlook on the various technologies we do see extends to the more radical approach that transhumanists envision. As he says of Gil's quantum computer, "I'm asking myself: have the creators of it lost themselves in a world of pure science fiction?"[20]

Perhaps so. True to Socratic form, at the end of the film Herzog remains unconvinced by the many knowledge claims of his interlocutors. "Of all the scientists we met, there was not a single one who could explain to us what a thought is, or what consciousness is," he remarks—while noting that they were sensitive to the ethical considerations raised by their work, which the final segments of the film consider.[21] Seen in this light, a previous portion of the film (and by far its most striking) takes on added significance.

Consciousness as a Mystery

This portion begins as Herzog interviews computer scientist Tom Gruber. After a lively exchange about whether the iPhone's Siri voice recognition can distinguish between "night" and "knight" (as Herzog correctly predicts, it cannot) Gruber explains the complexity of fish schools' swimming behavior. Behind them plays video footage of a school of fish which, we are told, was shot by Gruber during a dive. The striking scene distracts Herzog from their conversation, and we

[19] The "trees" in question are power line towers.

[20] In this connection, an apparently offhand comment by Herzog elsewhere in the film has striking prescience if we extend it to the rapid rise of generative artificial intelligence in our present historical moment: "The parrot is not actually in dialogue with us even though the parrot speaks."

[21] For reasons of space and thematic unity, I have (perhaps somewhat surprisingly) had little say about the prominent bioethicists who appear in the film. This should in no way be taken as disinterest in their contributions. Various of the ideas they express in the film are drawn from their own scholarly work, which I have engaged with in, e.g., Zuk (2024).

the viewer leave Gruber behind as the film drifts off into an extended reverie narrated by Herzog (complete with matching visual imagery):

> This looks like a fantasy world. This looks like out of a movie.
>
> Landscapes always played an important role in my films—in particular, landscapes of the soul. Do fish have souls? Do fish have dreams? Do they only dream this landscape? Do they think? Do they have thoughts at all, and if so, what are they thinking about? Is the same thought simultaneously in all of them? Is this what they dream collectively? Could they have visions of albino flowers that do not exist underwater?
>
> Are we floating in a strange and beautiful reality? Do we dance in our minds? Is this Paradise, where there are no boundaries anymore?
>
> We can observe the dance of fireflies. But their choreography is beyond our comprehension. For the firefly, the full moon casts its shaft of light on the floor of the forest. We can only stand by and marvel.

If we take this as a metaphor for the brain, we arrive at something like the following thought. We can observe the brain, and indeed, see it alight with activity (on, e.g., an fMRI). But the details of its processes and how they give rise to consciousness (the "dance" of the mind) are beyond our ken, at least for now. The appropriate reaction is astonishment at a mystery we do not fully comprehend and a corresponding posture of intellectual humility.[22]

Herzog here seems to be gesturing toward what Niikawa & Kriegel (2025) call "the sublime of consciousness."[23] This is the idea that, much as it is fitting to feel awe in response to viewing certain works of art and striking natural phenomena, so too can consciousness itself be the proper object of awe (and thus sublime). They highlight in particular the ungraspability of consciousness: how it can be that a world ostensibly composed purely of physical objects and processes also contains conscious subjects of experience (i.e., Chalmers (1996)'s hard problem). The recalcitrance of the hard problem of consciousness, together with a few other conditions,[24] lead Niikawa and Kriegel to affirm that we should indeed regard consciousness as a sublime object to be grouped with other, more traditional members of that category.

[22] Another pertinent moment of the film is the segment with brain-machine interface pioneer Eberhard Fetz, who switched fields from physics to neuroscience after having "the epiphany that consciousness was really the most incredible thing" in connection with psychedelic experience.
[23] See also Niikawa (2025).
[24] These are what they describe as the experience of the *overwhelming* significance of consciousness and the affective *ambivalence* we feel in contemplating it (a combination of "existential delight" tempered by frustration at our inability to grasp how consciousness exists).

Awe, astonishment, reverence—this is the family of attitudes that Herzog seems to be suggesting we should take toward the mind and its embodied activity in everyday life. This represents a clear counternarrative to neuroscientific postmodernism. As Herzog says after viewing a brain surgery performed by Edward Chang, "For me, most astonishing was to see the brain pulsing with a heartbeat." What better image could one imagine of the entanglement of the brain (and mind) with the rest of the body? This entanglement has a fascinating dual nature: there is the profound mystery raised by mind-body union, yet also the fact that, however philosophically perplexing it may be, it is for each of us an undeniable lived reality.

No Illusions?

It is perhaps most appropriate to conclude by finishing the discussion of our knowledge of that lived reality, which was previously cut short by our Dennettian digression. We can now return to Herzog's dialogue with Anikeeva and its emphasis on direct contact with the world around us, most notably the ground beneath our feet. There is a clear echo here of Samuel Johnson's so-called "appeal to the stone," which is considerably more succinct. Johnson, perturbed during a discussion of Berkeley's metaphysical idealism, is said to have exclaimed "I refute it *thus*" as he forcefully kicked a large stone (Boswell 1791, 143). Through this gesture, Johnson asserted the primacy of common sense and direct experience over what he viewed as metaphysical extravagance.[25]

There is here also something of a parallel with Moore (1939, 166)'s response to the skeptical hypothesis that he might lack a body. According to Moore, he both knows that he has a body and can prove it by holding up his hands while uttering, "Here is one hand… and here is another." And we can compare also Wittgenstein (1969, §151), commenting on Moore's proposal: "I should like to say: Moore does not *know* what he asserts he knows, but it stands fast for him, as also for me; regarding it as absolutely solid is part of our *method* of doubt and enquiry." While the various proposals here all differ in their philosophical details, each shares with Herzog an emphasis on the importance of remaining (quite literally) *grounded* in the world of sense-experience unmeditated, at least to some extent, by technology.

A more direct response to the specifically *Cartesian* version of the skeptical problem, and one with added relevance to Anikeeva's outlook, is offered by Bouwsma (1949). On Bouwsma's view, the world could be an *illusion* of the sort countenanced by Descartes only if there were some possible experience that would reveal this fact. Illusion presupposes the potential, at least in principle, of discovering the illusion. If there is no observation we could possibly make that would

[25] Of course, Berkeley never denied the existence of stones nor denied their materiality: he insisted only that *being a stone* and *being material* were, like everything else, themselves ideal.

falsify the world, then the world is real in the only legitimate meaning we can afford the word *real*.

As Anikeeva puts it above, "I haven't had a chance to perceive an alternate reality, so for me this is the only one that exists." Bouwsma adds that we could *never* have such a chance. There is no possible world-falsifying observation, no stepping entirely outside of both our individual and collective subjectivity to compare the theater of thought with a fully mind-independent reality (the point on which Kant insisted most emphatically as the cornerstone of his critical philosophy). The world, like consciousness itself, could *not* be a mere illusion, albeit for different reasons.

It is of course debatable whether this line of argument really solves the problem, and also whether Herzog himself thinks it does. The enigma of Herzog's intentions serve as a fitting counterpart to the philosophical enigmas with which he engages, and this is nowhere more manifest than the closing image of the film. Having just heard from human rights lawyer Jared Genser about the burgeoning neurorights movement, we see soldiers marching in footwear that clacks loudly against the stone surfaces on which they walk. This is clearly a callback to the discussion with Anikeeva and eerie sequence that followed. But what exactly is it intended to convey? Is it a nod to the Bouwsma-style solution to skepticism? Given the air of mystery Herzog creates around Anikeeva's habitual walk, is this instead a way of emphasizing the skeptical problem's intractability? Is it an artistic way of capturing the Arendtian insight of the fundamental *ambiguity* of our post-Cartesian plight? I can only follow Herzog's own lead in leaving this as an exercise to the viewer.

Acknowledgements

I thank the NeuroRights Foundation for providing me with access to Herzog's film for the purpose of writing this essay, and I thank Nyssa Juneau for her helpful comments and suggestions.

Conflict of Interest Statement

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Arendt H. *The Human Condition*. 2nd edition. Chicago/London: University of Chicago Press, 1998 [1958].

Arendt H. *The Origins of Totalitarianism*. 2nd enlarged edition. Cleveland: The World Publishing Company, 1962 [1951].

Bostrom N. Transhumanist values. *Review of Contemporary Philosophy* 2005; 4(1-2): 3-14.

Bostrom N. Are we living in a computer simulation? *The Philosophical Quarterly* 2003; 53(211): 243-255.

Boswell J. *The Life of Samuel Johnson, LL.D.* London: John Sharpe, 1830 [1791].

Bouwsma OK. Descartes' evil genius. *The Philosophical Review* 1949; 58(2): 141-151.

Chalmers DJ. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press, 1996.

Dennett DC. Postmodernism and truth. *Proceedings of the Twentieth World Congress of Philosophy* 2000; 8: 93-103.

Dennett DC. *Consciousness Explained*. Boston: Little, Brown and Company, 1991.

Lewicki R. Art as the parasitic process of thought: art-work, art-labour, and art-action in Hannah Arendt. *Journal for Political Thinking* 2025; 14(1): 119-137.

Lyotard J-F. *The Postmodern Condition: A Report on Knowledge*. Bennington G, Massumi B, trans. Manchester: Manchester University Press, 1984 [1979].

Plato. *Apology*. In: Grube GMA, trans., Cooper JM, ed. *Plato: Complete Works*. Indianapolis: Hackett, 1997: 17-36.

Rodriguez P. Talking brains: a cognitive semantic analysis of an emerging folk neuropsychology. *Public Understanding of Science* 2006; 15(3): 301-330.

Roskies A. Neuroethics for the new millenium. *Neuron* 2002; 35(1): 21-23.

Moore, GE. Proof of an External World. In: Baldwin T, ed. *G.E. Moore: Selected Writings* London/New York: Routledge, 2003 [1939]: 147-170.

Niikawa T. Consciousness is sublime. *Ergo* 2025; 12(28): 726-748.

Niikawa T and Kriegel U. The sublime of consciousness. *British Journal of Aesthetics* 2025; 65(1): 113-130.

Strawson G. The consciousness deniers. *New York Review of Books*, 2018 Mar 13. Available at: https://www.nybooks.com/online/2018/03/13/the-consciousness-deniers/.

Ryle G. *The Concept of Mind*. New York: Barnes & Noble, 1960 [1949].

Wittgenstein L. *On Certainty*. Paul D, Anscombe GEM, trans. Oxford: Basil Blackwell, 1969.

Zuk P. Mental integrity, autonomy, and fundamental interests. *Journal of Medical Ethics* 50(10): 676-683.