THE TRANSMITTER | BRAIN INSPIRED

What do neuroscientists mean when they use the term ‘representation’?

A group of neuroscientists and philosophers discuss the use and misuse of the term “representation” across the cognitive sciences and how it influences the way we interpret the connection between neural, behavioral and mental activity.

4 JUNE 2025 | by PAUL MIDDLEBROOKS

This transcript has been lightly edited for clarity; it may contain errors due to the transcription process.

Luis Favela
Does everyone know what people mean by representation in the cognitive sciences? Do neuroscientists even know what they mean by representation? We're looking at various literatures within the neurosciences, the cognitive sciences, and thinking, wow, it just seems like there's very different uses of the term.

Frances Egan
What's at stake are the sorts of conclusions that philosophers, among others, draw in thinking that we've made a lot of progress in understanding these mental capacities and understanding consciousness and rational decision-making, and so on.

Rosa Cao
It seems to me that those people should be allowed to keep on using the term "representation" for these things that don't have very much to do with mental representations, and the only bad part is equivocating between the two and pretending like we have an explanation when we don't.

Edouard Machery
I think neuroscientists do actually distinguish these two types of terminology and they're just not using one as a way of talking about the other.

John Krakauer
I think the idea that you're going to ask them, who basically are saying there's nothing at stake, let's just play around, is going to open--

Paul Middlebrooks
Hold it, hold it, John. Hold it for a sec--

Edouard Machery
This is not what they want you to say, John.

[laughter]

Rosa Cao
That's hard.

[music]

Paul Middlebrooks
This is “Brain Inspired,” powered by *The Transmitter*. What do neuroscientists mean when they use the term "representation"? That's part of what Luis Favela and Edouard Machery set out to answer a couple years ago when they surveyed a bunch of folks in the cognitive sciences and they concluded that as a field, the term "representation" is used in a confused and unclear way. Confused and unclear are technical terms here, and Luis and Edouard explain what they mean, what those terms mean in the episode in a moment. More recently, Luis and Edouard wrote a follow-up piece arguing that maybe it's okay for everyone to use the term in slightly different ways. Maybe it helps communication across disciplines, perhaps.

My three other guests today, Frances Egan, Rosa Cao, and John Krakauer, wrote responses to that argument. On today's episode, all those folks are here to discuss that issue and why it matters. Luis is a part philosopher, part cognitive scientist at Indiana University Bloomington. Edouard is a philosopher and the director of the Center for Philosophy of Science at the University of Pittsburgh. Frances is a philosopher from Rutgers University. Rosa is a neuroscientist turned philosopher at Stanford University, and John is a neuroscientist among other things, and co-runs the Brain Learning, Animation and Movement Lab at Johns Hopkins.

I'll link to all of their information and all the papers that I mentioned, plus some other papers that are mentioned throughout the episode. Also, some books that Luis, Edouard, and Frances have written. Frances in particular has written a recent book called *Deflating Mental Representation*, which deals with a lot of what we discuss in the episode. Anyway, you can find all that jazz in the show notes at braininspire.co/podcast/213. I am Paul Middlebrooks. I represent "Brain Inspired." I had to do it. I'm sorry. Now, I present our discussion, which picks up right where those intro clips left off. Enjoy.

[transition]

[laughter]

Edouard Machery
Hold your fire for a minute. Come on.

[laughter]

Paul Middlebrooks
All right, it's already gotten ugly. Here we are. Thanks, everyone, for joining me. We're here to discuss a back and forth that everyone here has had in the literature on the topic of representations. I thought we could just start maybe, Luis, with a high-level view of what we're talking about and then we can talk about why we're talking about it, what's at stake, why does this matter?

Luis Favela
Great. Thanks for having us, Paul. It's great to see you and everyone else as well. Edouard and I, a number of years ago, we're chatting one day and I was expressing discontent, which is really surprising for a philosopher to be discontent about anything. I'll say, Edouard, I hear my peer philosophers say things like, "Everyone knows what neuroscientists mean by representation," or, "There's a really well thought-out sense of representation in the cognitive sciences. Let's just take that for granted and move on."

Or, here are my key examples that I like to go to that represent, no pun intended, my view on representation in the sciences, and so we're going to work with that because what the scientists do should ground how philosophers of science and philosophers of mind think about representation. Edouard and I were looking into the puzzle and thought, does everyone know what people mean by representation in the cognitive sciences? Do neuroscientists even know what they mean by representation? We're looking at various literatures within the neurosciences, the cognitive sciences, and thinking, wow, it just seems like there's very different uses of the term "representation".

Our publication came out a couple years ago now on this, we decided to do a sort of empirical project on this in which we would try to find some evidence, systematic evidence of how the term "representation" is being employed or deployed, whatever your favorite term is in the actual practice. We sent out a survey to a few thousand, maybe more than a few thousand people around the world, and we ended up getting about 800-ish responses from philosophers of mind, from cognitive scientists, psychologists, neuroscientists, and the survey included questions like do you think cognition involves representation?

A binary yes, no. We asked the questions later on about some of the foundational issues in cognitive science and neuroscience. Do you think representations can be embodied? All these other kinds of questions. The meat of the project was to present these vignettes that were supposed to be depictive of your old school neuroscience research that's looking at neural activity in relationship to some sort of stimulus. We presented these vignettes that might be something like a face. We have someone sitting in an FMRI and they're presented with an alternating stimulation between the blank slides, but then also pictures of faces, pictures of cars, all these sorts of stimulation.

Here's the part of the brain that's lighting up, and we present this picture. Here's a time series recorded. Now, here are some options in terms of your responses. Would you describe what's happening with the neural activity in relation to the stimulus as being about the stimulus, or conveying information about the stimulus, or representing the stimulus, and then the standard Likert scale, 1 through 7 of how confident they were in their responses.

We have these different vignettes that try to look at sub questions like, do people tend to describe with more confidence that there is activity localized in a single neuron, or that there is activity distributed in a network, or that that network needs to be embedded within other networks to have the relevant activity that we're interested in? That was the basic overall vignette that we presented to them. Edouard, do you want to add anything?

Edouard Machery
No. I think the only thing that's worth mentioning is actually a few things about the three aspects you just described. One was about the scale at which representations are supposed to be found in the brain. One might wonder, are representation found at the level of neurons or small number of neurons, a dozen of neurons connected in some way, or are representations supposed to be found at the level of brain areas, hundreds of thousands of neurons, or maybe even distributed throughout the brain, millions of neurons? When neuroscientists use the word "representation", do they have any expectations at the scale of organization at which representations are to be found?

Or actually, are they totally ignorant of that question, or have no commitment about that question? The second one was about the causal relation between a stimulus and brain activation. Must this causal relation have some specific properties for it to call this a representation? Must it be fully

reliable or just somewhat reliable for the activation to call this a representation? The third one was the embedding in a broader network, which we conceptualize as of having a function being used by the rest of the brain. That's a little bit the three question that were driving the project and that Luis was alluding to in his description.

Luis Favela
Thanks for elaborating on that. What did we find? Of course listeners can check out the article, the primary article. Then Edouard and I wrote a target article version of that for a philosophy venue, and Doctors Egan and Cao, and Krakauer kindly present their views on that as well. Then we were able to give a response. What did we find? What was the hot conclusion from Edouard's and my work? To put it gently--

Paul Middlebrooks
Everyone uses representation in exactly the same way. That's what you found, right?

Luis Favela
Exactly. What we found was that this work was utterly unnecessary, that there is conceptual clarity and consistency across all the sciences. We are inching towards solving the mind-body problem. That was it. Really, Frankie, Rosa, John had nothing to comment on other than to shower us with praise. Unfortunately, that's not what happened. What happened was that Edouard and I concluded that the concept is, to put it gently, unclear and confused in its application across the various relevant sciences. We interpret a lack of clarity and a state of confusion by an interpretation of the results from our experiments.

One kind of result we found was a hesitancy to take a strong stance on how this activity was described. Do you think the neural activity is about a face? For example. People would pick the middle score, the 4 or 5, that they were not super confident that it did, they weren't unconfident that it was. We took these findings as indicative of a sort of-- there's not clarity on how to deploy these terms as well. Through our other findings, we found some confusion as well. I don't know, Edouard, if you want to elaborate a little bit on the lack of clarity and confusion interpretation we gave?

Edouard Machery
By lack of clarity, we somewhat have a semi-technical notion. We mean that when people use the word, they don't really have many commitments of what follows from that. When they use the word "representation", you might think, if something is a representation, then it follows that blah, blah, blah. We know at the scale at which it is to be found, or we know that it stands in a specific causal violation with something in the world, or we know that it has a specific function.

What we found is that because of this ambivalence that we were describing, neuroscientists, by and large, really don't seem to have any commitments about how the word is used. They just use it in a somewhat free-floating manner, one might say. That's what we mean by unclear. The confusion was somewhat different. We were interested in whether neuroscientists and psychologists are willing to say that the brain misrepresents, or an activity in the brain is a misrepresentation of something.

What we found out is that neuroscientists are extremely unwilling to do that. They seem to be thinking, no, that's actually not the kind of things you can really say. Even if the brain fire, so to speak, when someone sees a house instead of a face, it's not a misrepresentation of a face as a house. That led us to the conclusion that neuroscientists confuse, and it's not necessarily a criticism, it's just somehow it's not a distinction that matters for them, but confuse representations and what philosophers call natural signs.

Natural signs is smoke, it's a sign of fire. When you see smoke, it indicates there is fire, but the sign itself cannot misrepresent fire. By contrast, a map, a representation, can misrepresent a city. It can be a very poor map. It's two different types of symbols, a sign and a representation. Our conclusion was that that distinction does not play much of a role in neuroscience. When neuroscientists use the word "representation", they don't care about the distinction between signs on the one hand and representations on the other. Which goes at the confusion, because confusion just means you use a single word to talk about two different things. That's the idea that we used.

Luis Favela
Great. To finish up. What do we do moving forward? That's our descriptive project, which is, again, the motivation. We were wondering, do we have any systematic evidence for the use of these terms and the relevant sciences? That's one part. Next part is how we interpret the results. Then it's the prescriptive part, which is, how do we move forward? Edouard and I, at least to get the conversation started, suggested three possible ways to move forward. One is we can do-- I think Edouard can speak on this more clearly. We could reform the concept. Edouard, do you want to say a little bit about concept reformation?

Edouard Machery
There's three options. Reforming, eliminating, and understanding, one might say. Reforming is, you take a notion that's not fully clear in science or in everyday speak, and you try to make it more precise. You try to specify the commitments one should have when one uses the word, or you try to draw distinction between different thing in the world. Make it clear that a sign in that representation. That could be one project. The other one is just eliminating, and I know that at least John is sympathetic to that view for neuroscience.

The third one, which at the beginning was an afterthought in the original paper, but became more important in our exchange in mining language, was to try to understand why a notion that's actually in precise, unclear, and maybe confused happened to be playing such a role in science. Part of

the thought is it's actually not an unusual situation that scientists use this kind of notion. It's not bad. The crucial idea is that there's actually a virtue of the lack of clarity and of confusion.

It is, in some sense, functional for how science works. That's a project we, so much in a speculative manner, I think that's-- it's more of a gambit, one might say, rather than a full answer. We suggested, maybe that's the right way to think about the concept of representation. It's unclear and confused, but that's functional, maybe. We need to understand why is that functional? There could be a lot of interesting work there, both about quantifying neuroscience and the history of neuroscience, to try to understand why this notion has the feature that it has. We suggested maybe there's a very interesting project there to be fulfilled.

Luis Favela
The last thing I'll just say is listeners, it might be helpful to think of the term "gene" in biology as one of those concepts that has been potentially unclear, confused by our standard. Where imprecision is a virtue, as Edouard was talking about, maybe the lack of clarity and strict definitions, it's helping conversations happen across different kinds of biologies, where the molecular biologists can talk with behavioral biologists or something like that. They have a loose sense of what gene means, and that's good. It helps them understand each other. It feels too precise, they might not be able to have those conversations that are really fruitful for multi-scale investigation.

The last thing I want to say, Edouard said it in passing, the option of elimination. It is one of my favorites. In the history of science, we see this move to just eliminate, to just murder a term. Just put it in a bag and throw it in the dump. Terms that don't refer to either anything real in the world, empirically supported, or maybe that are so conceptually confused that it's not worth using. My sense, I think I hold it may be stronger, less ecumenical view than Edouard - he's much kinder than I am - my sense is, let's just get rid of this term altogether. That's my last line.

Edouard Machery
It's the first time anyone has said I was kind in academic setting.

[laughter]

Edouard Machery
I'll take the compliment.

Paul Middlebrooks
Are there examples that come to mind, and anyone can chime in here, of the first two options in the history of science of reform or elimination?

Edouard Machery
Germination Phlogiston is an example, obviously.

John Krakauer
Yes, Phlogiston is one. The either.

Paul Middlebrooks
Those terms still exist. I guess it's just they've been discredited as referring to something real. What about reform then? The reason I ask is because language has semantic drift. Trying to reform anything is a fool's errand, perhaps. Was that just off the table because it's just impossible to do, or are there examples where reform has been a successful project?

Edouard Machery
It's an excellent question. I'm very interested in that very question about scientific concepts and how they work in science. I used to be a supporter or a fan of reforming projects because I think there was a job for the philosopher there. It was a way to give me something to do. I take a notion, I clarify it, I propose an amelioration. I think exactly for the reason you're mentioning, I've become actually a little bit concerned about that kind of project. You propose something, and then the semantic drift and your proposal gets ignored or transformed by users very quickly.

It's actually not so easy to find very successful interventions of reform where the original intention has stuck throughout the years after its introduction. I've been looking at a bunch of case studies. Either they don't work, they're not taken on, or actually, they get very quickly corrupted. I think that's exactly for the reason you're mentioning of semantic drift. That's unavoidable in language use. Even in more formal context like science, I think that's just totally unavoidable. I've become actually a little skeptical that reforming is actually this normative, a top-down approach. It's a very successful approach for concepts in science.

Rosa Cao
I think John mentioned in his response article, the invention of temperature. That seems like a case where maybe there was bottom-up reform based on people who are actually using the concept in thinking about it in science, rather than top-down or from outside by info-driven intervention from philosophers.

Edouard Machery
That's a great example. Notice also it took, according to Chang, hundreds of years for the process to work. If you believe Chang, it was just such a very long process, but that's a very nice example.

Paul Middlebrooks
Then maybe we should just give overviews then of the responses to this. I will say something, and then you guys please correct me. John doesn't like the idea of being kind and accepting various uses of the term "representation", doesn't think it's a useful way to proceed, and wants to keep the concept of mental representation alive and to, I guess, reform it. Rosa doesn't accept the results because she pushes back and says that people can use representation confidently in various ways, but the survey didn't necessarily get to those ways.

Frances, more or less, says representation, it's okay to use it because it's a-- you'll have to clarify for me what a causal thin gloss is in terms of relating some measure of brain activity to a mental representation. [chuckles] Where did I go wrong there? Please correct me. Whoever is more vehemently against what I just said can jump in first.

John Krakauer
I'm going to just because I genuinely find this whole thing startlingly irritating. The reason is that things will get better once we actually think about the science properly. It's not about terminology, semantics, or anything like that. It's a failure to think hard about the phenomenology, what's going on. In other words, as I've said to you before, Paul, spend one week with me on a neurology ward and you'll never question mental representations for the rest of your life. They exist. There is representation-rich behavior. This conversation would not be happening without representation.

Paul Middlebrooks
Does anyone here disagree with that statement?

Rosa Cao
I don't think anyone disagrees with that. We just disagree about whether other people who aren't dealing with those--

John Krakauer
I'll keep going. I'm just saying, I'll keep going.

Rosa Cao
-representation get to use the term.

John Krakauer
Mental representation. Representation-rich behavior cannot be disputed. It gets lost selectively. That's why everyone is so terrified of Alzheimer's, why everyone is so terrified of schizophrenia, is because that representational rich, meaning rich capacity that we have gets cruelly and selectively targeted. That's undeniable in my view. The real question is, if you have that amazing capacity, bats fly, mice don't, flying is amazing, what's amazing about humans is this rich mental representational capacity, which can be selectively targeted by injury and disease.

Now, what happened? That's just as difficult to explain as consciousness is. There was a time in the past where consciousness in that form of cognition were put together. Like Tyler Burge, for example, when he talks about mind, talks about the ability to do a particular form of representation and consciousness. Now, because consciousness was so hard and had first-person ontology, it got shunted aside and said, "Oh, we can do the cognitive part because that's more third person. It's more algorithmic. Maybe we can deal with it." That divorce was imposed.

Then, ironically, what was done to consciousness until recently was done now to this particular form of cognition. Let's take it down a peg or two. Let's sensory motorize it. Let's embody it. Let's find a way to get rid of its special qualities. Now, totally in parallel, the word "representation" is used by the neuroscientists. You go all the way back to Hubel and Wiesel, where informational content, sometimes causal, mainly correlational, but sometimes causal, got called representation. It has absolutely nothing to do with mental representation in the kind that we're talking about, but it just took on a life of its own.

I think Luis and Edouard are right, that representation just got its second existence for informational content in neurons that somehow correlates with external stimuli states, blah, blah, blah. Fine. I would even be willing to say you call one representation star and you call the other one representation real. Whatever you decide. What happened, though? Maybe we can tell a neural story with the "word representation" about mental representations. That's what happened, is the word was allowed to overlap because maybe the answer to how mental representations happen is neural representations.

You basically place the same property happening at the holistic psychological level and you stick it on the neural level and go, "We're on the way to an explanation." Complete disaster. My objection is don't use the word "representation" for just informational content because that is not at all going to explain mental representation of the rich kind. Like Frankie [Frances Egan] says in her talk, in her book, if you're going to talk about mental representations, they are going to have to have the properties of external ones. She makes a list of those properties that maps and pictures have.

If a mental representation is going to be worth the name, it's going to have to have those properties. Fine. That's all fine. I agree with her about that. There are real mental representations. You can really lose them. You don't want to get Alzheimer's. They have the same properties mysteriously in our brains that external ones have. All true. Neuro representation, the word should never be used for information content. The final problem then is can we use neurons to explain mental representations if we talk about structural representations in the brain, some kind of isomorphism between the neurons that are mapping onto the real world? That is allowed, neural structural representations.

That's also a non-starter in my view. In other words, the two neural stories that borrow the word "representation" for the true reality of mental representations are just not going to do it. Then we are left with a very interesting situation is we don't know what the neural story will look like for mental representation any more than we know it for consciousness. It may turn out not to have any easy, intuitive basis that information and structure have, and we should just accept it.

What doesn't help is, A, to deny mental representations, which some people do, and B, to prematurely use the word for neural data where it's actually not going to help whatsoever. That's where we're at. I believe in mental representations. Obviously, I believe that neurons and populations of neurons are ultimately causal for mental representation behavior, but not the way the neural data is talked about today.

Paul Middlebrooks
I'm going to let anyone jump in. I have things I can say, but this is your argument.

Edouard Machery
Rosa and Frankie.

Rosa Cao
I almost agree with everything that John just said, except for the part where he wants to get rid of people using the term "representation" in neuroscience because I feel like neuroscience is actually quite fragmented and people go into it from different backgrounds and there are different subcultures, especially once you include computational neuroscience in it. It seems to me that those people should be allowed to keep on using the term representation for these things that don't have very much to do with mental representations.

The only bad part is equivocating between the two and pretending like we have an explanation when we don't. As long as people keep their indices distinct or keep their representations distinct from their representation stars, from their representation with a capital rs then I think it should be fine. You might think practically, it's hard to be disciplined about the use of the term when we're all using the same word, but I think in principle, there's nothing wrong with having a diversity of uses for this word in a diverse--

John Krakauer
Just a little codis to that, Rosa, we now know from the survey that Luis and Edouard did, that there is a consequence to it. We wouldn't even be having Paul's podcast. If the same word wasn't oscillating between these two qualitatively distinct meanings, there would be less confusion. I think that mere podcasts happening is indicative that having the same word being used in utterly different ways is hugely confusing. I've had students come up to me saying they are confused by it.

Paul Middlebrooks
What does it matter that people are confused? Seriously, what's at stake here moving forward if people are using the terms in the ways in which they understand them?

John Krakauer
The distinction I made was people don't understand.

Paul Middlebrooks
Why would it matter to understand it?

Rosa Cao
I think there's a huge consequence, for example, philosophers who are reading the neuroscience literature and think, oh, of course there are neural representations, and that licenses our talk of cognitive representations or first-person accessible representations. Of course, we know that science tells us that. That's a problem.

Edouard Machery
Sorry, Rosa. I think that's not just an issue of philosophy. It is an issue for philosophers relating to neuroscience. I also think it's also an issue for the kind of explanations that we are getting from neuroscience and the kind of models that neuroscientists are actually gravitating toward. There's really different ways of explaining neural processing or neural dynamics, and using the concept of representation, pushes people toward some specific model and away from other models. Luis is probably someone, given his commitment to dynamic models of cognition and of the brain, is very sensitive to that kind of question.

I do think using the word, even if you use it in a very empty way without much content, prime people toward a specific form of explanation of behavior. Indeed, somewhat sympathetic to what John was doing toward assuming that they can get a very easy explanation between neural processing and psychological dynamics or psychological processing. I think that's actually really what's at stake is a kind of explanation we should look for neurodynamics and for the relation between neuroscience and psychology. The stakes are actually not just trivial here, and not just for philosophers. We don't matter a whole lot.

Frances Egan
I'm going to jump in here. I agree with a lot of what John says, but I don't think that this shows that neuroscientists are confused about the notion of representation. I have a different diagnosis for why I think that John is right, that the notion that they have in mind is a purely information-

theoretic correlation causal notion. Then the question is, why do they persist in using representational talk? I have a different diagnosis for that. It's going to take a little bit of setup. In general, my view of mental representation, whether it be what John is calling mental representation or the use of representation in neuroscience, is that such talk is always motivated by pragmatic concerns.

Representational talk is always pragmatically motivated, whether it be in our everyday lives, talking about people believing this or that, or neuroscientists characterizing a structure that they're positing as representing maybe an edge in the world. What are some of the purposes that might be served by neuroscientists using representational talk, intentional talk with a commitment to misrepresentation, all of that stuff that philosophers are really committed to? Why might neuroscientists, if they really mean information theoretic or causal or correlation, why are they using that loaded term?

Here's, I think, the main point of they're using that loaded term. It's because characterizing a structure or a process as representational allows us to evaluate it for accuracy. If you say that something's a representation, then you can ask, is it accurately representing whatever it's representing, or is it misrepresenting whatever it's representing? I think there's a big benefit to that, to being able to evaluate the structures and processes for accuracy. Now, it's not really a benefit that's intrinsic to the neuroscience. It's not really something that neuroscientists care about. I don't think they care that much about evaluating these states and processes as being accurate or not.

The explanatory targets of the project are cognitive capacities. They're not just arbitrary sets of behaviors. They're manifestations of cognitive or rational capacities. We've got the normativity, the intentional characterization built in from the beginning. Given that explanatory target, then the neuroscientist is interested in characterizing how it's possible. How can we detect three-dimensional structure in the scene? How can somebody grasp a coffee cup without knocking it over? The neuroscientists are interested in explicating the causal processes underlying the target capacities that are pre-theoretically characterized in normative or intentional terms.

One maybe craft reason why they might do that is because they have to justify their research and grant proposals. Everybody recognizes that what they're supposed to be doing is character-- or rather, the relevant committees recognize that the explanatory targets are rational capacities, cognitive capacities. By characterizing these structures that are explicated by neuroscientists in purely information-theoretic terms, by characterizing these causal processes in representational terms, by attributing content to them, then that's a connective tissue between their causal mechanical accounts.

That's what they're trying to give, and the pre-theoretic targets that are characterized as successes. What neuroscientists has to do is explain our successes and our occasional failures. That gets done by attributing content, by construing the states and structures that are causally explicated in the theory as being representing this and that, getting it right generally, but occasionally getting it wrong. I think that's the explanation for why neuroscientists might seem to be confused or unclear about the notion of representation. They don't really care about it that much. I think John's right about that.

They're trying to bridge what they're doing with mental representation in the sense that John's happy with these rational capacities that organisms have to succeed at various things and occasionally fail at others. Just one point to conclude here, wrap it up. Neuroscience, unlike say, the science of digestive systems, is answerable to what we might think of as an intelligibility constraint. At the end of the day, the process that the neuroscientists characterize it has to be that the outputs, we can see them as being rational, given the inputs to the process. Digestive science doesn't have such a constraint. I think that's why we find representational talk in neuroscience and not in other branches of physiology.

Edouard Machery
I like a lot of what Frankie was saying. I don't think I agree with a lot. I'm not sure I agree with everything or even much. I don't know that much. I definitely don't agree with everything. One thing, just a few small points. The first one is it's not quite right that neuroscience does not care about assessment. There's a very long tradition within neuroscience, a very influential approach that develops optimality models of what brain neuro-processing is.

You can't do a Bayesian model of the activation in, let's say, V4, if you don't assume that actually it's optimal in representing whatever it is that V4 is about, and so on and so forth. That's really one of the most important traditions in neuroscience. Assessment is actually crucial to some part of neuroscience. Go for it.

Frances Egan
I'm not denying that neuroscientists are concerned with assessment, but I am denying that-- It's a particular kind of assessment. Take the frog and the fly. The frog typically catches flies when a fly moves across its visual field. They do need to explain the success of that process. Optimality considerations, they bear on explaining the success of what's going on, of these behaviors. That's different from semantic evaluability. That's the characteristic notion that comes as a package with representation. To characterize something as a representation is thereby to admit the possibility of misrepresentation. This intentional notion is a very particular way of assessing.

Edouard Machery
I agree with that. If you have a Bayesian model of, for example, color perception, or a Bayesian model for some visual illusions, the assessment is not just at the level of the whole process. You're going to say, for example, that there's a transformation that's accurate or not. I think the assessments will be much more fine-grained than the one you have in mind there. In that specific way of representing neuroscientific models, neuroscientific processing, I don't think this is just, are you getting the fly that gets to be assessed? It's a somewhat different form of assessment as well.

[crosstalk]

John, give me a sec. I just want to make a second point, then I'll give you the thing. One concern I have with your proposal, Frankie, in light of our data, is I think it's clearly part of the story about why people use the word "representation", that they want to be connecting it with this, maybe pre-theoretical or maybe pragmatic, as you nicely put it, explanatory goals. This is no doubt about that. I'm actually quite convinced by that. I don't quite think that when people use the word "representation", they just mean that, because our data seems to suggest that people actually don't treat causal relations the way they treat representation talk.

They're totally happy to use causal terminology to describe what the brain is actually doing. They have no issue with that. They say it's processing a stimulus, it's reacting to a stimulus, and that's totally fine. However, when it's intentional vocabulary, representing, being about, somehow, they become really ambivalent. I think neuroscientists do actually distinguish these two types of terminology, and they're just not using one as a way of talking about the other. That's actually the place where I want to push back. I'm not sure if that's an objection to you, but I think that's a little bit of a wrinkle. It's not really what they mean when they use representations. They just don't mean causally connected to.

Frances Egan
I want to address that, but I know John's trying to get in here. If I could come back on that point.

Edouard Machery
Go ahead, Frankie. Okay, John.

John Krakauer
I just fundamentally disagree with you, is obviously the case, if you ask people and make them suddenly go, "Ooh, maybe there's a distinction here that in my everyday life, I don't actually entertain myself." I agree completely with Frankie that when they use that word, they are basically hedging their bets. They're using it in the informational content sense, and then they've just been infected by the everyday mental representation language as well.

Edouard Machery
This is possibility.

John Krakauer
I can tell you, I spend a lot of time with neuroscientists, being one. They do it all the time. In other words, you're giving them far too much credit. Now, what happens is that when the neural data go from being sensory motor to being more cognitive, in other words, the actual neural work being done is on a more cognitive topic, there's more danger of that infestation happening because now you're dealing with more representation-rich tasks.

It just comes more naturally to take the information content job and give it the word "representation". It's just a consequence of what Frankie's talking about. You can't help yourself. The idea that they have a qualitatively distinct idea of what the neurons are doing when they're representing versus when they're just correlating or representing a stimulus, there's the word, they don't, Edouard. There's no neural--

Edouard Machery
That's my point. John, that's my point. We're the exact really. [laughs]

Frances Egan
Could I jump in here with my point I was going to make to Edouard? I think that's right. I think that there's more to the story about why they use representational talk. One important aspect of it, and this is something that I think Rosa highlights in her really great paper, *Putting representations to use*, and that is that characterizing a structure or a state in representational terms, saying that it represents, say, an edge in the world or a fly for the frog, that has the result of high-- in the causal process is really super complicated. It's very complex.

There's a bunch of things going on distally, then there's proximal stimulations and a bunch of things going on in the brain. Really complex causal process. Saying that this structure represents a bug, stick with the simple example of the frog and fly, selects or highlights a particular aspect of that complex causal story that's important, that's salient given the explanatory target, given that the point is to explain how the prey catching mechanism works.

The point I'm making here is that it's more than just addressing these pre-theoretic explananda. The content attribution or representational talk is still motivated by pragmatic concerns of the theorists. That is to, in a sense, highlight certain aspects of this really complicated causal process, highlight them because they're important or salient given the explanatory target. We're back again to motivating or justifying representational talk in terms of what is trying to be explained.

Edouard Machery
John was saying I was giving too much credit to the neuroscientists. My sense is that you are giving too much credit to the neuroscientists. My sense is actually the notion of representation is much bigger and looser, and it's much more like a free will that actually use with so little content. That I do feel you have a somewhat really restricted and very functional, if not referential, but functional use of the word "representation" of the

concept, in your sense. I do think that might be part of the story, but I think it exaggerates its regimen, one might say. In fact, it's much looser and vaguer.

John Krakauer
All I can say to reconcile the two of you is if you read Sherrington on the stretch reflex, it's very unlikely the word representing. It doesn't get used a lot. You don't say that the muscle spindle represents position and velocity. What happens if you now do Sherringtonian-like work, same kind of work, but you do it in the prefrontal cortex, and again, you're just finding a correlation between something in the world and something in the brain, because you're in the prefrontal cortex and probably doing a more cognitive task than the stretch reflex, the word "representation" is used for the same kind of result, which is just a correlation between the outside world and the inside.

Because it's a cognitive task, to Frankie's point, it gets infected by mental representation talk. Much more likely than when you're down in the spinal cord, even though the neuroscience being done is exactly the same.

Edouard Machery
John, that's a really interesting empirical question, which I want to flag. Here's an empirical question. The use of the word "representation" is going to vary across experimental traditions or parts of the neural system that's examined. Actually, we could easily test that by using text analytic methods. I would get really interesting results.

John Krakauer
You see my point that-

Edouard Machery
I totally see your point.

John Krakauer
-the nature of the work, the nature of the neuroscience work itself has not changed in the spinal cord or in the prefrontal cord.

Edouard Machery
Totally.

John Krakauer
Yet the language has changed. In neither case has the genuine entity realism of mental representation been addressed.

Edouard Machery
I'm happy with that. I'm just want to flag out as an empirical prediction here, which I think would be worth examining whether it's true or not about how the concept happened to be used. I think that's a plausible empirical claim, so I'll take that.

Rosa Cao
I'm trying to think about the disagreement between Frankie and John here because it sounds like you agree about what neuroscientists are generally doing. It's just that Frankie thinks that it's mostly harmless or somewhat helpful because it [crosstalk] what you think the relevance is, and John thinks that it's pathological original sin.

John Krakauer
I don't just agree with Frankie about where it might be coming from. I think where we seem to be is I believe that mental representations are real entities. It's a genuine thing that we need to explain that gets lost in disease. One day there will be a neural story, but the neural representation story, whether informational or structural, will not be the story.

Rosa Cao
Does Frankie actually disagree with that?

Frances Egan
I don't think I do.

Rosa Cao
[crosstalk] in first-person representations, right?

Frances Egan
Yes, but I think always content is pragmatically attributed, even at the personal level. That's a whole other story.

Paul Middlebrooks
Can I ask Luis and Edouard? I don't remember, in your study, did it distinguish between different stages of researchers' careers, their ages? I know that it's underpowered to start asking those fine-grained questions. The reason why I ask is because I was going to make a comment. I'm a

neuroscientist. The people that surround me, and because of that, because of the culture, I probably do this, too, often neuroscientists, when they use the term "representation," are merely talking about the shape of whatever neural activity they're measuring.

When you talk about a manifold, that is a representation of the neural data. They're not trying to connect mental representation with the neural representation, they're literally talking about the neural activity and the kinds of different shapes that you can make out of it, whether you're reducing dimensionality, et cetera. I'm not sure if that's a younger generation that's doing that. Where would that sit? If we're just talking about neural activity, the shape of neural activity, I think everyone would be okay with that, as in the activity of deep learning networks as well. Those representations are just saying something about the shape of the neural activity.

Edouard Machery
We have the data, but our sample size, as you mentioned, is really way too small to look at variation for age. Actually, it was part of the original project that we would be looking at sub-disciplines within neuroscience because Rosa mentioned computational neuroscience, and you might think that it has a different use of the word representation compared to system neuroscience. Actually, we are very excited, Luis and I, by the possibility of actually much more fine-grained empirical work. It turns out that neuroscientists aren't the most easy sample. They're very busy creatures.

We sent, I don't know, 12,000, 15,000 emails just to get a sample of 800, which is not unusual in this kind of work. It's actually what we should be expecting as a response rate. We couldn't look. I think that's a very interesting question to examine whether there's variation as a function of age or as a function of cohort. I don't have a strong intuition. You have one, maybe John has one, or Rosa. I'm not sure.

Rosa Cao
Can I jump in?

Luis Favela
Yes, I agree. Oh, go ahead.

Rosa Cao
Oh, sorry. Just like the particular question, I think I would speculate that it doesn't have to do with age directly, but it does have to do with the subfield, whether people came into it from computer science or from a more computer science tradition. If you go back to David Marr, he had a use of representation that was pretty different from what you see in systems neuroscience. I think in systems neuroscience, representation often means whatever the thing-- a current neural activation, the shape of neural activity. From computer science, it can be anything that carries information about the stimulus.

It doesn't even matter what the further context is, whether it's being used downstream, and so on. Everything is called a representation. Any activation in a deep neural net is called a representation. I think there, the motivation is not the one that Frankie highlights in, we want to explain some kind of representation-hungry activity. It's rather just a way of talking about the way that information is transformed at different stages in some processing system.

Luis Favela
I agree with Rosa's points of view, equals features, especially the kind of Marian tradition. Also, I agree with Edouard, the sample size is too small. It would be great because to say neuroscientists think-- a molecular neuroscientist to some kinds of computational neuroscientists, why did they even go to the same conference? There's a lot of variation there. In a follow-up study, we'd love to see more variation and maybe generational issues. I wanted to respond to that as well, but also tie together, Paul, you asked earlier what's at stake here with these debates.

I think, also, to speak to the theme of your podcast, neuroscience and AI and that relationship. One of my worries is that if people get too comfortable- I'm putting up scare quotes with my fingers- using mental kinds of words to describe non-mental phenomenon, I think that's a bit of a problem. In one concrete case where that's a problem is what we're seeing with contemporary- again, I'm putting up scare quotes- AI. The public at large and the grant funding agencies, they love hearing these terms like “the machine, the system is creative, the system is hallucinating.” They're using these kind of words that we only use in application to minded things.

Maybe the dog is creative, the human is hallucinating, or something like that. We use these terms like representation. It's imported from the cognitive sciences and cognitive psychology and stuff that use it in relation to mental representations, the juicy stuff that John doesn't want us to ignore. We slip back and forth, and then we start using that as a computer scientist to describe nodes in a network. We're so used to applying these terms in the way that it conveys some mental activity. Then we slip into this quickly wanting to describe these systems that, hot take here by Luis Favela. They're not mental. They're not thinking. They're just input/output pattern finding and all these kinds of good stuff. Sorry if I offend anyone.

To describe them as hallucinating, it's just ridiculous. You might say, "Well, but I just need a technical term." Yes, but the public doesn't know that. The grant funding bodies don't necessarily know that. I see that as what's at stake here, it's the importing of this mental talk and applying it to things that are not mental.

John Krakauer
Also, to that point, just one thing to your point, Paul. Oh, "Ask younger, sophisticated generation, we're just using the word representation for the shape." Is this false? I know all those people. Let me give you a quote.

Paul Middlebrooks
I just had someone on my podcast I can point to. Of course, there's variety, but there are people who use it in that way.

John Krakauer
Here's a quote, "Where neurons represent the formation of decisions." That's from a manifold paper. I will not say who it is. In other words, where the psychological thing, decision making, is attributed to the manifold. It happens all the time.

Paul Middlebrooks
I agree that that's a problem. I completely agree with you that that's a problem. To Luis' point, also, it bristles me still when people call an artificial unit a neuron. It's something as simple as that. AI, I feel like because AI is so popular, has essentially co-opted these terms, like hallucination or whatever mental capacities you have. Even at that neuron, I really hate it, but we can't control that. We can just bitch about it.

Edouard Machery
I will just make a [audio gap] -logical point, I think, mostly for the listeners. We are here a somewhat biased sample of opinions about neural representations. I think it's quite striking that despite our disagreement, we all agree that there are neural representations in a very realistic, hardcore sense that we can collect to psychological processing. I think there's no one here on this panel who makes that claim.

There are actually plenty of philosophers who actually are committed to that view that we can actually easily move from a psychological story about reading and then look at changes in the brand or steps in the neurodynamics and say, "Oh, this is a detection of the phoneme. Oh, this is a detection of grammatical structure." Here, it's somewhat a select sample of individuals. All these agreements are among friends in some sense, people who are in various way, skeptical of the notion of neural representation. That's maybe just for the readership. Actually, there's other sets of philosophers and neuroscientists.

The second thing I want to say, and then Frankie will follow, is just also small comments about the sociology. I do think neuroscientists do care extensively about what representation means, about what the world representation means, about what representations are. You can just go on social media on Bluesky and every third week you're going to have a long thread with 75, 100, 200 neuroscientists going on and going on about what representation are. Every third week it is the same thread with somehow different words that happen to be used.

There is actually a lot of demand, I think, in neuroscience to clarify these notions, to help somehow build some form of consensus about the way we should have this terminology or the way we should use this term. It's not just philosophers who are just trying to just look a little bit, I think neuroscientists themselves are actually really interested in understanding well, what our foundations and so what do we mean by that? Thank you. Sorry, I interrupted. I went on.

Frances Egan
Thanks, Edouard. Just to clarify my position, I agree that there's lots of misuse of representational talk by neuroscientists. I think it can play a useful role, and so I'm trying to characterize what's going on in the cases where there's some point to it, where it is playing a useful role. There's lots of cases like that, of egregious misuse of representational talk. There, it's not serving any of the functions that I try to characterize in my book. What's at stake? I agree with both of you, with everybody I think here, or at least with Edouard and John, that what's at stake are the conclusions that philosophers among others draw in thinking that we've made a lot of progress in understanding these mental capacities and understanding consciousness and rational decision making and so on.

Using the same vocabulary can prop up that idea, can support that idea, they're talking about the same thing, so there's much more progress. It's connective tissue. There's an upside and a downside that you need to connect the work with the explanatory target. You shouldn't make it look like, "Oh, we've solved the problem of intentionality. Now we know how it is that brain states can refer to." I'm not saying that at all. I think philosophers are guilty in appealing to neuroscientific work, in support of their particular theories of content, thinking that these views are solved, that the neuroscientific work is actually solving the problem of intentionality. I think that's probably the most flagrant misuse of the work.

John Krakauer
I was just at a meeting on representation that Kenneth Aizawa held at your alma mater basically, Frankie, right? It was interesting that it's not that we shouldn't be doing neuroscience on these hard questions of mental representations. I'm not a nihilist; in other words, there has to be a way of using neural evidence to update our theories, algorithmic theories of what's going on in representation-rich behavior. I've always said you should send out all your hounds at any level of evidence. The critical issue is what is the explanation that is constructed out of these different forms of evidence? What I'm saying is, is that the explanation using all forms of evidence isn't itself going to be articulated in neural language.

The mistake is to hope that you see a homunculus of the phenomenon you want to explain in the neural data. That's what's not going to happen. Just like when you close your computer, there's no Word document in your computer. There isn't. It's a Word document when you open it and you use it, and then it takes on that true representational format for use. Now that transformation that occurs when you open your computer and you get a genuine representation that you can look at, neural data may one day tell us how we open up the representation in our heads to use, like we

use in external representation. Just like computer science work, we will need such an explanation but you don't go looking for a Word document in the computer before you open it. That's [crosstalk]

Luis Favela
Are you actually [crosstalk]

Rosa Cao
I totally disagree with that. Fighting words. There's definitely a document in your computer.

Luis Favela
It's a thing, [crosstalk]

Rosa Cao
Exactly. This is so bizarre.

John Krakauer
It's not an external representational format. When Frankie says-- of course there's information, there's information in your computer that can be turned into a representation but you're doing it now, you're saying it's a representation. You have stuff in your head right now, Rosa, that you're not saying, like phone numbers from past homes and things that you've seen. You're not using them right now, you're talking about--

Rosa Cao
They're not current but I think I still have them.

John Krakauer
They're not in that external representational format that is used when it's represented. It's in a stored format. What I'm saying is information through transformation can be turned into representation, but just because transformation of information can turn into representation doesn't mean you call the initial information representation before the transformation.

Rosa Cao
I agree with that, but I think if it's sitting in a system where there's a well-established mechanism for consistently turning it into a representation every time, turning into a current representation every time, then you can say that it's a non-current representation that's still there.

John Krakauer
Well, then, that's fine. Then we're just back to representation; Representation R and Representation Star. I'm just saying that what we need to do with the neuroscience is how do you get information plus transformation or process into a representation for use. That's what the neuroscience should be after, not prematurely providing the property to be explained on the information.

Rosa Cao
I think there's a difference between totally unstructured information that could never be used by the system versus something that is already structured. It's in like a standard format for retrieval and you could retrieve it at any time. Those just seem like very different things. I don't think we should mush them together. I think that the thing that is in a standard format that can regularly be retrieved as much closer to the current representation than it is the unstructured information.

John Krakauer
Close is a squirrely word. All I'm saying is there is a difference that makes a difference before and after retrieval.

Edouard Machery
I do think you really don't want to use computers as your test study to argue that there is no representation at the level of the hardware. What's remarkable about computers is that you're compiling high-level language into lower-level language, up to machine language. That means literally that when you've got an instruction in a high-level language, that instruction exists at the level of the hardware.

John Krakauer
Exists as information.

Edouard Machery
It exists as a representation. This is how companies brand, a computer. Computers have representation. Brands don’t! It's a main distinction between a brand and a computer.

John Krakauer
Who's looking at the representation in the hardware?

Edouard Machery
It's not very possible.

John Krakauer
You're not understanding my point, this is the same mistake again.

Rosa Cao
It sounds like you're saying it doesn't exist unless you're looking at it.

Edouard Machery
You just stipulate. That's also just strange idea, of course it exists.

John Krakauer
Of course, the information exists, but representation is a process. It's a thing that you do, it happens at the moment of use.

Edouard Machery
You are just stipulating it. You are just stipulating the use and you're basically out of thin hair, just stipulating a use here of the word representation that there's no reason that anyone would be granting. Look, I have a map. No one's using the map. It's hidden. It's actually, I buried it in my garden. No one will ever use it ever again. It's still a map of Paris, even if it's not used right now.

John Krakauer
That's true. I would just make the point. The point there is that what happened, I gave this example. Let's say I'm trying to draw a map for all of you guys right now for how to get from my apartment in Lisbon to a really nice café. I'm drawing it and half of it comes out of my pen, and then you distract me and I stop drawing. Half of it has come out onto the paper and the other half is about to be drawn by me onto the paper. What I'm saying is that I'm in the act of drawing it and I'm halfway through drawing it. It's in representational format as I'm getting it coming out of my pen onto the paper to join the representational format on the piece of paper.

Now, when I draw that whole piece of thing and I expressed it as a representation, and I drew it, I completely agree with you, Edouard, it froze. It's like a frozen accident. That representational behavior is now locked onto a piece of paper. What I'm saying is that if I was distracted from writing that final half of the map, and I thought about something else, I went, "Oh my God, I have to do this," and it's now gone, it's no longer in my head in that representational working memory-like format that it was being used to complete the map.

All I'm saying is the format it was in when I was finishing before interruption, versus now no longer doing it, and I could complete it the next day, those are two very different things. I'm saying one is representational, a time of use, gets frozen on the paper, but when it's in my head before I conjure up again, it's no longer map-like like the map stored in your cupboard. It's in a different--

Rosa Cao
Wait, John. It sounds like you think that memories when they're not actively being recalled, are not representations?

John Krakauer
I think they are in a format that can be transformed into a representation.

Rosa Cao
You want to stipulate that they don't count as representations because--

John Krakauer
That's right. Yes.

Paul Middlebrooks
How was the whole representation? How was the Krakauer representation different from just a mental phenomenon? It sounds like when you think it, that's the representation. How's it different from--

John Krakauer
That's right. I think you're right. I think representation is a conscious, overt moment of use.

Paul Middlebrooks
That's a very particular, specific definition of representation.

John Krakauer
It is because it's the one that we all are implicitly hinting at when we talk about thinking, planning, creating, and AGI.

Rosa Cao
I don't think it's the only kind of first-person representation that we care about. If we were worried about Alzheimer's and not being able to remember things, we worry about them not being there to be retrieved, not just--

John Krakauer
Again, I'm not concerned about that. I'm just saying that implicitly what we are genuinely interested in when we talk about AGI, for example, is conscious, overt, representation-rich behavior that's happening. Of course, we're interested in all the things that go wrong in the substrate and how it makes that transformational ability go away. Of course we are. I'm just saying that I have a problem using representational language, which is at least tacitly a lot of the time referring to that kind of behavior. Then we imbue sub-personal processes with the same property.

Edouard Machery
I think Luis wanted to jump in earlier, and I cut him. Maybe we should give you an opportunity to jump in.

Luis Favela
No, that's okay. I was sitting here. I don't know if any audiences that can see this video, but I was doing like Mr. Burns, "It's good." I was just enjoying watching the heat flowing through people. [laughs] Just kidding. If I had to describe John as a meal, I would say it's a meal that has a label that says very spicy and hot, but then when you take a bite into it, it's actually cool and creamy. John says things like, "I totally disagree with all of you." Then he'll say things that are very sympathetic or along the same lines.

One thing that came across as spicy, but I think ended up being cool, was I'm still not quite clear what John thinks the role of a neuroscientist is for explaining these kinds of phenomena. John sounds to me like a methodological solipsist, someone who thinks we can theorize about mental states on their own, and they are this kind of domain in which we need to explain that kind of phenomenology or those kinds of relations, to other ideas and conscious states, and things like that.

He says, "Of course there's neural activity that's related to it. Maybe the job of the neuroscientist is to look for those correlations," or the strongest term that I picked up from him was, "Look at the transformation to neural activity, to those higher-order mental states." That just seems like extra crumpet on the side for him but doesn't really tell us anything about what mental states are like or what representational mental states are like.

I was still not quite sure what neuroscientists are doing or explaining that psychological scale. I think they're useless for John Krakauer. We should all be psychologists, we should be Fodorians and just theorize about the mind.

John Krakauer
Why can't I say the same thing to you? When Sherrington was doing the stretch reflex, why wasn't he doing particle physics to explain it? In other words, why is it that circuit neuroscientists are not forced to be particle physicists when they do their work? Why do psychologists have to be neuroscientists? Are neuroscientists also methodological solipsists because they're not doing particle physics when they do their circuit neuroscience?

Luis Favela
Well if it came across as hot and spicy like you. Yes, probably.

John Krakauer
I'm just saying that that's what emergence is. I find it ironic that the best [crosstalk] effective theories, and basically psychology is an effective theory for mental phenomena, just like economics and sociology are effective theories. We don't ask people. Anderson's main point in *More Is Different* in 1972 is that you have [crosstalk]

Edouard Machery
I think there's really interesting questions here about levels of explanation and the role of notion of representation in tying, or maybe actually should be untying, these levels of explanation. I think that we would learn a lot of people who are interested in neuroscience and psychology, by looking at how levels of explanations are articulated in other domains of science, for example, in physics. In physics, we also have a different level, a different scale of energy, different models that can be related to one another.

It's very unusually the case that one notion that's going to be very much in line with what John was saying, that one notion at a higher level, is actually found at a lower level. When you have a different model, a different scale of energy, usually, the explanatory primitives stay at one scale. It's not that we don't have any understanding of how the scales are related to one another. In fact, we often have very precise understanding in at least some domains of physics, not everywhere. In some domains of physics, we have a precise understanding about how model at one scale of energy can emerge or can be the result of what's happening at lower scales of energy.

John Krakauer
Which is why I said transformation. In a complexity science, one has to distinguish three different disciplines. There's discipline on level X with its vocabulary and explanatory primitives, and then there's the discipline with its explanatory principles at level X minus one. Primitives remain at their levels. There can be a domain of science where you look for a congruence, a transformation between level X minus one and X, yes.

What people mustn't misconstrue is that if you have that transformational understanding, as you correctly say exists in some areas of physics, that that transformation will make the vocabulary of X no longer necessary, that it will be **[inaudible 01:17:55]** way. That is what I think Luis is actually trying to infer, and other philosophers, that if we just had the transformation and we had the lower level X minus one, we could change our primitives for level X. That is what I think--

Edouard Machery
I don't think anyone here wants to say that. I think another attitude, which is not the one you're describing, is actually hoping that we can get, so to speak, a reduction of a higher-level primitive, let's say, representation. You can say, "Ah, representation just happened to be this quiet, complex thing that I can describe in lower-level primitive terms."

John Krakauer
It seemed that was what Luis was saying because he was saying, "I'm a methodological solipsist." Neuroscientists are methodological solipsists when it comes to physics.

Edouard Machery
I don't know. Luis, what do you think?

Luis Favela
Sure. Why not? It still wasn't clear to me if we're doing "study higher-order phenomena," like what John's calling cognitive states with rich phenomenology, I still don't know what the job of a neuroscientist is. Is the neuroscientist supposed to elicit the transformations or illuminate the transformations that are related to the psychological states, or is the psychological researcher supposed to illuminate both transformations or it is supposed to be both [crosstalk]

John Krakauer
I said, send out all your hounds. I said you can have confirmatory evidence at the neural level. We wrote this in the behavior paper in 2017, that can break the tie on psychological theories and help you update psychological theories. When the update occurs, the language of the update will still be in the original x primitives that borrowed updating from the neural data. Confirmatory updating evidence for neuroscience to break the ties of psychological theories, rule out others, and update them is fantastic but the explanations themselves will not be in neural primitives.

Luis Favela
Does the updating go the other way, too, for you, John?

John Krakauer
I think so.

Luis Favela
We can update our neuroscience theories based on psychological scale?

John Krakauer
Well, that's where temporal difference learning came from. Temporal difference learning was a mathematical theory long before we went looking for the neural data for it.

Luis Favela
I don't know that literature, but we have these levels that have not been explained by a lower level, of course.

John Krakauer
I looked at those traces and went, "Oh, that looks like temporal difference learning," basically. I'm bastardizing the story a little bit, but was able to use mathematical theory and abstraction and see it in the neural data. It was one of the most fruitful moments in neuroscience in recent day. If it hadn't been for his psychological mathematical insight, it would not have been as obvious when the neural traces were looked at.

Paul Middlebrooks
I'm missing how that updated the neuroscience primitives.

John Krakauer
I'm not saying the primitives. I'm saying that I think the question was, what happens when psychological-level theories can provide insight on how to look and interpret neural data?

Luis Favela
I was making a stronger claim, which is how I interpreted what you were saying, that you can arbitrate psychological theories by looking at neural data. I was wondering, can we do that the other way as well?

John Krakauer
Look, I thought that maybe that one doesn't meet what you want. I would say I would be very surprised if it can't be a two-way street. I would be surprised.

Luis Favela
Now, that's hard for me to reconcile with this emergentist effective theory approach, because it seems like there's some unidirectional relationship.

John Krakauer
I don't think that's in pleasant effective theories at all. It'd be very odd to say to economists that you really would be much better if you were chemists. It'd be very odd to say to a basketball coach, "You did a good job bringing them to the NBA Finals, but if you'd done some brain imaging, you'd have been even better." It would just be very odd. All I'm saying is every area, every discipline has its primitives, its ontology, and we should allow that, even though there can be another discipline that moves between them that can help them but that's a different science. They're just different disciplines.

Paul Middlebrooks
Isn't the grand goal to connect different disciplines? Yes, you use the same language at your own level. What you called earlier, I think the phrase was, "A complete disaster," in terms of relating neural activity to mental functions, I would call an ongoing project because it still is a goal to relate neural activity to mental functions, whatever language we're using in whichever level it is.

John Krakauer
Relate is a bit of a filler term. What does relate actually mean? Correlate?

Paul Middlebrooks
Have some explanatory power in terms--

John Krakauer
What does that mean? What does that actually mean? That's what I'm saying. It's very easy to utter such sentences but what I'm saying is--

Paul Middlebrooks
You're fine. Correlation. Just correlation is fine.

Rosa Cao
I think you want to look for difference-makers. You want to see what changes at the psychological level if I make these manipulations at the neural level. That's one explanation that's stronger than correlation.

John Krakauer
It's causal. A stroke. I'm a stroke neurologist, and we know exactly where the lesions are to make you aphasic, the different kinds of aphasia. I can tell you all of that, exactly where you will have a causal consequence. That in itself is not going to tell me much about language and how it's computed at all. It's a start, but it's hardly very richly explanatory for me to say that I can make you aphasic by putting a lesion in the inferior frontal [crosstalk]

Rosa Cao
That's right. That seems like a non-sequitur. It seems like it helps explain strokes. If you want to explain language, then you want to know what kinds of manipulations do you do that allow you to change particular things about the [crosstalk]

John Krakauer
It does.

Rosa Cao
The more specific, more fine-grained things.

John Krakauer
Okay, but I think in the end of lesion analysis.

Edouard Machery
I agree with Rosa. It's definitely more than correlations. It's of the right kind. Also, if we look at, again, I'm pushing back for things not a myopic view about the relationship between neuroscience and psychology, what's happening in other areas of science. Here's a common relation between models at different levels. Derivation on the assumptions. You take a model at a specific scale of energy and you assume, for example, that it has an infinite number of particles, that model can be derived from a model at a lower scale of energy. By derived here, it's mathematical derivation.

In some other parts of science, more successful, I would say at least at this point than neuroscience and psychology, because they have a longer history, maybe because their object is simpler, easier to understand, you have literally formal derivation of models at different scales under values

idealizations. You need always to idealize. For example, you need to assume an infinite number of particles. You need to assume that some quantity goes to infinity, and so on and so forth. This is really what gives you the understanding in these other areas of science. It's not just a manipulation here give you some outcome there. It's literally, "Oh, that model is true at this level higher over there."

John Krakauer
I totally agree with that. Sometimes that works. As my brother said, sometimes you look under the hood and then you don't. I do find it interesting, though, that whenever these conversations are had, it's always statistical thermodynamics and the ideal gas laws and the kinetic theory of gases.

Edouard Machery
No, that's not true.

John Krakauer
Always the one that comes up. Give me another one, then.

Edouard Machery
For example, you can explain the kind of work that Bob Batterman has been doing about emergence phenomenon where they have a critical point. That's also of the same kind. We have a mathematical understanding about why the higher-level description depends on the levels--

John Krakauer
I'm not against that. Again, I hear these arguments all the time, for 20 years--

Edouard Machery
No, of course. They're not new.

John Krakauer
All I'm saying is that I think we just have a very interesting discussion here, which is, what do we think in 20 years' time, when a psychological language and things like mental representations and all the cognitive science of people like Chaz Firestone make, all the people who do lots of great work in psychophysics and psychology who don't refer to neurons, are we saying that other than this transformational knowledge of how you get one from the other, that when you construct explanations of these higher-level cognitive acts, that you are going to now start uttering sentences like you do for the stretch reflex with neurons in the sentences?

That's the question, is will we put neurons in the sentences naturally, like we do now for stretch reflexes?

Rosa Cao
I think this happens in popular culture now, actually. People talk about their dopamine levels.

John Krakauer
But it's completely wrong.

Luis Favela
Yes, it's wrong.

Rosa Cao
Maybe not very accurately, but it seems like they are happy to expand their everyday language to include terms that had previously only been in neuroscience.

Paul Middlebrooks
That's actually a case, I think, where people are imputing dopamine. They're talking about dopamine on a psychological level, and they're sneaking it in. I don't think they're talking about dopamine as a brain process in most of these popular examples. There is, I'm surprised, John, you're saying in the language of neurons, but we're beyond neurons. We're in manifold land. We're in topology land. We're in shapes of neural activity that are getting closer as explanations, line attractor land. If you can map a decision along a line attractor in a dynamical regime. That's a closer explanation.

John Krakauer
That's true. When I wrote the paper with David Barack, again, the line attract is not making a decision. That's a mereological balancing. It actually does get used. I can again give you quotations. What I'm saying is, I think that the idea that there will be primitives, dynamical objects, as David Barack called them, that you can combine to do a computation and that we get intuitive insight, like the beautiful work that David David Sussillo and XJ and others are doing, where you have some idea of these dynamical primitives. Yes, but we are looking from the outside in, going, "Oh, that's a delayed match-to-sample task. That's an alternative for its choice task," which we understand qualitatively.

Then we see how you would do combinatorials with those dynamical primitives to construct that. When I asked David Sussillo, "Who knows that to combine them in that way? Who frames the task to then build it. Ah, Stefano Fusi, beautiful paper in *Nature* late last year where there's a geometry for switching between two contexts. The amazing thing in that paper from late last year was in one minute, you could instruct a human,

"These are two contexts," and one, rather than thousands of trials of training to construct that geometry, it was instantly configured. The question of what's the upstream cognitive process that allowed the construction of that geometric structure, completely unknown.

In other words, the actual stuff that we're interested in, this cognitive understanding is upstream of the phenomena that you're talking about. That's what's puzzling. We don't have a neuroscience to that upstream conscious understanding bit. We see its product downstream, which are fascinating. Don't get me wrong, I love that shit. I'm just saying it's downstream of what we're talking about when we talk about mental representation.

Luis Favela
That's something that Edouard pointed out earlier towards John. That's another great stipulation of that relationship. You could have an identity relationship. I think with Rosa, the example with dopamine, the details might not be right, but let's tell a story that is right. Patricia Churchland, neurophilosopher, has this story that she talks about where she says, "I came home and if I used my full psychology, I would talk to my husband and say, I had a really stressful day. I need to have a glass of wine to relax." She said, "There's no real principal reason why I couldn't say to my husband, 'My cortisol levels are up. I really need some alcohol in me right now to lower the dopamine.'" It's not the same thing. It's an identity relationship. It's identity between stress as the cortisol levels, or whatever is the right [crosstalk]

John Krakauer
As [crosstalk] pointed out a long time ago, saying pain is just C fibers. Well, no. The C fiber side has certain properties that a third person and you can characterize, and pain is something else. They're not equivalent. They're not. Pain has the ontological subjective first-person part, and yes, it can be causally related to C fibers, but to say that C fibers and pain, the same thing is beyond bizarre. It's beyond--

Luis Favela
You're not a philosopher, John. This is not bizarre to--

John Krakauer
[crosstalk], he was the one that a philosopher pointed out the fallacy of that [crosstalk]

Rosa Cao
John, now you're starting to sound a lot like some kind of mysterion. You said we ought to treat representation the way we treat consciousness. It also seems like you're saying we couldn't even in principle make sense of representation in neural terms in the same way that heart problem people in consciousness say we can't even in principle make sense of consciousness in terms of physical.

Luis Favela
Yes. Rosa, you made my point better than me.

John Krakauer
We'll come back with very interesting necessities and sufficiency conditions. We'll update our psychological theories with neural data, of course, but just like we don't go around saying we should have physics explanation for stretch reflexes. Why don't we have this? Why don't we have a podcast on that, Rosa?

Paul Middlebrooks
Why is it always down to corks with levels? Think about a neuron.

John Krakauer
What I'm saying is what happened here is that whether it's consciousness, mental representations, or stretch reflex, they're all part of the nervous system and they're all made up of neurons. The idea is because it's made up of the same stuff and because we've done quite well thinking about the stuff at the level of individual neurons and then populations, we should get the whole story up the neuroaxis. In other words, the explanations should cash out in the same way because it's made of the same stuff. What I'm saying is that just not true. It's not true in physics.

Astrophysics and particle physics are different disciplines with different objects, and we have great trouble, as you all know right now, reconciling gravity with quantum mechanics. In other words, all I'm saying is that the way in which neural data will be used to think about higher-level cognition will just be used in a different way to the way that neural data is used to explain CPG, saccadic eye movements, and stretch reflexes. It's just going to be a different way we use the data. I'm just saying right now we don't know how we're going to use it.

Paul Middlebrooks
Agreed. What did you say? Gravity versus some other?

John Krakauer
That we haven't yet got a unifying theory of the very large and the very small.

Paul Middlebrooks
The very small. That will be Part 2 of our conversation today.

[laughter]

Edouard Machery
We're going to have to bring new host, I'm afraid, a new guest. I also wanted to push back a little bit. There's been in the discussion, and maybe accidentally somehow in, I think that's part is the way John frames the issues at a personal level, there's a bit of an alignment of a bunch of different notions, personal, psychological, and representational. I think these are different notions. A merch of psychology is actually not partly concerned with the conscious, of what people are aware of. Many of the representations that cognitive scientists postulate are not representations that people are aware of having. They're very different.

If you work, for example, on reading so transformation of Grapheme onto Phoneme, there's literally 40 years of work in cognitive science on that topic, you're going to be postulating a bunch of presentations that explains how a Grapheme can result in a Phoneme. They're all representation. None of that is meant to be connected to, in a way of conscious life. Much of psychology, when they're talking about representation, it's really not what John seems to think they're talking about. It's not the first personal experience of the world. It's not imagination. It's not memory, first-personal experience. It's really something quite different.

John Krakauer
Just to be not mischaracterized. When I spoke to Tyler Burge talking about mind, he talked about a particular form of representation and consciousness, and it could be either or. I have no problem with there being unconscious representations. I'm writing a paper with Jake Quilty-Dunn right now on language of thought, which can be implicit. All I'm saying is that it's very important to distinguish between those cases where there seems to be implicit psychological representation, Tyler Burge makes the same case, versus the existence of implicit policies. What I'm saying is a lot of what is called representational that is implicit can just be explained in terms of policies. They're not the same thing.

Edouard Machery
That's a completely different question.

John Krakauer
Please don't say that I don't believe that--

Edouard Machery
Oh, I didn't say anything about you, John. I just say it's important to distinguish issues such as your experience of the world that involve, let's say, a first-person grasp and the kind of things cognitive scientists are doing where they postulate representation. The two are quite different from one another. It's not quite clear to me what justifies a postulation of this second type of representation. On your view, John, for example, is very clear why we want to postulate imaginations. We have a first-person experience.

John Krakauer
Just to be clear, I just want to say it's absolutely true that when we talk about reflective mental representations of the kind that everyone would like to go after intuitively, they have that form. I have no problem talking with Steve Fleming-- We're writing something now on perceptual representation and how that might have been the base for conscious reflective representation. Representation is a term that should have properties that survive them being conscious or not. What I'm saying is that the kind of stipulations of the LOT framework of representations, the perceptual stuff that Tyler Burge, Ned Block have done, neuroscientists, I can say those distinctions made by those philosophers, they never even talk about those very much. That's what I'm saying.

They just go straight from V1 through to prefrontal cortex and use representational language all the way through. I'm saying that there has to be a point where you go from sensation unconscious to perception and representation, which can also be unconscious, like blindsight. There is a division there, even in the unconscious realm, when you use the word representation.

Paul Middlebrooks
I want to just jump in here. Sorry, Edouard.

Edouard Machery
Go ahead.

Paul Middlebrooks
Frankie, I don't feel like I know you well enough to be calling you Frankie yet at this point.

Frances Egan
That's okay. Go ahead.

Paul Middlebrooks
Everyone else is, so I will also. You've been fairly quiet. I just want to tell you to jump in if you have comments or what you've been thinking in the past 10 minutes of conversation. Of course, if you don't want to chime in, that's totally fine. I cut Edouard off there.

Frances Egan
Okay. We've moved on now to personal level. Let me just use that umbrella term for the sorts of representational talk that we use in ordinary lives. I'm not talking about the science because I think if you look at the science, the point that I want to beat is that representational talk and content attribution is always pragmatically motivated. The issue then is, what are the various purposes or functions that representational talk can serve in these various domains? I think I've talked a little bit about the sciences, but I think that in belief attribution and explaining behavior of ourselves and of others, in thinking about perceptual experience, we talk in representational terms. This doesn't necessarily connect with the dispute that we've just had, but I think that we're modeling mental processes in the various domains in terms of representational notions that apply primarily, I think, in language.

What we do is we model inaccessible mental states and processes in terms of these public uses of representation that we understand. I think that belief processes, we think about them as inferential processes. They're not inferential processes. They're causal processes that can be modeled by relations among sentences. Those are inferential processes. With respect to perceptual experience, I think we do talk about representing the world. I've got a completely different account of that. It's a type of adverbialism. I think that we model our internal experiences, our perceptual experiences in terms of what's going on out there.

We use the external world model, our ways of conceptualizing the external world, to talk about our perceptual experiences. That serves our purposes pretty well. Again, one of the general points of content attribution or representational talk is allowing evaluation for accuracy. With respect to modeling mental processes that underlie action and the explanation of behavior, in linguistic terms, we can model perceptual processes, what we call perceptual inferences in terms of-- Let me put it this way. We can characterize our experiences, what kinds of experiences are like what, what kinds of experiences lead to other experiences, and so on, in terms of categories that are based on our understanding of external objects and properties. Again, representational talk, we move from what's out there. We model our inner lives in terms of what happens to our external lives.

With respect to the question, how does the science hook up with either our ordinary ways of understanding high-level experience, high-level thinking, or how does it hook up with the fairly primitive sciences of those things, I think that's an open question. I'm pretty convinced that representation is going to play the same sorts of roles as we make progress on understanding at the level of personal experience, at the level of thought and action. The same sorts of roles that it plays in the sciences, namely, simplification, being able to talk to each other about it, about private experiences in terms of our understanding of the public shared world, evaluating our experiences and our thoughts for whether they're accurate or not, and so on.

I see representation as playing the same role that it plays in the sciences in everyday life, in thought about personal level experience, and so on. That's just my two cents on how the sciences and these higher level--

Paul Middlebrooks
I just wanted to bring this up in case-- Many of the ideas that you were just discussing will likely be in your book. I'll talk about it in the intro also, but I just thought I'd do a fancy screen share real quick to show people.

Frances Egan
The bottom's cut off, so it's-

Paul Middlebrooks
Oh, sorry.

Frances Egan
-waiting mental representation. Thank you for that. Anyway, that's all shoot over in the book.

Edouard Machery
Quick question for you, Frankie, following on what say. It's really interesting. When you say that in everyday life, in a non-scientific context, we explain behavior by means of presentation, what do you exactly mean? Of course, we provide intentional explanations of what people do in terms of beliefs, desires, in terms of what they want, what they need, what emotions they might have, which are the self-intentional entities, but I would've thought this is not quite the same as explaining in terms of representations.

One might have a view that, of course, there's intentional explanations, but that's another step to just say that, thereby you're committed to anything like representations. My colleague, now retired, John McDowell, had that kind of view. Of course, intentional explanations are just fine. People have beliefs, desires in some sense. You explain them, but this is not representational. When you say it's representational, you're actually bringing something else. Any thoughts about that, or do you just want to collapse intentional and representational for--

Frances Egan
I think that we explain behavior by positing in terms of beliefs and desires. I think that those explanations are typically true. I think that what we're explaining our behavior in terms of are either complex causes or complex sets of dispositions. We model those in terms of the sorts of relationships that hold among sentences. I don't think commonsense psychology is representational at its base.

Edouard Machery
I see.

Frances Egan
I think that what we attribute are certain very complex causes, and we characterize those. We can't get in the brain and discover those causes, but we characterize them in linguistic terms in terms of that clause attributions, content attributions, and we model causal relationships that hold among beliefs and desires in practical syllogism, for example. Why the relationships that hold among linguistic objects.

John Krakauer
Frankie, just so I understand it, when you're drawing those fantastical ibex on your farm and you draw one--

Frances Egan
Oryx.

John Krakauer
Oryx. My bad, sorry.

Frances Egan
Very important distinction.

John Krakauer
Sorry, I know.

Paul Middlebrooks
Jesus, John.

John Krakauer
I know, Chuck. When you're drawing art for me, is that not representational behavior?

Frances Egan
Yes. Behavior is representational, for sure.

John Krakauer
The act of drawing a picture for me to correct my egregious error, that seems to be very psychological behavior to me.

Frances Egan
Certainly, my drawing of the oryx is a public representation.

John Krakauer
You did it, right?

Frances Egan
I did it. I produced a public representation. How did I produce it? Now, we're in the realm of what processes were going on when I produced it. We can characterize those internal processes in representational terms in the way that I've been talking about, but there's no question that I produced a public representation.

John Krakauer
That's like that analogy--

Frances Egan
I think representation has its home in public representation in language. That's the home base. We use that to talk about all kinds of things that are not public. We can talk to each other about our perceptual experience, and we can talk about them in terms of stuff that's out there.

John Krakauer
People with brain injury can selectively lose those representational behaviors. I think we're going to have to have some neural explanation for the tissue that is responsible for being able to do that remarkable construction of a public representation. We've done work showing that you can lose the ability to represent shapes that you can subsequently draw.

Frances Egan
I agree with all that. Yes, that's right.

John Krakauer
That is a unique representational ability, unique in just the same language is, and it's got nothing to do with language. Just to kick back at Luis, I believe in a neural story one day for that ability, but I think that the word representation should be for that ability. You shouldn't describe that ability to the neural explanation itself. It's going to have to come in different language.

Frances Egan
I agree with that. We agree on that.

Edouard Machery
Quick question. Why do you think it's important to describe it in representational terms? Isn't that just an intentional activity? I'm not sure what it adds to set a representational activity. I'm not sure what the word representational here is.

John Krakauer
I'll tell you, it's what Frankie said in her book. The thing about representation is, and I can't remember your list of three attributes, Frankie, I'm not going to say it, you're here, but there are three things that they have these properties. What I'm saying is that those mental representations have exactly those properties. You can imagine a shape, you can operate on that shape. You can operate on it just the way that you can amend a drawing. It is literally a representation that you're operating on. It just happens to be in your head rather than on the page. It's literally, Edouard, a representation.

Edouard Machery
It's surely not a representation.

John Krakauer
It is.

Frances Egan
I agree that it's not a picture. I agree that it's not a picture in the head, but--

John Krakauer
It's not a picture. What I'm saying is, we are able to open the computer, the Word document in our head. We can look at the Word document in our head in the same way that we can look at it on the screen. That is odd that we have that ability that we can literally conjure something up, operate on it the way that we do on a drawing. Now we don't know how that works, but that's what I mean by representational behavior is that's our superpower.

Gaudi imagined the Sagrada Familia in his head before it ever laid the first stone. Now, how is it that that's possible that we can do that? That is what I'm saying.

Edouard Machery
I agree. This is a very interesting set of capacities. I agree that it has some interesting similarities with the use of external representation. I'm actually quite happy with that. I think that's really interesting. I'm not quite sure that everything you want to call representational has this feature. I think the imagination here might actually be somewhat an unusual case. Probably a lot of different things that don't quite have all these nice properties that you find in imagination where I can move things around in my head.

John Krakauer
Well, what I'm saying is, it's like people like Dan Dennett when I would have arguments with him, language got all the credit for why we have Zoom and books and arguments like this, and language was an infection in a regular primate brain that made it suddenly do all this work. I'm saying the representation is as powerful a concept. I agree with Bill Ramsey as language is. In other words, we really do have this ability to represent abstractions and operate on them.

I think that that is why I defend it, because it seems to get less credit as a concept that we have that language does. Language is something we use as an excuse for our abilities far more than our unique representational ability. I'm just trying to address that balance. It doesn't mean that there aren't unconscious ones. Absolutely, that's a fascinating topic. I'm just saying that representation, because it's become just neural correlates and information, it's lost that weird thing that you've just acknowledged, which I can see disappear in patients with language remaining completely intact.

Edouard Machery
I think, Luis, you wanted to add to John.

Luis Favela
Maybe an aspect of what Edouard was getting at was a question that I had which is I just wonder for people like John, is all mental life expressible and representational terms? Are there nonrepresentational features of our mental life going from the conscious to the non-conscious, to the cognitive to the non-cognitive, the imaginary, the non-imaginary, the vertical, the non-vertical? Are all of these to be described in representational terms? Frankie just put her hand up. I'm happy to hear what you have to say, and then I can finish up later.

Paul Middlebrooks
Let me just say we're coming up on two hours, and so maybe in a minute we'll do a wrap-up. Rosa, this is probably way too much for a new mother to bear for so long here, so we'll ask you first, but sorry, Frankie, go ahead.

Frances Egan
Just a really quick point. I don't want to leave the impression that I think that all of these capacities and abilities, what we might think of as the personal level, our linguistic. The point rather was that we understand them. In theorizing about them, we model them as linguistic processes, but it's a different question. It's an important question. What properties of mental states are actually getting modeled as linguistic properties. That's a big issue.

Paul Middlebrooks
John, you were going to respond, I think, to this also.

John Krakauer
No. I was going to say that I think you are right, and Edouard made the point too, that if we're going to talk about external representations of our model for internal ones, whether they're words, drawings, pictures, algebra, numbers to the degree that a lot of our thinking operates on numbers, symbols, pictures, sentences, I think language of thought notion of representation is a useful universal way of thinking about it.

Yes, in a way, to the degree that there are external clues to the representational work that we do in our heads, whether it's pictures, math, or poems, I would like to think that the generative process that led to those external representations were themselves representational. I don't know whether there are things off the top of my head where they're never going to have an external version of themselves, whether those should be called representational or not. I don't want to be a completist about it, but I think to the degree that many thoughts can be expressed in different types of external representations, they have their analogs inside the head.

Paul Middlebrooks
Does anyone want to comment on that before I demand that we-- Okay. Rosa, I'm going to put you on the spot first here. Closing thoughts in general, but did we make any progress here? Did any of us move our opinions? Did we learn anything? Are we going to agree to all disagree? What are your thoughts? That's a lot to ask of you. You can pick and choose among any of those things. Some closing thoughts.

Rosa Cao
I feel like I got more clear on what John's view was. I think I agree with what Luis said, that it sounded outrageous and authoritarian at first, but actually, it's quite reasonable once he articulates it in more detail. I guess what I took away from this is I'm still quite resistant to the idea of trying to legislate how people use the term representation, even though there are these knock-on consequences of equivocation and confusion.

It just seems to me that there's such a diversity of enterprises where people use the term representation that I think it really makes sense to try to get clear on how it's used in each of those domains and to try not to equivocate between them, to not offer a promissory note for a future explanation as an explanation that we already have. That seems bad. I agree with John there, but really trying to get clear on how it's used in different domains is really useful.

I guess I'm just going to use this opportunity to advertise. I was part of this generative adversarial collaboration at CCNA a few years ago, where a bunch of neuroscientists, computer scientists, cognitive scientists, and philosophers got together and tried to come up with a menagerie of different ways that the concept is used, and that collaboration is finally publishing its thing. We're trying to give a taxonomy of different uses of representation, like why they're useful in different ways. The one thing that they all have in common is that they have to be used and they have to be usable.

I think that relates to the thing that John was saying about the interesting processes being upstream of the stuff that you find in neuroscience, because what makes something used or usable is that it's part of this larger system that we can think of as doing these interesting things, these representation-hungry activities. It's only in that context that it makes sense to talk about representation.

Paul Middlebrooks
Frankie, I might ask you to go next because that sounds very pragmatic. I don't know if you have anything to add to that and/or closing thoughts. Have you changed your mind about anything? Did we make any progress, et cetera?

Frances Egan
I don't really have a lot to add to what Rosa said. I agree pretty much with everything that she said. I'm looking forward to seeing the results of the work she mentions. I've learned a lot, and I'm interested in just continuing to find out more about what's going on in the empirical sciences. Thanks for this opportunity, it's been a lot of fun.

Paul Middlebrooks
Well, thank you for being here. John, did you learn anything today except the correct animal on Frankie's ranch or farm?

John Krakauer
Oh, I did know it. I'm never going to be forgiven, and I'll never get the invite that I was waiting for now. What I learned is actually that my concerns

are somewhat valid, that I think that there's a usefulness to the notion of mental representations, which can be got at with effective theories of cognitive science. My friend Chaz Firestone always says, "Don't call me a cognitive neuroscientist, I'm a cognitive scientist."

You can do cognitive science without referring to neurons because you can have effective theories of representational behavior. I certainly have done work on that, but I don't want anyone to think I'm anti-neuroscience. I love neuroscience, and I think that neural evidence, mainly confirmatory, will have a lot to say about mental representational behavior.

My only concern is a misunderstanding when it comes to manifolds, connectionism, single neurons, where somehow, by using the representation word, on those neural data that somehow, you're much closer than you actually are to a true neuroscience of mental representation. That's why I feel like, in fact, the use of the word-- which I would never police- in fact, we wrote an article for *The Translator* on this. I never would suggest policing it. I'm just saying that I think I'm correct that it leads to deep conceptual confusions about how we're going to, in the end, properly link neuroscience and psychology.

Paul Middlebrooks
Edouard, that's now surprising to hear John say he thinks he's correct, wasn't it? [laughs]

Edouard Machery
No, I saw that's right. Something that I said earlier and I want to highlight again is how much agreement there is among us about some of the crucial issues, right? We should not be naïve about neural representation. We should just not assume this as very trivial link between the neuroscience or what I call neural representations and psychology. I think there's a great consensus here, which I don't think is reflected in the whole of philosophy, and I suspect also in the whole of science. I think that's really quite remarkable here and I think that's worth highlighting. We all agree about that much.

I think the place where we might disagree, but I'm not entirely sure, is about the value of having imprecise and loose concepts. I think there's some of us who- and I think John might be one of them- Rosa might be on the other side, I'm not sure about Frankie- I think loose concepts have a fundamental place in science. I think they allow knowledge, information to circulate in a somewhat unregulated manner. It can breed mistakes, confusion, but in the grand scheme of things, this is actually the way science tends to make progress, by ideas jumping from one area to the others by means of equivocation. I think that's extremely important for the good march of science.

I tend to be here a bit more for yes, let's 1,000 flower bloom, less confusion, do its nice trick, and that actually might be for the greater good. I think that with a ton of thesis in science where it has not been. I think that's a place where there might be disagreement about this normative take on confusion and ambiguity in science.

Paul Middlebrooks
I could be wrong, but that's a surprising thing for me to hear from a philosopher. It tends to be the other way, right?

Edouard Machery
That's right. I've changed my mind, actually, over the last five to six years on that very matter. I used to think restricting and regimenting was actually what philosophers should be doing, but I've become actually convinced that it's actually not the case. That actually, there's a great value for semantic drift and actually worth using somewhat vague manners, I think that's both in everyday speak and in science, too.

Paul Middlebrooks
Luis, I hope you're happy with yourself, bringing all of us together here since this was your inspiration, your doing. Closing thoughts from you.

Luis Favela
Thanks again, Paul. I'm an emotional vampire, so when I see people get worked up, it just really feeds to see people get upset. I feel like I've been buffeting today, so thank you for this opportunity. I'm going to go an opposite view of Edouard since we're a team on this. He can be good cop with the Kumbaya and everyone. We can let a thousand flowers bloom. I just want to let the good flowers bloom. I think some of them might actually be weeds that we're thinking are flowers. I want to weed out the weeds and really facilitate making room for the flowers.

Edouard and my, our project, we're the embodied people who did it, but I think there has been want and concern about trying to get some systematic evidence for how these terms are used. This conversation has further illustrated how people working on these topics continually talk past each other. There's a lot of work that needs to be done with saying, "What do you mean by your terms in this case?" That doesn't mean we have to police all universal uses of the term, but when we're trying to have a debate, getting clear on the terms is essential, and I further learned that.

I say we go scorched earth and get rid of representations, maybe just talk about other things. I'd like to plug Rosa's paper that was a reply to Edouard's and my paper, where she lays out. It's a very nice list of different kinds of uses of the term representation. As I said to Rosa in the past, "Fine, representation can mean this. Representation can mean that. Why not just say this or that?" That's my view and take from it, but yes, just happy to spend this time chatting with smart people about this.

Paul Middlebrooks
All right, guys, we did it. Thank you so much for joining me. This was easier than I thought it would be, so that was nice for me. Rosa, you hung in there, so thanks for hanging in there for so long. I know you've got some other things going on right now.

John Krakauer
Congrats, Rosa. That's awesome, by the way.

Rosa Cao
Thanks.

Paul Middlebrooks
Okay, take care, everyone. Thanks again.

Edouard Machery
Thanks, Paul, for having us. It was actually really, really nice [crosstalk]

Frances Egan
Thank you.

John Krakauer
[crosstalk] see you in the pub, Frankie, and you, Luis.

Rosa Cao
Yes.

Luis Favela
Yes, we should do that.

Eduoard: We should get lunch, Paul, after, we're neighbors.

Luis Favela
[chuckles] No.

Paul Middlebrooks
Sounds great. I want to hear who's been talking about me behind my back.

[laughter]

Eduoard: Only nice things. All right, take care, bye.

Rosa Cao
Bye.

[music]

Paul Middlebrooks
"Brain Inspired" is powered by *The Transmitter*, an online publication that aims to deliver useful information, insights, and tools to build bridges across neuroscience and advance research. Visit thetransmitter.org to explore the latest neuroscience news and perspectives written by journalists and scientists. If you value "Brain Inspired," support it through Patreon to access full-length episodes, join our Discord community, and even influence who I invite to the podcast, go to braininspired.co to learn more. The music you hear is a little slow, jazzy blues performed by my friend, Kyle Donovan. Thank you for your support. See you next time.

[music]
