

## DO MY SIGNALS DECEIVE ME?: An Interview with Jorge Morales, PhD



### ABOUT JORGE MORALES, PhD



Jorge Morales is an assistant professor in philosophy and psychology at Northeastern University. He also directs the Subjectivity Lab, where he and his team look for theoretical and empirical answers that help us better understand the mind and the functions of consciousness. Born and raised in Mexico City, he completed his undergraduate and master's degrees in philosophy there before coming to the United States. In 2018, he earned a PhD in philosophy from Columbia University, after which he was a postdoctoral researcher at Johns Hopkins University. Morales' research occupies a unique niche at the intersection of philosophy, psychology and neuroscience. It focuses on understanding visual experiences, how the brain creates them, and how we introspect our own mind. His scholarly work has appeared in top philosophy and scientific journals, and his findings have garnered attention from both the national and international press. In 2023, he won a Rising Star Award from the Association for Psychological Science for his contributions to the discipline.



STANCE: Thanks for talking with us. We really enjoyed reading your work. We found it interesting that you work in both philosophy and psychology. How did you end up with that kind of job?

MORALES: In the last few years, I started doing more empirically based research, where I was not just reading about empirical findings but also trying to produce them. My job is mostly based in the psychology department. I guess I'm still fifty-fifty, but I flipped from full philosophy to a little bit more psychology than philosophy. I'm just very happy to be talking with you. Thank you so much for inviting me.

STANCE: Oh, thank you so much for coming! We're particularly interested in what got you into studying philosophy. What in your educational journey inspired you not only to study it as an undergraduate but also to go on to graduate school to study philosophy of mind and philosophy of perception?

MORALES: I started being interested in philosophy in high school, actually. I was in Mexico City, where I was born and raised, and I was lucky to take a class about the history of culture and philosophy. It included everything from the Pre-Socratic era and St. Thomas Aquinas to Islam, the Industrial Revolution, and philosophy in the twentieth century from Adorno to Wittgenstein. It was a very comprehensive class, and it really made an impression on me. I decided to pick philosophy as my field of study in college and eventually I focused on consciousness.

Grad school was always on my mind, almost since the beginning of college. I had a lot of luck. I worked together with a few professors who were really passionate about research and that helped me get a closer look at what a career in academia might look like before deciding to apply to a master's program. After a couple of years of

IT WAS  
THERE  
WHERE I  
HAD MY  
FIRST REAL  
EXPOSURE  
TO SCIENCE

doing that at Mexico's National University, I had the opportunity to spend a semester at Indiana University working with Colin Allen on a thesis on animal minds and theory of mind. That's where it became very clear to me that I really wanted to keep going and get a PhD. It was there where I had my first real exposure to science as well. Colin Allen, even though he's a philosopher, worked really closely with psychologists. Then I started reading empirical papers about animal minds. Eventually, I ended up going to Columbia

University to get my PhD, and I thought I was going to write my dissertation on animal minds. But I had a chance to take a class about

consciousness and attention with Ned Block at NYU. At the same time, I volunteered at the Hakwan Lau laboratory in the psychology department at Columbia. These two big influences really shifted what my PhD was eventually going to look like. Basically, I went back to my interests from undergrad and ended up focusing on perceptual awareness and introspection.

STANCE: Awesome. So that relates really well to our second question. Throughout that journey, what got you to move more to the empirical side of research? Was there a particular moment or a study that really inspired you to say, "Okay, I want to contribute to this body of empirical research"?

MORALES: I think I acquired the first seeds of interest in empirical work when I was an undergrad. And it was through the history of philosophy and popular science books, actually. I went to a small liberal arts college in Mexico City, and they were very focused on ancient Greek Philosophy, medieval Christian philosophy, German idealism, and especially Kant. So, I took several classes in Latin, Greek, and German, thinking that this is what I wanted to do because that's what the people around me were doing. But it was during my sophomore year that I just happened to check out a couple books from the library—one being by anthropologists Niles Eldridge and Ian Tattersall. That book was called *The Myths of Human Evolution* and one of the things that it talked about was the importance of cranium size changes across hominin evolution and how that allowed newer species to develop a more cognitively sophisticated apparatus. The other book that I checked out was by the neurologist Oliver Sacks—a pretty famous guy, some of you probably know him. It was called *The Man Who Mistook His Wife for a Hat*, which again made a huge impression on me about how a little deficit, a little injury, a little disease in the brain can massively change how the human mind operates. And by total coincidence, I was reading *De Anima* by Aristotle at the time, and I don't know, something just clicked by having these very disparate set of readings—it made me think, "Look, all this stuff that Aristotle was saying is so important and it is so interesting to try and understand how the human mind works, but it cannot be done without a thorough understanding of the brain and psychology."

I think that just planted a seed in me. I wrote my undergrad honors thesis on consciousness, including something of an empirical approach. I abandoned Greek, Latin, and German and started reading analytic philosophers from the twentieth century. But that really set me on a path of caring about science, which was eventually



heavily triggered by Colin Allen and looking at actual scientific papers on that topic. And I guess what sealed the deal was when I started my PhD. I knocked on Hakwan Lau's door, a neuroscientist at Columbia, and I just asked him, "Hey, I want to write my dissertation on these topics that you work on from the empirical side, but I feel a little bit silly just reading philosophy. I think that I should know more about the brain and psychology." I was kind of naïve at the time, so I asked which books or which articles I should read to get more informed. He responded, "Hey, it doesn't work like that in science. If you really want to understand these papers and not just glance over the abstract, you have to do the work. Why don't you come work with me?" And you know, one thing led to the other. I started as a research assistant and then I guess I was never able to stop. For my postdoc, I went to Johns Hopkins University to work with Chaz Firestone—a fantastic psychologist with deep interests in philosophy—with whom I continued to weave these threads of science and philosophy. So now I do the empirical work as well. That experience really made me value the importance of doing both types of research at the same time.

*STANCE:* So, when it comes to both at the same time, we've heard a bit about the role of the research in the scientific studies. What if a scientist, a psychologist, or a neurologist questioned the value of philosophic foundations or a philosophical approach to scientific discovery? What if they say, "I don't need it. I've gotten by just fine doing research from my perspective." Would you argue that they're missing something?

THERE'S  
A HUGE  
NUMBER OF  
SCIENTISTS  
THAT VALUE  
AND CARE  
ABOUT  
PHILOSOPHY

*MORALES:* I love this question because I actually get it a lot from real psychologists and neuroscientists. I think there's a huge number of scientists that value and care about philosophy, but there's still a good number that just don't get it, or even worse they think it's a waste of time or even pernicious for science. And my first thought is that philosophical reflection is pretty much inescapable, right? As soon as you're asking the type of questions that scientists ask, if they want to actually argue as opposed to just stating that philosophy is useless, they actually

have to provide philosophical argument for why. Maybe there is a demarcation issue where science and philosophy should be separate, or one is useless for the other. But in virtue of making that argument, they are actually doing philosophy.

But maybe more importantly, both are theoretical reflections and in the empirical pursuits that they follow, they actually make

philosophical assumptions. In fact, they offer philosophical arguments even if they don't describe them as such. Not every single sentence in an empirical paper is based on empirical evidence, right? There is theory, there is reasoning, there are metaphysical assumptions of how the mind works, and so on. Sometimes if researchers reflect on those assumptions with philosophers, the assumptions can be improved a little bit.

THERE IS  
THEORY, THERE  
IS REASONING,  
THERE ARE  
METAPHYSICAL  
ASSUMPTIONS  
OF HOW THE  
MIND WORKS

Just to give you an example, Philosopher Ned Block, who I mentioned before, introduced the distinction in consciousness research between phenomenal and access-consciousness. Access-consciousness has to do with availability of information, with rational use of that information, whereas phenomenal consciousness is what it's like to be in a particular mental state, its qualitative character. This distinction was kind of there in the literature; not to take it away from Ned Block, but he did not come to this conclusion in a complete vacuum. Scientists and philosophers were using it before, but he, with careful philosophical reflection, introduced this distinction, making it very useful for philosophers and scientists. These days this distinction has become something that guides empirical research as well. So doing philosophy well is very helpful. Knowing about what philosophers think about is very helpful. Philosophers do a lot of distinctions that help not only consciousness, which happens to be my field, but also geologists, neuroscientists, physicists, and so on. So yeah, I don't think that scientists can really be fine without philosophy, even if they don't accept it.

*STANCE:* Alright, well, moving on a little bit to your work. We're interested in this concept of mental strength. Could give us a brief explanation of mental strength, and in particular why you think it's a domain-general property of experience?

*MORALES:* Oh, great. So, mental strength is the intensity of a conscious experience. Pains, for example, can be more or less strong. Mental images can be more or less vivid. Perceptions can be more or less striking, emotion can be more or less intense, and so on. All these variations, according to my view, are variations in the degree of how strong these conscious experiences are felt. They can all be traced down to a property that I call mental strength, which is just a way of describing a phenomenal magnitude: how much of a phenomenal character and the degree of phenomenal intensity that experiences have. And when I was thinking about these issues,



one thing that caught my attention is how little recognition of this property you find in philosophy and in cognitive science these days. It's starting to change, but it's still kind of a secret that people don't talk about.

Philosophers from the past recognized it and psychologists like William James thought it was very important, but it has flown under the radar in philosophy—at least until recently. I think this is partially because philosophers get anxious to think about degrees of consciousness. They often think that it is an all-or-nothing phenomenon and once you start talking about degrees things will get metaphysically murkier, which might just as well be true. But I think that we can't deny that we have different experiences with different degrees of intensity. I take this evidence as a point of departure to build on the theory. I think that these variations in intensity have to be attributed to an intrinsic property of experiences. It's how much it's felt as opposed to what they are about or what you're representing.

A popular, alternative view in philosophy would be that these changes in intensity are really just a consequence of what the mental state is about. On this view, if I imagine something very vividly, that's because the representational contents of the experience are sharp, not because the experience itself is sharp. The experience itself is not necessarily something that some philosophers think that you have access to; they think that we experience the world through our representations, but the representations are transparent.

But I think that this is wrong. I think that we experience at least some of these phenomenal properties of experiences, like their intensity, which can eventually be determined as distinct from their representational content. And I think mental strength is domain-general. It's either domain-general or domain-specific. So, on the domain-specific view, pains have one type of intensity, mental images have another type of intensity, perceptions have another type of intensity. But on the domain-general view, intensity is the same property shared across different types. And I think the domain-specific view has a hard time explaining “blinding” experiences, or those where a reduction of intensity in one experience takes place when another stronger experience interferes with it.

So, if you're in *excruciating* pain—I hope none of you have experienced it—but if you have, it's really hard to experience other things, right? It's almost as if your whole existence becomes that pain and what

you hear, what you see, and what you feel in uninjured parts of your body kind of cease to exist. You're not aware of them. It's as if the pain is blinding or blocking other types of experience. If you have a vivid imagination and you're daydreaming, it can get so vivid that you perceptually decouple from the world. You almost stop experiencing what is in front of you. Maybe you even start acting out that imaginary conversation you're having with your partner or whoever it may be.

I think that these kinds of interferences are nicely explained by the fact that mental strength is a limited resource, when it gets spent on one type of property, other states have less of it. You just can't experience everything at the top of its possible degree of intensity. I think that this just speaks in favor of the domain-general view.

*STANCE:* This quality of one strong experience being able to blind another, this is the kind of relational aspect of mental strength that you bring up, correct?

*MORALES:* Right, yeah.

*STANCE:* As we read your explanation of mental strength, particularly in your dissertation, you describe how the mind kind of self-structures in the sense that it has it a series of priorities where it prioritizes mentally stronger experiences and deprioritizes mentally weaker experiences. Could you explain that further?

*MORALES:* First of all, I think that you're probably like the sixth person that read my dissertation besides my advisor, which is fantastic [*laughs*]. But, yes, exactly! That's exactly right. I think one very nice consequence of the mental strength theory is that it really portrays our minds, our conscious minds in particular, as being active and self-structuring. I think that other views might have a little bit more trouble getting this picture, and they require the subject to be way more active than I think we actually are in building what our experiences are like.

I think it's true that when you pay attention to different things, you can voluntarily control your attention and so to some extent, we are definitely responsible for contributing to the structuring of our minds. Think about this interview, right? Maybe my voice is kind of prominent in your conscious experience at the moment as opposed to the experience of feeling the chair against your back or the light in the room. Before I mentioned them and your attention was captured by them, they weren't prominent in your conscious field, right? It was just the strength of the external stimulation. Yes, what structures the mind is a little bit of what you're paying



attention to, but I think that attentional effect is secondary. It is one mechanism that we have for structuring the mind, but it is really the mental strength that makes mental states fall into place. And, you know, it's nice when you find that other, smarter people kind of agree with you. William James held a similar view. Like with many other things, I think he got it right when, while he was discussing attention, he said that there is no need for attention to drag ideas before consciousness when we see how perfectly they can drag themselves there.

We don't need to make a voluntary effort to organize our minds. Our minds are already organized. That organization happens at the conscious level, according to me, thanks to mental strength. And that explains why there are ranks, such as why we experience a background and foreground. Everything is set on a foreground or a background, right? The Gestalt psychologists got that right too; it's a necessary trait of our minds. It would be very weird if every single conscious experience that we have is at the exact same level, making it look kind of indistinguishable from everything else. It's very malleable levels that structure how we experience the world.

*STANCE:* We were interested in taking this self-structuring into an ethical dimension. We're curious if the kind of automatic self-structuring that you describe, the secondary nature of attention, threatens or provides some sort of argument against libertarian accounts of free will. Does the self-structuring that you're talking about prevent agent causation, or something like that, or does it make it more difficult to defend?

*MORALES:* Yeah, that's a really interesting question that I confess I haven't given enough thought to before. I usually try to avoid ethics because I just think it's very complicated, so I study consciousness instead [laughs]. Although my gut reaction is to think that the self-structuring by the intensity of our experiences does shape the space of possibilities that we can freely act upon. You know, out of sight out of mind. If something is not a part of your consciousness—or if it is, it's not particularly intense—it limits what is available as an actual possibility for you to act upon. It's not just that we can't act, freely or not, upon what's not in your mind, but our decisions are also going to be impacted by what's at the center of the totality of our moment-to-moment experiences. So, I don't know if it makes us not free, but it limits the range of mental states that we can act upon or base upon.

That said, I think that this primacy of salient mental states should be defeasible, right? I think that in many cases we can choose to ignore



our more intense experiences in favor of the less intense ones. It might be harder to act upon the weaker ones, but it doesn't mean that we can't figure how to ignore strong feelings such as hunger in order to carry on with whatever we are doing because we deem it to be more important. But, of course, there's a limit, right? For example, there is extreme pain, hunger, passion, euphoria, sadness—any emotion can become very intense. It's easy to lose ourselves a little bit, or a lot, and struggle to be in control of our actions. I think this is a good result because we know that when people are put to extremes, it's hard to attribute freedom to them. There is a book by Primo Levi, a survivor from Auschwitz, titled *If This Is a Man*. The premise of the book is if someone's body and will have been broken to the extent that prisoners of Auschwitz were, is that a man? One of the questions explored by the book is "Is there free will?" Levi also thought that there was no morality anymore. For example, the stealing of a spoon from another prisoner in Auschwitz might not be immoral because it's out of the realm of morality once the human condition has been broken to that extent. And I think that it's a terrible consequence, but it might be in line with when certain mental states are so extreme, we lose some kind of control and maybe even responsibility. So yeah, I think we can use mental strength to explain, at least a little bit, what the margins are when it comes to operating under freewill.

*STANCE:* I think you had a great point there, that a blinding effect can occur with mentally stronger experiences. Can we connect this blinding effect to the popular intuition that crimes of passion are less blameworthy? I mean, even the law is structured so that premeditated murders generally receive higher sentencing. So, do you think that the intuition that an absence or reduction of control accounted for in the law could be explained with the blinding effect of mentally stronger experiences?

*MORALES:* Yeah, I love this question. I was only partially joking when I said that I tend to not think about ethical questions in my professional work often. But yeah, this is a good case for thinking about it. Because being consistent with the self-structuring thesis, yes, you are in control of what you expose yourself to. Just as you can avoid looking at the sun to avoid blinding yourself, you can try to avoid having extreme experiences to avoid making them too dominating, right? But sometimes we fail at that, or

**MAYBE WE  
ARE FREE TO  
CONTROL  
OURSELVES  
AND NOT ACT  
UPON AN  
OVERPOWERING  
EXPERIENCE,  
BUT WHAT IF  
LITERALLY THE  
ONLY THING  
THAT YOU'RE  
EXPERIENCING IS  
ANGER?**



life just makes you have these very intense experiences. So, I think that you're right, we do tend to excuse, both morally and legally, actions that stem from extremely intense dominating experiences: maybe bursts of rage; excruciating pain, either physical or mental; extreme sadness; burning lust; blinding desire of vengeance; maybe even intense happiness, as people do silly things when they're extremely happy. There might be fairly well-studied biological and neurological reasons why we can experience an atrophy of reason, but I think at the conscious level, when we talk about the phenomenology, the blinding experience is useful too because an overpowering experience is something that we can't really help. We can't change too much moment-to-moment. Maybe we are free to control ourselves and not act upon an overpowering experience, but what if literally the only thing that you're experiencing is anger? Think about Iago's jealousy in *Othello* and how it made him act in completely irrational, even self-damaging ways. I think it's not just that the jealousy is very powerful and becomes the only thing in front of you, but it can be so overpowering that it dampens other experiences—it prevents you from having other normal emotions or to be able to focus on anything else.

Sometimes in our lab we ask people to focus their gaze on a small dot at the center of a screen. Some people find it very difficult because there is just this gray screen in front of them with this tiny dot and when everything is so homogeneous it's very hard to keep your eyes fixed because we're used to moving them around and shifting our attention. And I think that an analogous problem happens with the mind and with conscious experiences. If you're so dominated at one particular moment, maybe it's very, very hard to look away. You don't have anywhere to move to. I would say that this is a theory of moment-to-moment changes, like mental strength. So, of course, this wouldn't apply to long-term planning or anything like that, but I think that there is some room for mental strength and ethical reflection to take place in the actions that affect you in the moment. I should think more about this.

*STANCE*: The way that you describe it made me immediately think of full-body relaxation techniques, where the trick is to imagine a particular part of your body and to tighten it and relax it. The idea is, if you're having an overwhelming blinding experience lying there, to focus on points to bring those experiences up and reduce the blinding effect of that anger or the anxiety you're feeling.

*MORALES*: Yeah, absolutely! I think you're right. That's exactly the effect. And it has been shown that if you focus on your pain, you



can bring that sensation up; and if you shift your focus away from the pain, you reduce how intense it feels even though the stimulation is similar. So, in terms of moral actions, that has to play a role, right? How many options did the defendant have? To understand the *mens rea* and the desire of causing damage, we must ask "To what extent did this person have options?"

**STRONGER EXPERIENCES ARE EASIER TO INTROSPECT AND HARDER TO MAKE MISTAKES ABOUT**

We usually think of having options and freedom to be crucial for assigning moral responsibility, but if you're blinded by an emotion, maybe that's why we are a little bit more lenient with people who didn't have options mentally speaking.

*STANCE*: Related to this topic of moral responsibility, we're interested in emotional experiences. While coming up with questions, I was thinking about the times I have had incredibly mentally strong emotions. I was looking back and thinking: "How does the strength of these emotional experiences affect my ability to understand them and to understand how my emotions are affecting me?" Do you think there is any relationship between this kind of introspective success, not just locating that the emotion is there, but understanding how it's affecting you? Or do you think that this is unrelated to the mental strength of an emotion?

*MORALES*: Yeah, yeah. Again, a super interesting question because I do link mental strength with the accuracy of introspection. In general, I think that stronger experiences are easier to introspect and harder to make mistakes about, whereas a very weak experience might be harder to decipher, and we may lose some of its detail. But your question raises an interesting possibility, which is that there is a limit to how strong an experience may be, right? Maybe it's harder to miss the presence of an extremely strong experience, but then, it's so blinding that it's hard to make out what it's about, almost like looking at the sun. You won't miss the fact that there's a bright source of light in front of you, but you won't be able to make out any of the details of the sun because it's so powerful that you can't actually see it properly.

In philosophy there have been huge debates about introspection and whether we can actually introspect mental states. Emotions are important in that sense, but also other kinds of mental states. The issue is that it seems like if you introspect, then you are—in virtue of introspecting—affecting that mental state. It's very hard to know what an un-introspected mental state feels like, right? When we attempt to introspect it's always a little bit in retrospect.



That is very different from what it feels like when you're not paying attention. I don't know if that means that we wouldn't understand those emotions or those mental states in general, but I think that it's definitely a problem that we have with introspection, that in virtue of using it, we transform the state a little bit. Which, on one hand, is a problem.

On the other hand, it's not that different from how other things work, like perception. You might think that looking at something doesn't change the thing that you're looking at. It's true that the object doesn't change, but it's not true that the way your brain processes that object doesn't change, whether you pay attention to it or not, right? It's hard to know what something looks like when you're not paying attention to it. So, in that sense, I think introspection both affects the state that you're introspecting, but it also helps to boost the signal that you're trying to get. So, as long as it doesn't get too strong, at least on average, you should get a better sense of what your mental state and emotions look like. This is why meditation tends to help people.

*STANCE:* Kind of related to this topic of emotions and our ability to reckon with them, we're interested in how your concept of mental strength might relate to, or explain, another concept in psychology: stereotype threat. The concept of stereotype threat relies on Cognitive Load Theory, which is similar to, but has important differences with, your theory of mental strength. So, we were wondering if mental strength might help to explain stereotype threat, or if it offers new challenges or problems?

*MORALES:* On an intuitive level, I would say, of course. Distraction, for whatever reason, might take away important processing resources that affect our performance in a task. Now, with that said, I don't think this is a case where mental strength can help that much, and I think that is because the theory is cached out at a phenomenal level. An experience that is intense might prevent feeling other things as intensely, as we've been discussing. But cognitive processing, reasoning, doing math, and other things that I think the stereotype threat literature has focused on, are centered on cognitive abilities, and not on experiences or their qualitative character. So even though cognitive load and mental strength may have similar implications for how they limit other cognitive capacities and other experiences, respectively, I think that they operate in parallel and in different levels.

Of course, if someone is having a strong experience of anxiety while taking a test, it might be hard for them to do well. But, stereotype



threat, as I understand it, goes beyond just feelings of anxiety, right? It's supposed to explain how cognitive resources are taking away from the task because of what effectively is multitasking. In this case, by trying to spend energy thinking about and trying to avoid fulfilling the stereotype, you fulfill it by underperforming. So, I think there are different levels of explanation. I would be remiss if I didn't mention that, as interesting as the hypothesis is, there have been lots of problems replicating stereotype threat's effects in the last few years. But maybe there is something about stereotype threat: when you become aware of your position in a social hierarchy, it's distracting, and that makes you underperform; but, given all the problems of replicating the effect that we've seen, I think we should be careful. Of course, caution doesn't take away from the everyday struggles that minorities and people who have been discriminated against experience, and the extra cognitive load that they have to put into doing tasks. It just seems that the reasons for underperformance, although not always, are more complicated than the simple activation of the stereotype. There's more research to be done.

*STANCE:* Thank you for that, I wasn't familiar with the replication issues surrounding stereotype threat. This other component of your work that we're interested in talking about is how you use a Signal Detection Theory model to explain introspection and introspective success. Could explain how you translate the model from perception research into introspection?

*MORALES:* Yeah, I can say a little about the origins of both mental strength and introspective Signal Detection Theory. I was very impressed by Signal Detection Theory when I was studying it in the lab. I thought this idea that there's noise everywhere—that the jobs of perceptual detectors are to separate signal from noise—was so powerful that I started to find parallels. In the history of science, it's called model migration. We use what we understand well to try to understand what we don't get very well yet. This happens in development when we're children. We try to teach children easy things so they can start understanding harder things,

and it happens in science too. Think about Michael Faraday, who used the physics of mechanical fluids to explain electromagnetism,

THIS IDEA THAT THERE'S NOISE EVERYWHERE— THAT THE JOBS OF PERCEPTUAL DETECTORS ARE TO SEPARATE SIGNAL FROM NOISE—WAS SO POWERFUL THAT I STARTED TO FIND PARALLELS.



a new phenomenon. And Signal Detection Theory, as it is used in perception, is already a migrated model. The math was developed for assessing radar performance—how to know if a radar is actually detecting targets, missing them, or just false alarming to non-targets. It's a huge problem, so scientists and engineers in the United States created Signal Detection Theory to solve this issue, introducing two main concepts that we use in perception today: sensitivity and response bias—how good you are at detecting something, and how conservative or liberal you are at calling it the thing you're detecting? One is how good your visual system is, and the other is how willing you are to say that you saw something. This is just such a powerful model that I thought that those formalisms could be used in introspection as well, and they can guide us in using something that is very well understood in perception to understand something that is harder to study, which is introspection of conscious experiences.

Just like in perception or radar technology, the mental states are out there in your mind, and introspection's task is to detect them. Is there hunger or not? Is this a craving for pizza or burgers? Am I experiencing crimson or scarlet? It's just hard to study introspection because our experiences are inside our heads, and experimenters cannot look at them. So, the idea is that introspection must operate as cognitive faculties that we know, and there are many of them. Many work as signal detection theory, many use signal and noise separation to perform their tasks. If we can use that, maybe we can understand introspection better. The idea is that mental strength plays the role of stimulus strength, and that role is modulating how likely it is to have an accurate introspection, or how seeing with good light is more likely to yield an accurate perception. Not always, but it's more likely. When the light is off, you are less likely to find what you are looking for. You could find it still—you can still see in the dark, just not very well. So, introspecting a strong relative state should also be something that is more likely to yield an accurate introspection, and weak mental experiences are just a little bit more likely to end up in error.

*STANCE:* One of the commitments of your theory is indeed the possibility for error, and we can introspect, experience, and detect something that is not there. We debated this idea, and some of us have the intuition that, for the most part, if a person thinks they are in pain, they are in pain, in the sense that felt pain is identical to pain. What would you say we are missing here?

*MORALES:* I get this question a lot. I totally get it—the idea that someone can be mistaken when reporting that they are experience

pain is counterintuitive. But I think this is a case where intuition can lead us astray. Granted, I think it's really hard to imagine being wrong about a strong pain. But I think that when we consider weak emotions, or weak experiences in general, our confidence should actually go down, and if our confidence about what we're experiencing goes down, we should be more open to error. Imagine that you go to the ER in extreme pain, and the nurse asks you if you're sure that you're in pain. You will probably be very mad and offended by the question and just demand to get a painkiller,

THE IDEA THAT  
SOMEONE CAN  
BE MISTAKEN  
WHEN REPORTING  
THAT THEY ARE  
EXPERIENCING  
PAIN IS  
COUNTERINTUITIVE

but after the painkiller starts reducing your pain, there is a point where it is actually harder to introspect if you're still in pain or not. "Is this pain or am I just numb?" Or "Is this just uncomfortable and I'm actually not in pain anymore?" Or "The pain is actually gone but it was there for so long that I still kind of feel weird, but it's not pain." Even if you think that you can't be wrong, I think these examples illustrate that you can't be equally confident in what you're experiencing in these two extreme cases. What I argue when thinking about Signal Detection Theory is that because there are these weak experiences that are harder to classify, it is possible that we might be wrong about them. It might be that we make a judgment that we're in pain, that we really believe that we're in pain, when there is actually no experience whatsoever of pain.

I think philosophers really don't think of introspection as just another cognitive capacity. They have provided introspection with the special status and infallible ability to get our mental states. But my view is that this is just another thing that our brain does and if it's just another thing that our brain does, why would it be infallible? This view is part of a family of theories called Inner Sense Theories, where the idea is that introspection works like perception—it's a detection mechanism and philosophers hate this. They literally have described inner sense as a repugnant theory. So, I guess I wear that as a badge of honor in the sense that if our brains are doing it, they can do it wrong. At least to me, introspection is like all other thinking, and I don't think our brains do magic. So, I just propose that introspection is another detection mechanism that is liable to error. Of course, one of the problems is that those

I JUST  
PROPOSE THAT  
INTROSPECTION  
IS ANOTHER  
DETECTION  
MECHANISM  
THAT IS LIABLE  
TO ERROR





errors are hard to detect. Like every time you say, “there is some pain,” there is no way of confirming that you’re right or wrong, and this lack of confirmation gives you some sort of authority over everyone else. But, at least in some cases, we should be a little bit more confident that this infallibility is very implausible, like in a case where the dentist hasn’t even touched you yet. You were so certain introspectively that what you experienced was pain, but maybe you just confused vibrations for pain—or something like that. So at least, theoretically, these errors should be possible.

STANCE: So, with that, do you think that it gets a little bit more complicated when the introspective error is long term? Would you consider experiencing placebo effects from a prescription medication a case of on-going introspective error?

MORALES: Well, I guess there could be two types of placebo effect. One is where only your judgments change, but your experience is the same. Another is that the *belief* that you’re being given medication has an actual causal effect on reducing the experience of pain, so your introspection is accurate. Your pain has been reduced, you’re just wrong about the cause of it. Feeling like you’re dealing with the problem or feeling less stressed now that you got the medicine both might help reducing pain in real ways. So, I guess it’s hard to know which one it is.

STANCE: Switching gears a little: we’re interested in your value for collaboration, especially between scientists and philosophers. How would you suggest that we foster this kind of collaboration, especially at the undergraduate level? You talked a bit about going to your professor and him inviting you to be a research assistant. Would you recommend that we, as philosophers, put ourselves out there in a similar way?

MORALES: I definitely do. The value of expanding your horizons is undeniable. I think that it is very important, even within philosophy, to read broadly—not just the ancients. Read contemporary philosophy too, or not just contemporary philosophy, read some Germans for a change. And I think that is good to expand your fields in general in life. This is why people travel and try new things. I think exposing yourself to scientific reasoning, to the scientific results, is really a way of broadening not just how you do philosophy, but what you can do philosophy about. I think it creates a very virtuous cycle.

How to foster this? It might depend on your interests and what’s available around you; but I would go from reading pop science books that introduce general audiences to a topic that you’re interested in to randomly attending science talks at your university. Even if you don’t get everything that is being said, that’s okay. I think if you’re



interested, you just circle around enough until you get it, until you start seeing problems, or learning the language, which is a huge part of doing interdisciplinary work. It’s not just learning results as a list of facts, which science is not. Science is not a list of facts, it’s a practice. It is an endeavor and in order to get it one of the very important things is to see how it is done. This could be in popular books or in an article. It will be hard to understand half of it, but next time you do it, it will be a little bit easier. Talk to friends or peers who are majoring in scientific fields. Just talk to people. Expose yourself to these things. I think it’s just so important to have the tools and a broader sense of what happens. Sometimes it might end up with you getting into an interdisciplinary research program and sometimes it won’t, but it will broaden your understanding of the world. And of course, scientists should do the same: they should take philosophy classes

SCIENCE IS NOT A LIST OF FACTS, IT’S A PRACTICE. IT IS AN ENDEAVOR AND IN ORDER TO GET IT ONE OF THE VERY IMPORTANT THINGS IS TO SEE HOW IT IS DONE.

STANCE: I think reading a lot of your work, especially the experimental work, gave us that experience. A lot of us came into it and when we started reading, it was very new for us. The experimental methodology, different theories about the neural correlates of consciousness, it was all very complicated for us. But I think you are right in that we collaboratively grew and understood a little bit more of the experimentation. We’re not quite there yet, but I think that we’ve at least started the kind of collaboration that you’re talking about. Maybe we need to get more scientists in here.

MORALES: That’s good to hear, it never works perfectly on the first try, but maybe some of you will be more interested and read a follow-up or another paper on that. It’s about being in the same room. Either physically in the same room or in the same head space as people working in another field. That’s what it takes, sharing concepts, language, and theories. Eventually you just get it.

STANCE: Alright, thank you so much for coming. It’s been a wonderful interview.

MORALES: No, thank you so much. I really, really appreciate it. I feel very lucky that you all read my work, thought carefully about this, and made wonderful questions, it’s great. Thank you so much.

STANCE: Thank you, too!

