

The Paradox of Self-Consciousness | Edge.org

A Conversation with

THE PARADOX OF SELF-CONSCIOUSNESS

One of the central questions I'm asking myself is how to fit the human being into our current understanding of both natural scientific fact and the social and general mental and interpretative facts unearthed by the humanities and social sciences. Where do we locate the human being and what we know about ourselves from humanistic, historically oriented research vis-à-vis contemporary technology, the digital sphere, cutting-edge research in physics, neuroscience, etc.?

Philosophy's central object is the human being and its position in the mindless universe. How do we fit into reality with our perspectival minds? That's fundamentally the kind of question that I'm working on, using various tools hopefully suited for trying to tackle that very hard question.

One of the tools I'm deploying is the whole category box of contemporary theoretical epistemology. Epistemology asks the following questions: What is knowledge? How far does knowledge extend? What can we *really* know about the universe and ourselves as knowers of the universe?

Many people think there's a problem with how we can fit consciousness, or the mind, into a mindless universe. But there's an even deeper problem at the outset of this enterprise, which is how can we know ourselves as knowers and our position in the universe?

Knowing anything about the universe requires being causally in touch with it. We cannot know anything about the universe without, to some extent, intervening in it. We don't know anything relevant about the universe a priori, that is, by just thinking about it. We didn't discover bosons by thinking harder about the composition of physical reality; we had to run thought experiments and develop the right mathematical tools in order to check whether our understanding of the universe matches the facts. Checking thought experiments requires causal intervention. The limits of causal intervention are, thus, real, physical obstacles to human knowledge. Currently, we do not know where exactly (if ever) we will hit a knowledge ceiling. In any event, modern science tells us that it is basically impossible to know absolutely everything about physical reality. There simply are scientific reasons that underpin research into meta-physical issues, i.e., into facts beyond the ken of physics and physical intervention.

There is a feedback loop between certain hypotheses that we have vis-à-vis the functioning of the universe, given what we know from prior research and certain expectations about the future of science. In this feedback loop, we are currently discovering that there are limits to what we can know about the universe. For instance, we can only know anything about the universe to the extent to which it's observable. We also know that parts of the universe are not observable in the same way with our current instruments. We do not know enough about dark matter and dark energy, for instance, apart from the fact that they exist and that there's a certain ratio of dark matter and dark energy underlying the principles of the observable universe. There's a

huge chunk of the universe that is currently not accessible via experiments.

Even when we discover more sophisticated methods, there might be other parts of the universe that are inaccessible to possible research projects carried out by entities, such as humans, within the universe. Knowing anything about the universe requires changing the universe, even if in very slight degrees. Running an experiment means interfering with the target system that I'm investigating because we are *in* the universe.

Similar things apply to the human mind. Take the overall mental state that I'm currently in. I feel a certain way, I have certain thoughts, I'm trying to answer certain questions, I'm looking at my mental history, and I'm weighing reasons that speak in favor of what I believe (as well as against it) so that I can get a reasonable point of view on my own point of view. As I do this, I change my overall mental state by virtue of thinking about it. This is what is called *the paradox of self-consciousness*. If I'm conscious of consciousness, if I'm consciously thinking about thinking, I change the state that I'm targeting because my overall mental state is now different. I might use yet another third-order state in order to think about me thinking about thinking, but this makes things more complicated and does not help me to get outside the skin of my thought, as it were.

There is an associated language problem which is often overlooked in contemporary mind science. If you look at the history of writing and literature in the 20th century, the "Letter to Lord Chandos," by the great Austrian author Hugo von Hofmannsthal, makes exactly that point. When we state our self-conception in any kind of linguistic or technical code (such as a scientific model) by describing us as "conscious," "rational," or whatever kind of thinking animal, the language itself does not guarantee that anything in physical or biological reality corresponds to this exact concept. How do we know that we are conscious without knowing that the English word "consciousness" picks out something in a reality that is not made of words (but say of neurons)?

Since Plato and Aristotle, philosophy has tried to solve the paradox of self-consciousness. My current work is focused on the consequences the paradox of self-consciousness for the interface of humanistic and scientific research on the nature of the human self.

Think, for instance, about artificial intelligence. I think about artificial intelligence, roughly, in terms of models of thought processes. While our algorithms deploy models of how humans think (which we created even though their execution in cutting-edge computing machinery is ever more independent from our control), these models are not identical to the way we think. AIs are alien and weird to us precisely because they don't literally think the way in which we think. This creates the problem of how we can compare artificial intelligence, to the extent to which we understand its inner workings at all, and human intelligence. Working out a common basis for understanding different kinds of intelligence lands us precisely in the paradox of self-consciousness.

Using human, biologically grounded intelligence, how can I know what intelligence is without interfering with the system that I'm trying to study? Artificial intelligence has clearly changed the way in which humans think. Human intelligence has been significantly transformed by the recent digital revolution. I think it is a reasonable assumption that humans have become more intelligent due to our use of advanced technology. We do not have to merge with AI (by implanting chips in our brains, for instance) in order for our minds to change. AI can directly interfere with our minds, as it does in our everyday use of digital technology, which evidently

changes subsystems of our brains. Our brains are not simply a bunch of hardwired structures, they constantly rewire themselves. Given that our minds are not strictly identical to our brains and how we do not know everything about the specific relationship between certain mental events and physical events anyhow, we should not rule out that we are already merging with AI. That's one topic I'm working on right now.

The notion of intelligence is subject to many different definitions. Figuring out which definition captures the essence of intelligence is another instance of the paradox of self-consciousness. We need to know what intelligence is by using intelligence. There is no sideways-on point of view where we just compare intelligence as it is with our definition of intelligence. We have to work from within intelligence.

The definition I currently use is: Intelligence is the capacity (paradigmatically embodied in mammals such as the readers of these lines) to solve a given problem in a given amount of time. One system is more intelligent than another system to the extent that it solves the same problem faster. In this regard, we have psychometric methods of measuring intelligence, which corresponds to the age-old idea of IQ. Measurable intelligence is, thus, a form of biologically grounded efficiency. This does not as such rule out that non-biological AIs could be truly intelligent; it depends on what it is for them to have and solve a problem in the first instance.

Given that I solve my everyday problems much faster with my smartphone, I thereby have become more intelligent. And if that's the measure of intelligence, it's much more intelligent to Yelp myself to a better restaurant than to simply ask my friends. Thanks to digital infrastructure, I get information much faster, and my behavioral system has already adjusted to that infrastructure. In this respect, the human-machine interface has made us much more intelligent than we were before. This might already be a way in which we merge with AI without having to enhance our brains. They might be good enough for a certain form of merger, and as [Susan Schneider](#) argues in her book *The Artificial You: AI and the Future of Your Mind*, this might also be the only recommended form of a fusion of humans and AI.

I worry about this much more than the prospect of a coming superintelligence. Why are we so concerned with the question of whether our artificial intelligence systems might take over the entire realm of information processing? It's because we are interested in our position in the universe and our human intelligence. Otherwise, we might just let our computers run their software and reproduce themselves in the most intelligent way. Why is this even a threat? Precisely because it changes *us*. Superintelligence is a threat to us, therefore, we urgently need to figure out who that is: us.

The paradox of self-consciousness has interesting methodological ramifications that people do not take seriously anymore. There's a sense in which we began to forget something that used to be called the "linguistic turn," in which there was some truth.

Imagine we want to find out what the minimal neural correlate of consciousness is. What can be taken away from my body before I cease being conscious? Whatever it is that you absolutely cannot take away, that would be a good candidate for the minimal neural correlate of consciousness. We all believe, though we don't fully know, that there's no consciousness without a *neural* correlate. Thus, some part of the brain has to be there for me to exist.

But wait a minute! What exactly are we looking for? It's not as if the English word "consciousness" has just one meaning that we can look at and find the correlate for that

meaning in the human nervous system. Why would one specific meaning of the English word consciousness be represented in brains? Brains evidently evolved before the existence of the English language, so there's nothing English about brains. Remarkably, the meaning of the English word consciousness has no exact equivalent in all extant natural languages. The German word *Bewusstsein* and the French word *conscience* have other meanings and none of them exactly captures the meaning of "consciousness" in the intended sense of "subjective experience" or "what it is like to be someone" etc. Chinese has at least five words for what we call consciousness, which do not literally map onto the dictionary meanings of consciousness in English. Which meaning are we focusing on when we look for consciousness in the brain, or in the subsystem of the brain? There's no way around solving that problem before we begin investigating consciousness. Just picking one's favorite meaning does not guarantee that we are asking good questions.

Philosophy, together with linguistics and other humanities disciplines, could contribute significantly to overcoming that deep methodological problem, which I call the "hardest problem." To some extent, that problem is even harder than the "hard problem" of how phenomenal consciousness—feelings, sensations, and the like—fits into mindless reality. That's a good question, once the meanings are settled, but there's the even harder methodological problem of how we determine what to look for in the central nervous system when we're talking about the minimal neural correlate of consciousness.

There's a real problem of how we can bring linguistics and other humanities that study natural languages and human self-conceptions together with cutting-edge research about the human being and its position in cosmos in the natural sciences. Here's a reason for doing this: We better have some kind of answer to what a human being is. My proposed answer to that question is that a human being is the kind of animal that sometimes leads a life in light of the question of how it fits into the mindless universe. We don't do this all the time. The philosopher perhaps does this more because we are paid to think about this.

Every human being has an account of what it means to be a human being. Some people mistakenly believe that they have an immortal soul, and that to be a human being is to be thrown into an evil cosmos so that God can test your mental states. Billions of people think that some version of the soul theory is true. Other people think that they are sophisticated killer apes designed to spread their genes. There are different ways in which you can think about what it is to be human. But they all better have something in common, otherwise the person who believes in the immortal soul and the evolutionary psychologist who thinks that mental states are fundamentally adapted to certain fundamental biological purposes wouldn't be human in the same sense. The great [Richard Dawkins](#) and Pope Francis are both humans, so that can't be the distinction between the two of them.

My proposal, then, is to say that the invariant of the human being is precisely that capacity to give an account of how you fit into the largest possible domain—reality as a whole. If you do that, you are engaged in the activity of turning yourself into the kind of animal that we happen to be. Humans are not just any old animal, they are animals trying very hard not to be animals. As the philosopher Stanley Cavell once put it, "Nothing is more human than the wish to deny one's humanity."

There is significant research produced by the humanities, social sciences, and philosophy about how people think about themselves and design languages to capture their sense of being a self. We need to take that research, including that of theology departments, into account, not

because we believe necessarily that there is a God and we have an immortal soul, but because we need to look at how and why people think about themselves as being equipped with (or not) an immortal soul. It is not enough to point out that the soul theory is an illusion if we cannot explain in humanistic terms how the illusion arose. For the illusion cannot be hardwired and impossible to overcome, otherwise great philosophers of mind like Daniel Dennett could not even claim to override it by means of an interpretation of scientific research.

What we often do in the 21st century is bracket these questions and pretend that we already know everything about the human being in order to be on safe ground establishing our animality. We are animals, but what exactly does this mean? We are also not animals in literally the sense in which all other animals are animals. There is a distinction. We built New York City, computers, but more importantly, we think about what it is to be an animal.

To be sure, we share fundamental behavioral traits with other animals because we evolved via the same kinds of principles, but our exercise of mental capacities significantly and in structure goes beyond anything that we have so far observed elsewhere in the animal kingdom.

~ ~ ~

Philosophy fundamentally is the highest-level metascience. It is the discipline that studies the rational grounds of the division of labor in academia. Academic disciplines come in departments, and the structure has changed over the centuries. Among other things, philosophers ask how this kind of knowledge acquisition in different fields ideally hangs together. What are the rational, including scientific, reasons for why there is a physics department, a biology department, a chemistry department, a German literature department, or a Japanese history department? Why do we have that? Philosophy's role in academia is to answer that question in conversation with all other disciplines.

We are trained in thinking about thinking, and in arguing about the structure of arguments; that's why logic originated in philosophy. Many other disciplines, modern and more recent, have a very interesting philosophical history and prehistory. Think about Alan Turing's engagement with philosophy. I don't think that we would have a digital age without significant breakthroughs in symbolic logic in the 19th century, prepared and, to some extent, carried out by philosophers such as Bertrand Russell, Gottlob Frege, Alfred North Whitehead, and George Boole. They prepared the ground for the digital age and the kind of high-level computer science that is a major driving force of contemporary global civilization.

Philosophers do everything meta. We go one level up. Someone asks a question, runs into a deadlock, and then philosophers ask why there is a deadlock. That's why philosophy can also be useful on the spot for other disciplines, because it helps us critique modes of thinking and come up with new ways of thinking about the human being, about intelligence, etc. That's how I think of philosophy, in general.

In the contemporary climate, philosophy has many roles. One is to bring postmodernism to a final conclusion. The other one is philosophy's function in various crises which I subsume under the label of "reality in crisis." People worry a lot about the accessibility of facts. Even neuroscientists worry that we might not be able to understand reality due to inherent limitations of the brain, or that our entire perceptual reality is an illusion forever masking a thing in itself—like reality. The idea that reality is an illusion, or the fact that we need certain forms of instruments—including a more or less healthy brain—in order to process, makes it very hard for

us to understand how there can be objective knowledge of facts, as the cognitive psychologist [Donald Hoffman](#) has argued in his book *The Case Against Reality*, in which he says that evolution has provided us with a mental tool structure that hides reality from us.

There's this paradox in the current sociopolitical climate where on one hand, we have never known as much as now. The last 200 years that is modernity is basically just an explosion in knowledge and technology. We clearly live in a knowledge society. On the other hand, people question our capacity to know the facts, which even led to crises in democracy. Just think of post-truth and fake news and all that bullshit, or the various forms of denial of scientific fact, conspiracy theories etc., which spread in social media and which would not exist without major scientific breakthroughs of modernity.

There's a two-fold threat to objectivity. I would define objectivity as the feature of human minds to get things right or wrong. We often get things right. For instance, thanks to nuclear physics, we know more about nuclei than, say, Isaac Newton or the pre-Socratics. We get the atom better than Democritus. We know it has a nucleus, and we know stuff about quarks and how they bind together. In virtue of our capacity to get things right, we sometimes get things wrong. We are manipulable precisely because we can know reality. There's a two-fold threat to that idea. One is an intrinsic threat coming from science itself, and the other is an extrinsic threat coming from postmodernism. As Robert Proctor argues, there is a whole system of what he calls "agnotology," which consists in methods of the making of ignorance. Humans can be talked into not knowing what they know, and this is a huge problem in the digital age.

The intrinsic threat has to do with certain discoveries about the human being; for instance, our cognitive biases, the fact that our reasoning in real-world circumstances is not at all ideal so that no human being ever fully behaves rationally for extended amounts of time. Every human being is a bunch of contradictions, which we know for certain from recent breakthroughs in behavioral economics, psychology, and neuroscience. We are limited in remarkable ways. But are we also limited in knowing about our limits? Another crucial instance of the paradox of self-consciousness!

This is where philosophy as *the* metascience comes in and points out that the knowledge the behavioral economist has about the human being is not in the same way biased as the biases he or she uncovers. It's not a biased claim that we have biases; it's just a fact. There is objective knowledge. Nothing that we know from neuroscience or psychology should ever stand in the way of recognizing our capacity to know how reality is. Otherwise, what are we claiming? This would undermine scientific objectivity itself by its own means.

There's a widespread idea, for instance, that our whole conscious mental life is a kind of illusion. If consciousness were an illusion, then what am I doing consciously making knowledge claims as a scientist? Looking at my instruments, I am consciously engaged in the activity of making and substantiating knowledge claims about natural reality. If I deny that I'm doing this, with the help of my instruments, then the whole thing breaks down. It's a "performative contradiction." There are ways of defending this view called "illusionism," but it comes with deep problems. Clearly, mental life cannot be an illusion all the way down and all the way up. That's the intrinsic threat to objectivity, stemming from human self-consciousness.

There's another threat to objectivity that is equally widespread in contemporary culture, not just in academia. You would find this threat in journalism and politics. This threat is postmodernism. According to postmodernism, our knowledge claims are just expressions of a will to power, as

Friedrich Nietzsche and his French follower Michel Foucault has put it. When you claim to know something, an exercise of your will to truth or your will to knowledge, what you are really doing is nothing but asserting your power position. So, according to the postmodernists, someone who wins a Nobel Prize didn't discover anything, rather he or she is just the most powerful scientist in the community.

The community, according to the postmodernists, decides what counts as true, and there's nothing beyond that recognition that would be the truth. The American postmodernist philosopher Richard Rorty literally defended that view. It's a crazy view, but many people hold versions of it, implicitly or explicitly. Think, for instance, of the fictional President Frank Underwood, from *House of Cards*, who nicely puts this view in the line, "There is no justice, only conquest," remarkably expressed by Kevin Spacey. There's a whole series of problems with this character (and the actor playing him) having to do with our socioeconomic conditions which are studied by sociology, a humanistic discipline steeped in paradoxes of self-consciousness.

Many people are confused about the fact that there is something social about knowledge claims. MIT is a complex social system, without which they wouldn't be able to discover so many things. Why is MIT such a great university? Among other things, it's a great social system: It pays the right salaries, offers the right benefits, attracts the right people. But this does not mean that people there do not discover things. That's the postmodern confusion. They think that the sociality of human knowledge acquisition stands in the way of objectivity.

I'm trying to create a whole conceptual toolbox, a model of objectivity. I'm also doing something that scientists would do, which is to create models in order to account for the properties of my target system. My target system is the human being and the human being's origin of knowledge claims. The human being is the animal that knows remarkable facts, such as that it is an animal. That's my target system. I want to put objectivity in the right spot.

I came to this overall project, which I call "New Realism," as a consequence of my own philosophical education and academic upbringing. I got all my philosophy degrees, a PhD and *habilitation*, which you have to get in Germany in order to be eligible for tenured full professor positions, at the University of Heidelberg.

One of the strengths in that department was that people were trying to overcome a very problematic division in the field between "continental/European" and "analytic" philosophy. Continental philosophy typically just means postmodernism. It's basically the idea that we cannot really know facts, that science doesn't matter or is some kind of problem. I rejected all of that. Continental philosophy is like a continental breakfast in that it's a bad hybrid of elements from the European tradition.

Then there is analytic philosophy, which usually just means philosophy—giving arguments, providing reasons, and adjusting your belief system to the best evidence and the best counterarguments, just improving by letting yourself be falsified by better arguments and better scientific evidence. The actual various European traditions of philosophy did just that: Remember how Aristotle discovered logic, Leibniz worked out calculus and contributed to the invention of the computer, not to mention all the Renaissance thinkers etc. Closer to our time, Edmund Husserl was both a mathematician and a philosopher, and many of the founding heroes of analytic philosophy are German (people like Rudolf Carnap or Gottlob Frege). Similar things can be said about the so-called German Idealists who were extremely knowledgeable in

the science of their day.

I try to combine both, the European traditions without postmodernism, and analytic philosophy. This is what I was trying to do as a student. Then, the great philosopher Crispin Wright, who still teaches at NYU, came to Heidelberg for a series of seminars, which deeply impressed me. He gave seminars about the problem of skepticism, which says that we are not able to know anything at all because in order to know anything at all, we have to rule out infinitely many hypotheses, some of which cannot even be tested.

For instance, how do I know that I'm not in a mad house, hallucinating all of this? It is logically possible that I'm in such a madhouse right now. How do I know I'm in New York City and not back in Bonn, hallucinating being in New York City? Given that everything would look the same to me, there's a sense in which I cannot know this. This is the case for infinitely many other hypotheses. How do I know that I'm not in a simulation?

The "simulation argument" is arguably another skeptical hypothesis. Any reason we might have to rationally believe that we might be in a simulation run by superior beings, futuristic human engineers, or maybe a superintelligent AI trying to study human psychology in order to colonize us, is immediately undermined by the fact that the alleged reason is merely a simulated reason. If we are in a simulation, we simply cannot figure this out by any means whatsoever. Otherwise, the programmers would be idiots if they are interested in keeping us in the dark!

I started worrying about the skeptical problem and came to New York City in order to work with some of the best epistemologists and metaphysicians on the market here at NYU. At the time, I was talking to Thomas Nagel, who became a mentor for me for several years, and the philosopher Paul Boghossian. These people here deeply impressed me—Thomas Nagel, Paul Boghossian, and other figures at NYU such as Paul Horwich, Stephen Schiffer, and Béatrice Longuenesse. That was a very good intellectual climate for me, as someone who had just finished his PhD.

Ever since I've been coming back to the United States, I got in touch with other first-rate philosophers and scientists; in particular, a whole community of Japanese physicists who teach on the West Coast such as Hiroshi Ooguri, who runs Tokyo's Institute for the Physics and Mathematics of the Universe, and Yasunori Nomura, who became one of my interlocutors in philosophical problems of quantum mechanics.

I have been trying, under the banner of New Realism, to reconcile various philosophical and scientific traditions. I'm looking for a third way between various tensions. There's more to a human being than the fact that we are a bunch of cells that hang together in a certain way. Humans are not strictly identical to any material energetic system, even though I also think that humans cannot exist without being, in part, grounded in a material energetic system. So, I am rejecting both brutal materialism, according to which we are nothing but an arrangement of cells, and brutal idealism, according to which our minds are transcendent affairs that mysteriously peep into the universe. Both are false, so there has to be a third way.

Similarly, there must be a third way between postmodernism, which denies the objectivity of human knowledge claims and science altogether, and various trends in cognitive science, which also threaten objectivity without, of course, fully undermining it (for instance, research on cognitive biases better be immune to second-order biases). Similarly, I believe we urgently need to reconcile so-called continental philosophy—European traditions, broadly construed—

and analytic philosophy, which means philosophy at its best when practiced in Anglophone context; there has to be something in between. That space in between is what I call New Realism.

New Realism is indeed a series of research projects. My own contribution to it consists of two fundamental tenets, which I have spelled out so far. Everything else that I'm working on is connected to these two claims.

Claim number one is that we can know reality as it is, in itself. If I know that I have two hands, then I have two hands. There is no gap between my successful claim to know that I have two hands and the fact that I have two hands, even though I need to deploy complex neural machinery in order to perceive my hands. My sensation and conscious perception of my hands is distorted. In perception, there's always an element of illusion, which we can adjust to the facts. I turn my hands around, and it turns out they have two sides, so I now know I have a full hand. That's why we vary objects in perception. We walk around them to gather more perspective in order to verify our assumption. Our eyes do this by themselves in the form of saccades, and our other sense modalities equally gather information about reality by shifting perspectives. This is why we are essentially dynamic, organic, animal thinkers whose cognitive capacities evolved over long stretches of biological time before we became explicit self-conscious knowers.

Once we overcome the element of illusion in a legitimate knowledge claim, we know how things are. That's claim number one. There is actual objectivity. We can get things right or wrong, but more often than not we get them right because we are trained at getting things right. We, modern humans, have less overall cultural illusions on this basic level. We don't hallucinate too often, unless we have a drug problem.

The second claim, which sounds sexy and paradoxical, is that the world does not exist. What I mean by this is that there is no single unified account of all the facts. Let me give you an example of different kinds of facts: It is a fact that there are infinitely many prime numbers. It is a fact that there are different orders of infinity. It is a fact that there are Hadron Colliders. It is a fact that there are bosons and fermions. It is a fact that there is just one US President. It is a fact that the European Union is a system of states. It is a fact that I'm feeling a certain way. It is a fact that the ancient Greeks attacked lots of other countries etc. We cannot demonstrate in principle that we ever have a full account, a complete list of all the facts.

What I'm claiming when I say that the world does not exist is that there is no single overall theory for all the facts. There might be a theory of everything about the physical universe; I'm not necessarily denying that. I don't think there is one, but that's not the center of my agenda. We might be able to come up with a maximally unified physics. I'm not saying it's impossible to unify quantum mechanics and relativity theory. It would be absurd for a philosopher to make such a claim. Yet, even if we had the grand unified theory about the physical universe, that wouldn't give us information about which party to vote for, or which artwork to appreciate, or whether there are transfinitely many numbers. The continuum hypothesis in mathematics is not solved by a physical theory of everything because numbers are not physical objects; you cannot investigate them with experiments, you cannot causally interfere with them in order to measure their behavior.

What kind of a physical experiment would you set up in order to falsify, or verify, or get evidence concerning the extension of numbers into infinity? Obviously, you cannot do that. Physical

instruments can only measure finite objects. Numbers are not objects in the universe anyhow. The number two is not located in a specific place; it's not in Oklahoma. It would be a misguided question to wonder where the number seven is today. Numbers are not spatiotemporally located.

My claim is that there's no discipline that would be capable of bringing all facts into view in one big world picture. That's the claim that the world does not exist. Otherwise put, no worldview or world picture is adequate. We should overcome the tendency to produce world views, which also means giving up the idea that science is a worldview. It is not. It is an important epistemic practice of achieving objectivity and of getting certain facts right, but it does not deal with all facts, as there is no such thing as all facts in the first place.

How do I know that a manifold of different kinds of things, which I take to be there in virtue of ascribing a certain kind of meaning to my words, are really there? The great philosopher and logician Willard van Orman Quine suggested that ontology should not be metaphysics, i.e., the investigation into reality; ontology should be the investigation into the ontological commitments of a theory. What does a theory need in order to be at its best? Which terms need to refer? "Electron," you might say, has to refer in a physical theory. But there are clearly elements in mathematical physics that are not intended to refer in our equations. This is why, before we speak about reality, we have to be able to become aware of the vocabulary in which we couch our knowledge, which lands us back in the arena of the paradox of self-consciousness.

It's not that we can just write down a catalog of objects that exist—numbers, grandmothers, pains, itches, brain states, etc. It's not that simple. There aren't any witches, for instance, but we talk about them. People dress like witches on Halloween and we say, "Look at that witch over there." We shouldn't infer that they are witches. They are people pretending to be witches. There are witches in my imagination, but witches in my imagination are not witches. If I think about bananas, that doesn't satisfy my hunger. Imagined bananas are not a kind of banana.

We need to be able to draw a good principled distinction between terms which are in the business of referring and those which are not. This brings us back to both the paradox of self-consciousness and the impossibility of just leaving language. New Realism tries to reconcile us with the insight that we can't just write down a list of entities. Physics is also not doing that. No one is doing that. You might think that this is what you are doing as a scientist, but you are wrong. That's not what we're doing. It's much more complicated. We cannot circumvent the meta-scientific, i.e., philosophical investigation into our ontological commitments and just "talk about things." That would be naïve.

Gertrude Stein says, "To measure is to treasure." One way of making sense of this insight is to say that when we measure something, we have evidence that our term meets ideal meaning conditions. For instance, if we measure the presence of the Higgs boson via the Large Hadron Collider, we know that certain elements in our equations have a high level of reference, as opposed to the term "unicorn."

Philosophers debate the question of whether there are unicorns in movies, or whether, according to a movie, there are unicorns. There are different views. One view is just a report of what a movie tells us, and then no one will buy that. You would say of course there are no unicorns, but according to a movie there are unicorns. This is like saying, according to Pope Francis there's an immortal soul, but of course there is no immortal soul. I have a different ontology for movies and artworks. I think there are unicorns, but that they are unicorns in

movies. I'm a fictional realist.

That's a complicated matter, which has interesting consequences, again, for our current age. Think of the ontology of virtual reality. David Chalmers is writing a fantastic book about that. Do characters in videogames exist? If you and I both play a videogame in which we both are present in a virtual reality scenario, in the form of our avatars, then the ontology of the avatar differs from an object that is merely invented, such as Donald Duck, because it relates to our actual behavior in systematic ways.

There are interesting questions to be asked about that. Virtual objects are hybrid objects, meaning not just figments of the imagination but grounded in something real. The behavior of my avatar is grounded in my behavior and my movements, so they're hybrid objects. That is why, for instance, we can so much as use them for scientific research. We can use virtual reality to cure people.

That's different from, say, watching a Disney production. If I watch a Disney movie, there is a sense in which I don't learn anything about reality. I learn something about the world of that movie, but nothing about anything outside of that movie. That's another problem of our age. Many people are confused about that. Arguably, many people think that if they watch a Netflix show, like *House of Cards*, they learn something about Washington politics, but they are wrong. A purely fictional depiction of events in Washington doesn't tell us anything about how politics functions in Washington.

There is this fiction-reality confusion in our wider culture. I've heard people refer to castles in Europe by saying that they just look like castles in Disneyland. It's the other way around—the castles in Disneyland look like those good old castles. They were there before. People even confuse what happens in Disneyland with historical reality.

These confusions can only be cured by bringing in philosophy and the humanities. Even our best physicists are not automatically good historians. Why would they know what the relationship is between an event in the Middle Ages and the depiction of that event in a production on Broadway? You need to ask historians in order to come to terms with the relationship between historical fact and fictional representation.

That's just not a target of physics. It's not that quantum mechanics should better deal with Donald Duck. It's just not a research project. Quantum mechanics can deal with what it takes on the level of electromagnetic radiation in order for the live play on the screen to interact with my nerve endings so that this gives rise to mental imagery. That can be studied, because the brain cannot perceive anything without being in the same electromagnetic field as its objects. There is a very meaningful and important contribution of quantum mechanics to perception, including aesthetic perception. But that is not identical to figuring anything out about the relationship between a Disney depiction of the Middle Ages and the Middle Ages.

~ ~ ~

Philosophy now is practiced in what many people rightly characterize as an age of science. Scientific discovery and technological application in a rapidly thriving global market economy changed the human life world and the human life form. Philosophy's expertise has always been to look at the conceptual changes that are created by changing socioeconomic and epistemological circumstances.

It's essential to philosophy that we are in control of our past too. That is true for any other scientific activity. Even though Newtonian mechanics is to some extent superseded, it's not the case that contemporary physicists would not be able to solve those equations; it's built into the DNA. Earlier stages of scientific and rational development always play a role in later stages. Philosophy is no different. I think of philosophy as a particular kind of science. The metascience that philosophy is is as much scientifically grounded as other disciplines. But philosophy has a different relationship towards the empirical.

For instance, the very idea that there are empirical concepts, and maybe non-empirical concepts, is something that can only be studied by philosophy. Some concepts are clearly empirical in that knowing how they function and how they can give us actual insight requires experiment, theory revision in light of incoming empirical evidence and the like. Take an empirical concept such as pain. We learned things about pain that we didn't know in the past by studying the evolutionary function of pain in humans or other animal organisms. That is an empirical concept.

The concept of a concept is not in the same way an empirical concept. There is the discipline of logic. Logic doesn't work by running experiments on how people think. Logic tells us how people *ought* to think. That's a classical philosophical distinction. For instance, if we think that it's an epistemic virtue to avoid contradictions, then we accept some version of the logical law of non-contradiction. If there is no reason for me to have consistent beliefs, then why on earth would quantum mechanics care for a mathematical apparatus that is free of contradictions? Of course, they do. It would be a disaster for science if contradictions were allowed *ad libitum*, because from a contradiction, everything follows.

Even though this has been disputed in philosophical logic, there are some arguments why the law of non-contradiction might not be universally valid, but even those arguments are not empirical arguments.

There are logical concepts, and the difference between kinds of concepts is laid out by philosophy. But philosophy cannot do this without taking into account what happens in science and the humanities. Philosophers can't just sit around and think about this harder. That wouldn't change anything. Philosophers need to be in the room together with people who ask philosophical questions in their own science.

Here's another way of thinking about the role of philosophy: There is pure mathematics, which as such is utterly useless. For instance, if someone does the kind of work of my distinguished colleague at Bonn, Peter Scholze, who won a Fields Medal for his work on perfectoid spaces, then this has no consequence whatsoever for scientific progress. He just figured something out about a particular set of problems in pure geometry. This is completely useless; however, it's pure mathematics and we value it.

There is also applied mathematics, the use that we make of mathematics, which in turn changes mathematics. Many problems and actual progress in contemporary science is a case of applied philosophy. That's why philosophers can contribute to the advancement of knowledge in some fields.

Think of [Daniel Dennett's](#) thought experiments, which have had a big influence on the development of cognitive science. He's thinking about a certain set of concepts, such as consciousness or free will. He brings philosophical expertise to a conversation with people who

work on the way these concepts are implemented, say, in human nervous tissue, and that conversation advances both.

I'm currently setting up a research center called the Center for Science and Thought, which I co-direct with a nuclear physicist. We run conferences on the frontiers of science, and we ask philosophical questions where you might not have expected them. For instance, there has been an important debate in fundamental physics about superconductivity and emergence. There's this famous paper by the physicist [Philip W. Anderson](#) entitled called "More Is Different," and there are different interpretations of this. There's an associated problem in physics. There is a limit to how deep low-level physics can probe into the universe, because at some point, you need more energy than is even available in the entire universe in order to dig deeper. We don't know if it's possible to go deeper than the Planck scale, because you may need too much energy in order to get there. Maybe more energy than might ever be available. And even if we got one level lower, we don't know if we hit the bottom rung of the universe.

This is a classical philosophical problem. Is there something that is the smallest entity in the universe? Or is there a smallest, absolute scale? This is what the philosophers have called an atom. The philosophical atom is not the modern atom. The philosophical atom would be the smallest scale. We don't know if there is a smaller scale than any scale we will ever empirically discover.

What we are doing in this research center is we look, for instance, at the role that effective field theories can play. They look at higher-level phenomena in order to study lower-level phenomena. This doesn't require strong emergence, it doesn't require the assumption that the higher-level phenomena are causally autonomous. It's compatible with that, but it's also compatible with weak emergence, and with all sorts of interpretations. What we do is look at those phenomena, and the physicists tell us what's going on. For instance, the group that I'm working with in quantum chromodynamics are working on issues such as questioning whether there are glueballs, which are structures made of gluons. They give me the evidence, they explain to me some of the mathematics, and I give them philosophical expertise—how I would think about that, the concept formation.

This is leading to results. We're editing a volume now on top-down causation, on the question of whether very high-level phenomena, such as human action, can causally influence low-level phenomena, and what this will take. It seems to be evident that it's possible, but this raises interesting issues which George Ellis, for instance, a cosmologist who is discussing in his book *How Can Physics Underlie the Mind?* and in a paper we are currently writing on together. Is determinism compatible with free will? There's an abstract, philosophical debate about that. We are trying to bring those tools to the table as well.

I have a big philosophical word for this project, which is non-transcendental empiricism. *Non-transcendental empiricism* is the idea that the reason why we value empirical concepts and ought to value them is because they give us scientific and technological process, but also because there is no fundamental layer of the human mind.

Many people who say that experiment matters, or who call themselves empiricists, assume something of the following sort: We fundamentally know of reality via sensations integrated into conscious perceptions; the rest is abductive inference. Here, I stop. Do we? One way of knowing things about reality is by being in sensory contact with it. Perception is a good way towards knowledge, but that's not the only one. Quantum physicist [David Deutsch](#) nicely

pointed out in his book, *The Beginning of Infinity*, that cosmologists know something about the universe as a whole. How do you know something about the universe as a whole? Certainly not by abduction, which does not give you total knowledge about the universe as a whole. The best way to know that I'm in New York City is to look out of the window and see the Flatiron Building. That's perceptual knowledge. It's empirical knowledge, but not all knowledge is of this sort. And don't forget that the idea that the best way to knowledge is perception plus abductive inference is not an application of that overly simple method.

My knowledge about the universe as a whole, or a cosmologist's knowledge about the universe as a whole, is not empirical in that sense. There will be empirical knowledge, but it's a different kind of empirical knowledge. This is why I postulate a *sense* of thinking. We know from recent science that we don't only have the five famous sense modalities. We have a sense of time, a sense of motion, pain and other sense modalities. A big team of life scientists at my university even consider the immune system to be a kind of sense modality. Why not include thought as an additional sense modality? If this were true, we would not be locked into our skulls, but our thought could as much be out there, a legitimate part of reality, as electromagnetic radiation.

~ ~ ~

The function that philosophy can bring to general debates of high public interest in modern secular societies, in my own experience, has been extremely fruitful. For instance, I published something about the relationship between the self and the brain, which a group of renowned German neuroscientists took note of. I argued that a human self is a compound of necessary biological ingredients and some non-biological elements, such as a self-representation of itself as being an entity in the universe that cannot be reduced to brain states. What they suggested to me is that it's even a hard problem to think of the unity of the brain. They told me that one of the deepest problems they are facing is not the mind-brain problem, but rather the brain-brain problem, as they called it.

What would it take to think of the brain as an organized entity in the human organism? When people say, "the brain," what do they mean? They don't mean the cerebrum, which is a subsystem of the central nervous system. They mean, roughly, what's in here (pointing to my head). What's in here is not unifying in the same way in which a liver or a heart is unified. There's no obvious single thing that the brain is doing. The heart is pumping blood for the organism. What's the brain doing? You can't say that it's the thinking organ because it does all sorts of things. It's not even clear whether there's a single organ—the brain. That's very questionable.

They told me that before we even begin to worry about the mind-brain problem, we should worry about the brain-brain problem. We talked about this, and it turns out that there is a new version of the mind-brain problem, namely that it might be the case that the mind is nowhere nearly as unified as we think, just as the brain might not be as unified. If the mind is not unified, if there is a sense in which there is no single self (maybe many selves in one animal), why would the brain be unified as the underlying material-energetic reality?

They revised some of their conceptions of the brain after the discussion, and I revised some of the conceptions of the self. I learned that parts of the self can truly be understood, and much better, by way of insight into brain function. My toy model for this is always puberty. If you think that adolescents are engaged in a revolt, you don't get the fact that this is an expression of hormone change. You treat adolescents very differently if you think of them as undergoing

hormonal change. It's not a revolt; it's a very different biological process.

Parts of the mind can be thought of entirely along the lines of what we know from biology. Other parts of the mind can't. My voting decision shouldn't be thought of entirely on those grounds because there are sociological reasons and causes why, for instance, I vote for a given party. There's no full biological explanation of my voting behavior, even though biological explanations are part of this.

What we can learn from full cooperation between philosophy and science about cutting-edge problems about the human being is how to distinguish between events and processes in the human life that follow a biological pattern and processes that don't. Parts of us transcend the biological realm, such as our capacity to think about transfinite numbers. We're in touch with objects that are not physical, even though in a physical way.

What philosophy can bring to the table at the frontiers of science and the humanities is the capacity to negotiate the various concepts that are out there, because we are all confused about those concepts before philosophical analysis. Philosophical analysis is both a form of therapy and a constructive enterprise. It helps us to understand our conceptual confusions better, clear them up, and come up with a new sense of what it is to be someone. That has always been the function of philosophy: the love of wisdom.

For instance, people worry about liberal democracy these days. Why do we even value it? There's no biological imperative leading to liberal democracy. The human being has been around for approximately 200,000 years in roughly our mental shape, but liberal democracy is a phenomenon of modernity. It's been around ever since the French Revolution, maybe a little earlier in various stages. Why we think that it is valuable, that by itself cannot be exclusively explained by reference to things that happened, say, 150,000 years ago.

This is where philosophy kicks in. It addresses questions, such as why is racism bad? Racism is bad. Fortunately, many people think racism is bad. The racists don't. But why is racism bad? You cannot look into biology. If you look into biology, you will find that many members of certain races fight against other races. Biology, as such, doesn't give you an argument against racism. You need arguments of the form that there are universal claims of morality, for instance. You need an argument why we should not privilege a certain group of people to the violent detriment of another group of people. Those arguments come from philosophy. Who else would give us those arguments?

These arguments are not free-floating. Once you formulate them in the right way, there will be possible tests, and they will line up with certain experimental results about, for instance, the evolutionary function of altruism. There is, of course, huge research about that.

The idea that there are morally universal principles that philosophy can uncover in an age of science, in cooperation with our best knowledge about the human being and other primates, and the reach of morality deeply into the animal kingdom—we need to bring this together.

We don't get sufficient justification for liberal democracy just from science. You can have good science in communist dictatorships. Science is not necessarily a contribution to moral progress. Science has also produced the atom bomb and the climate crisis. We wouldn't have that without scientific progress. Modern physics and chemistry are literally in the engines of fossil-driven mobility and modern biology has also (perversely) contributed to horrible eugenic

projects and racist phantasies. Right now, progress in AI research might lead to the complete destruction of humanity by superintelligent AIs, a problem that leading AI researchers are, of course, very aware of and which they tackle in cooperation with philosophers. But we need to bring other disciplines in as well, such as the history of science, sociology, literature in order to study the influence of science fiction on our concept formation vis à vis AI and so forth.

If we think that our commitment to liberal democracy and certain universal values of human equality is just a side effect of contingent historical modern events, then I see no reason why we should stick to this. We might still stick to this, but then we end up being in a cultural war against other ways of being human. There's no guarantee that we would win this. In terms of numbers, we are losing it anyhow. That cannot be a sufficient ground.

Philosophy can work out ways of looking at this open-mindedly. I don't think the outcome is going to be that communist dictatorship is better. I'm biased in favor of liberal democracy, democracy in general, but we have to have that discussion. How to conduct a rational discussion about moral principles, cannot be addressed without calling the philosophers in. That's what we train in. We are trained in idealizing rational reflection, pure logics. We need to bring that to the table, not because it's an exclusive form of insight. The implementation of human rational insight, and that's something we know from contemporary science, the implementation principles violate the ideality of it. But that does not mean that there is no reason to study the ideal type.

~ ~ ~

Fundamentally, my argument is against relativism and anti-realism, which are widespread cultural and scientific phenomena. Relativism is the idea that commitment to our life form is basically just an expression of group interest. I prefer my way of living, and I think that my whole value system is nothing but an expression of my membership in my group. That's relativism. It's incredibly widespread, when people think that there are American and European and Chinese and Russian values and maybe even a clash of whole civilizations.

The second opponent is anti-realism, also an incredibly widespread phenomenon. Anti-realism tells us that our knowledge acquisition is not really about an independent reality. All we do when we make knowledge claims is mirror certain internal operations. Maybe we are just creating mental representations or brain representations of our environment.

Fundamentally, science is not a neural image of external reality, that's not what it's about. I'm targeting that idea, that we are brains in a vat, and the anti-realist consequences of that. On that basis, I'm attacking the horrible moral consequences that postmodernism and anti-realism have left us with, namely a weakening of our universalist commitments to hardcore human rights and universal moral principles. Those are my two fundamental targets, and I'm setting up all my conceptual machinery and engagement with science in such a way that we have a real ground for continuing the project the enlightenment project of reasons and the creation of a fully secular human society not based on false ideas from the deep human past. That's what I'm fundamentally interested in.

I'm specifically arguing against Daniel Dennett and Keith Frankish, and their irritating view that the mind is a kind of illusion. I'm also arguing against cultural relativists who hold views of this form. More specifically, my target is Nietzsche and the entire Nietzschean tradition in moral philosophy who uses some of the arguments for illusionism in order to undermine the value of

rationality.

When I'm attacking relativism, the opponent is Michel Foucault and his contemporary followers, including some feminist epistemologists such as Sally Haslanger or Judith Butler, who think that values or even human individuals are socially constructed. Values are not socially constructed; they are a universal truth about human beings. This is why, for instance, we need gender equality, not because we're engaged in socially constructing feminist values. Feminist values should be built into the human life form because all humans are equal. We shouldn't confuse our fight for certain values with the form of those values. Those are my specific opponents.

Daniel Dennett is about one of the worthiest possible opponents. Even as a matter of methodology, it's good to look at Daniel Dennett's forms of anti-realism and illusionism. Quite specifically, I'm targeting ideas from his most recent book, *From Bacteria to Bach and Back*, according to which consciousness and other mental phenomena are illusions.

Earlier Dennett was on the better track, by thinking that we characterize mental states and others from the "intentional stance." Facts that we get from the intentional stance need in no way be less objective than facts we get right from the physical or from the design stance. It's a specific lacuna in Dennett's arguments against mental realism to move from his insight into the intentional stance to full-blown illusionism. Mental realism is the idea that our terms, like "consciousness" and "qualia," refer to a reality that's really there. Many of those terms refer to reality that's really there, but they only refer to such a reality from the intentional stance. I'm trying to defend that, and to demonstrate that illusionism cannot be coherently formulated. The idea that consciousness is a kind of illusion, a trick played by the brain on itself, cannot coherently be formulated.

Another opponent is Francis Fukuyama. In his *Identity* book, he defends the idea that social identity is socially constructed. Who I am—as someone who has certain values, who belongs to a certain class or group of people (like philosophy professors)—is a function of ascriptions by others. There is a sense in which I have no essence that goes beyond my relationship to others. That is the idea of social construction.

Yet, like it or not, my shifting social identity is a kind of essence. It's something that is me, it's historically contingent, I can change it in the right ways by engaging with others. There is a whole social dimension to personal identity over time. But that social dimension is not a social construction, as the saying goes, but rather a real social production, including, for instance, biological phenomena. There's confusion out there about the relationship between, say, biology and gender. We need much more joint scientific and humanistic research in order to figure out how exactly the biology of the human organism and the social conditions of production of human cultures etc. hang together. This cannot be settled by the usually purely philosophical ideas from mainstream social constructivist gender theory.

Let me give you another example of how my view has teeth in such debates about gender. A good reason to recognize more than two genders is the biological fact that humans are not just born in two genders. It's a biological fact about human infants that we cannot distinguish them neatly in just two genders that are recognizable by certain reproduction organs. There are biological reasons for a plurality of gender, and not just socially constructive reasons. The reason why we should recognize transgender humans as full members of our community is an objectively existing moral fact, not just a social construction. It's not that we just welcomed them into our community. They have always been in our community. The problem was that we didn't

recognize that. I'm changing gear in those debates with my realism. I don't think of justice as a consequence of pure activism. I think of justice as, in part, established by scientific fact.

There is a phenomenon, particularly in the United States, called theory. It's mostly a form of social activism grounded in philosophical work. A prominent thinker in that tradition is a monster that a friend of mine, Maurizio Ferraris, made up and called Foukant, which is a hybrid of the worst ideas of Foucault and Kant. This hybrid holds that social facts are always in the eye of a beholder. When you judge a social affair, the distribution of economic resources in Manhattan, and it's unequal to a certain extent—if you judge that, according to Kant you are an activist. You are either for or against it, depending on your class membership and Foukant thinks there is no neutral, scientific ground, such as economic studies of distributive justice. That's the idea that there is no objective knowledge of those facts. And it is deeply misguided and dangerously false.

If you think there is objective knowledge, such as statistical knowledge, you will judge already from your privileged position. That's how they argue. They would think that, for instance, objective, statistical arguments of the Steven Pinker shape always just express his privilege as a Harvard professor. But there is a huge lacuna in the argument. Why would the fact that someone who is a Harvard professor say something about social statistics, as such, mean that he cannot give you objective judgment? That is just a fallacy, an ad hominem argument. A lot of social constructivism and a lot of theory is based on that idea, a series of ad hominem fallacies.

Think of Judith Butler. Her very important, influential book, *Gender Trouble*, thinks of itself as an intervention. She's not giving you a theory of gender. She doesn't clearly lay out in the book, for instance, what gender is as opposed to sex, or whether there is a sex-gender distinction, or how this relates to human population, or the human animal to other animals. That's not there. She does not make falsifiable empirical claims in order to offer evidence, philosophical concepts etc. The book is an intervention in order to give a voice to people she thinks are repressed, for better or for worse reasons. That's what I call activism. Activism for the right cause is fine, but activism itself cannot tell us what the right cause is. I'm not saying don't be an activist. But it's better to be an activist if your activism is based on facts, then to be an activist about activism. If you think that the basic facts that should turn you into an activist are already constructed by your activism, then you think of the social contract as just a pure struggle of forces. It would just be a clash of one community against another community.

That is indeed a problem of contemporary liberal democracy, that people misrepresent political dissent; they think of it, the social contract, in terms of just a war—an information war, say. We should think of it in terms of, to some extent at least, evidence-based policy. Otherwise, postmodernism will become the nightmarish political reality TV show that is currently happening in Washington DC and many other places.

The mode of thinking that I'm propagating has, among other things, the consequence of rethinking the way in which we inhabit the social world. It's between what people call theory, the typically left-wing activism, and the pure scientific expertise of, say, [Steven Pinker](#).

The *Enlightenment Now* book gives you an informed picture of modern progress, based on a bunch of facts. But then, that is, by itself, insufficient to direct us. What are we to do with that? Those facts alone don't give us a direction, we need moral theory based on scientific, humanistic, and pure philosophical knowledge in order to determine rational courses of action.

What I'm trying to do is to come up with a model of enlightenment where we get a direction;

namely, we should strive towards the implementation of universal value. That universal value is grounded in facts about the human being, and not in social activism.

Nietzsche is very important for the theory people because he has prepared the ground with his work *On the Genealogy of Morality* and *The Gay Science*, et cetera. Nietzsche's work has largely contributed to the idea that the social domain is nothing but an expression of a clash of powers. He even thinks that the universe is a struggle of forces, which he calls quanta of power. He thinks that all social affairs are nothing but that. Basically, he's a sophisticated social Darwinist, who believes that the stronger force will just win the fight and force the loser to accept their value system. That's his model of history. That's why he's asking for new values. He just wants to come up with the values that would then dominate the coming centuries which is why he even wrote a kind of prophetic, religious book *Thus Spoke Zarathustra*. That's what he's up to. In that respect, as a writer, he is himself a social activist.

Instead of giving us arguments that are clearly expressed, Nietzsche merely suggests views. He is first and foremost a great writer. His texts are written in such a way that they persuade you, regardless of the quality of the arguments (if any). It convinces you by way of the presentation of the thought, not by way of an argument. I find that highly problematic. We should have less of that pure rhetoric, even though it's an interesting genre, and more scientifically based rational, philosophical argument.

The pure philosophical qualities of someone like [David Chalmers](#) and Thomas Nagel are just among the highest level. Daniel Dennett, of course, is also a great pure philosopher, but he has always been in the business of engaging the hybrid domain. Lately, he has just not been that interested in spelling out his purely philosophical views.

I am more on the side that we need pure philosophy in conversation between science and advanced theoretical philosophy. However, there is a difference between the pure philosophy of Chalmers in *The Conscious Mind*, or Nagel in *Mind and Cosmos*, and what I'm doing. The difference is that I want to take the pure philosophy immediately to the table with the physicists and see what they have to say about that so that I can revise some of my concepts before they become a kind of personal dogma.

By the way, I don't think that Chalmers' view is correct, that possible worlds semantics can deliver, via the zombie premise, an argument for dualism. I don't think this works, because any good scientist or mathematician will tell us that there aren't any possible worlds. If there aren't any possible worlds, then this kind of argument fails. This is a debate that I would be having with him. And I would be having similar debates in pure philosophy with Nagel. But I would always double-check this with science.

I'm proposing an intermediate level. Pure philosophy—that's just my expertise, but then I want to double-check this with what's going on in other domains. I don't even give privilege to the natural sciences. If a historian could convince me that my moral universalism is flawed for certain anthropological reasons, then I would take that information in.

My range of interlocutors that I'm looking for in that respect is, therefore, bigger than Dennett's focus. This is how I would place myself in that landscape, which if you ask me, is the most advanced philosophical landscape right now which is why it attracts my own attention and motivates me to come up with arguments for a new form of humanistic knowledge, a new account of the objectivity of the humanities and their role for real moral progress based on

factual insight.