JORGE MORALES

Research Statement

Overview

Cognitive science, like any scientific enterprise, aims to provide an *objective* understanding of its domain. Historically, this has often meant eschewing the more *subjective* aspects of the mind, including introspection, conscious awareness, and other private features of our mental lives. But in doing this, we may be leaving out a crucial and fascinating source of knowledge about ourselves. My research embraces the subjectivity of the mind and studies it directly. The goal of my research program is to **understand the mind's *subjective point of view***. In particular, my work is focused on understanding how subjectivity affects *what we perceive* (perception) and *what we take ourselves to be perceiving* (metacognition); and my work also aims to understand *subjectivity itself* (consciousness). My training is in **philosophy, psychology and neuroscience**, and it has given me the tools required to build an independent research program that explores and answers foundational questions. By taking subjectivity seriously and tackling it in a **fully interdisciplinary** way, my work opens an avenue to make progress in a problem often considered intractable.

1. Subjective Influences on Perception

Our own subjectivity can affect what we think, what we remember, and—importantly—what we perceive. As a case study, I focus on the subjective aspects of our mental life that affect **perceptual experiences and perceptual decision-making**. It is often argued that the visual system is optimized to represent the objective world in a way that overrides our own subjective idiosyncrasies—let these be perceptual (e.g. the perspective from which we perceive an object), cognitive (e.g. the priors we have), or social (e.g. the stereotypes we hold). **My work challenges these assumptions about the objectivity and optimality of perception** and offers a way of studying subjectivity in the laboratory.

Vision Is Perspectival

Look at these coins? Does the circular, rotated coin on the right look to you like the oval coin on the left? This question has generated heated debate from Locke, Hume and Helmholtz to current research in philosophy and cognitive science. Arguably the most foundational principle in perception research is that our visual experience of the world goes beyond the retinal image; we perceive the distal environment itself rather than the proximal stimulation it causes. When a circular coin is rotated in depth, for example, we experience it as the circular object it truly is, rather than as the perspectival ellipse it projects on the retina. Despite the ubiquity of shape constancy, what is our experience like when an object's shape does not match its perspectival shape? Does the rotated coin look like an elliptical object too? Whereas extant arguments rely primarily on introspection, I have provided the first empirical test of this core philosophical question. If the mind continues to represent the ellipticity of a coin, objectively circular objects should, when rotated, impair visual search for objectively elliptical objects. I demonstrated that this is so using highly realistic computer-generated images presented to hundreds of subjects online, as well as laser-cut "coins" presented to subjects in real life under completely normal conditions (**Morales, Bax & Firestone, *PNAS*, 2020**). These results show that **perspectival shape representations persist far longer than is traditionally assumed both in psychology and philosophy**. Objects in the world have a surprisingly persistent dual character in the mind: Their objective shape "out there," and their perspectival shape "from here." For demos, visit **www.jorgemorales.xyz/perspective**.

Inattention Impairs Optimal Perceptual Decision-Making

Another widespread assumption in vision research is that perceptual judgments are the result of Bayes-optimal computations. When you see well, you do not need to rely much on what you know about the world to make accurate perceptual judgements. In contrast, when vision is poor—like when there is fog or when you are not paying attention—you are better off relying more on you prior knowledge about the environment. Nevertheless, the visual system does not always behave like this and it can, in fact, be highly suboptimal. Leveraging robust

psychophysical and neurophysiological findings, I demonstrate that counter to Bayesian predictions, **subjects systematically fail to incorporate prior knowledge into their perceptual decisions when not paying attention** (**Morales, Solovey, Rahnev, Maniscalco, de Lange & Lau, *Attention, Perception & Psychophysics*, 2015**). Importantly, this otherwise counterintuitive result was predicted a priori based on signal detection theoretical considerations. When not attending, subjects can only incorporate their priors optimally if they shift their response criterion dramatically; but humans struggle to make extreme trial-to-trial criterion changes which makes them *predictably* suboptimal.

Stereotypes Impair Incidental Perceptual Judgments

Social stereotypes shape our judgments about people around us. A striking example of stereotype bias involves the treatment of people whose identities run counter to our stereotypes—as when women are assumed to be students, research assistants, or nurses rather than professors, principal investigators, or doctors. My colleagues and I demonstrate that **long-held social biases**, such as those implying doctors are men and nurses are women, **can affect *extremely basic* perceptual judgments** (**Baker, Morales & Firestone, *Journal of Vision*, 2019**; under review—manuscript available upon request). By leveraging statistical learning, when we exposed subjects to a straightforward regularity present in images of *doctors and nurses* (e.g. that the latter faced left and the former faced right), they judged basic features of those images as if we had taught them something about *men and women*! This work was awarded the best poster prize at the 2019 Annual Meeting of the Society for Philosophy and Psychology. You can try our demos at **www.jorgemorales.xyz/bias**.

2. The Mechanisms of Introspection & Metacognition

The subjectivity of the mind is not limited to mental states directed towards the world; it also pervades mental states about our own minds. Via introspection, we can have a view about what states we are in and what these are like. But, **how do the mechanisms that support introspection work**? One of the most fundamental questions one can ask about any mental process is the domain over which it operates. In particular, in psychology we often make a distinction between *domain-general* and *domain-specific* processes. For example, thoughts are domain-general because they can be about anything (from rocks and Santa Claus to atoms and Wall Street). Other processes are domain-specific: Consider vision and how it is specialized for processing visual properties and nothing else. This distinction has touched practically every process in the mind and, in my research, I take **introspection as a case study in the analysis of domain-generality and domain-specificity of mental capacities**. My work asks a fundamental and until recently unsolved question: Does introspection rely on a common, domain-general resource that is recruited across many tasks (e.g. memory, perception, etc.), or is introspection supported by domain-specific components?

Domain-General & Domain-Specific Metacognition

Can you easily tell when you have rehearsed enough before delivering an important talk and when you need more practice to guarantee a good performance? Can you tell confidently how many people attended an event or do you always hesitate when estimating the size of a crowd? Importantly, do you think that being good at one thing makes you good at the other? You can make these introspective evaluations about your own cognitive performance thanks to metacognition—the capacity to monitor and control the success of one's own cognitive processes. Combining novel stimuli, psychophysics, neuroimaging, and computational modeling, I show that human **metacognitive abilities differ in the perceptual and memory domains**. I also demonstrate that both **generic and task-specific confidence-related signals co-exist in the brain**. (**Morales, Lau & Fleming, *Journal of Neuroscience*, 2018**). Such an architecture is consistent with lower-level feelings of confidence being combined with higher-order contextual information to allow effective behavioral control and decision-making. Together, these findings reveal both domain-general and domain-specific constraints on human metacognition, reconcile previously conflicting results on the domain-specificity/generality of metacognition, and lay the groundwork for a mechanistic understanding of reflective judgments of cognition.

The Calibration of Introspection

Introspection, then, has domain-general and domain-specific components. An even more general question we can ask about introspection is whether it operates in a unique fashion (i.e. is it the only capacity in the mind that

operates in the way it does?), or does it employ existing mechanisms that are then recruited for its own purposes? Moreover, should we care either way given that introspection is often proscribed from psychology under the charge of being utterly unreliable? Rather than ignoring or rejecting introspection, we should, instead, *calibrate* it. If introspection operates under similar principles to the ones governing other cognitive capacities, then we should expect it to be somewhat reliable at least some of the time. In my theoretical work, I show that **we can calibrate introspection's reliability by modeling it as a signal detection process akin to perceptual decision-making**. While we do not literally *perceive* our own mental states, the general kind of signal detection that explains perception can be leveraged to explain introspection. *In a slogan: Introspection is signal detection* (**Morales, "Introspection is Signal Detection"**, under review—manuscript available upon request). Using Signal Detection Theory, we can successfully predict subjects' perceptual capacity (e.g. sensitivity in a detection task) by modeling the strength of the internal perceptual signal produced by the intensity of stimuli. A similar framework applied to introspection can predict subjects' introspective capacity (e.g. reliability in an introspective task) by modeling an introspective signal produced by the *intensity or mental strength* of conscious experiences themselves. Thus, we can make predictions of subjects' introspective trustworthiness across a wide range of situations based on how intense or strong their conscious experiences are.

3. The Scientific Study of Consciousness

How does subjective awareness affect how we behave? How is consciousness implemented in the brain? And, crucially, **how do we study consciousness scientifically**? A huge barrier to studying consciousness scientifically is that it is the ultimate subjective phenomenon, and relying on subjective reports alone is often deemed as an unsurmountable problem for a science that aspires to draw objective and general conclusions about psychological phenomena. Despite these difficulties, my research advances a theoretical framework that rescues *the importance of subjective reports* in the scientific study of conscious awareness. At the same time, I propose principled ways for measuring consciousness in the laboratory while avoiding well-known pitfalls. In summary, my work provides general theoretical and empirical tools that we, as a field, can use to study not only the behavioral and neural profile of consciousness, but also other subjective features that pervade human psychology.

Metacognition as a Window into Consciousness

Confidence and conscious awareness are intimately related. When you are fully conscious of seeing a stimulus, you are likely to be confident in what you saw; and vice versa, if you do not see it consciously, you are unlikely to feel any confidence. Accordingly, metacognition in general, and confidence ratings in particular, are often used as a measure of consciousness. For example, instead of asking subjects to introspect and report what their conscious experience of a stimulus was like, subjects are asked to introspect and report how subjectively confident they are in their performance in a task. Despite its intuitiveness, the use of confidence for studying subjective conscious states is not without problems and, if this link is to be meaningful, it requires solid justification. In fact, consciousness research has relied so much on metacognitive measures (which in turn depend on confidence ratings) that the stability of the field as a whole is at stake when addressing this issue. In response to this important challenge, I have recently argued (**Morales & Lau, "Confidence Tracks Consciousness", in *Qualitative Consciousness*, Cambridge University Press, Forthcoming**) that **metacognitive measures in general, and confidence ratings in particular, are a valid measure for studying conscious awareness in the laboratory**. Subjective confidence ratings are more interpretable for both the subject and the experimenter than direct reports on awareness and yet they are highly correlated with them. Metacognition indeed can offer a window into conscious experiences.

Distilling Subjective Reports & Cognitive States from Consciousness

Ideally, to find the neural basis of consciousness, we need to compare the neural activity elicited when subjects report being aware of a stimulus and when they report being unaware of it. However, critics argue that activity in the prefrontal cortex does not reflect awareness proper but just cognition in general or even just *reports of* awareness (which are also subject to cognitive demands and response biases). These critics propose using indirect physiological markers of conscious awareness to classify conscious and unconscious states, thus allowing to study awareness without the alleged contaminating influence of subjective reports. But, for all the ingenuity and potential advantages of these so-called "no-report paradigms," **the science of consciousness is ultimately about subjective reports and we should not get rid of them**. As I have argued (**Michel & Morales, *Mind & Language*, 2019**), simply

providing a subjective report does not explain away the prefrontal activity associated with consciousness that has been described in scores of animal and human neurophysiology studies (**Morales & Lau, "The Neural Correlates of Consciousness" in *The Oxford Handbook of the Philosophy of Consciousness*, Oxford University Press, 2020**). In fact, designing experimental paradigms that completely eliminate cognition when we probe subjects' awareness is harder than it may seem (**Phillips & Morales, *Trends in Cognitive Sciences*, 2020**).

Eliminating Confounds in Neuroimaging Research

While subjective reports of consciousness are ultimately not a confound in consciousness research, there are plenty of other real confounds that are both hard to eliminate and detrimental for the validity of scientific findings—especially neuroscientific findings. Perhaps the most egregious of these confounds is *task performance*. Everything else being equal, we are better at discriminating stimuli that we see consciously than stimuli that we do not. This is why, unless subjects are comparable *in their perceptual capacity* (i.e. performance in conscious and unconscious trials is matched), we risk finding just the neural correlates of perceptual *processing* rather than the neural correlates of perceptual *awareness*. But **consciousness *can* dissociate from perceptual performance**, suggesting distinct computational and neural processes underlie these psychological functions. In fact, using psychophysical methods and targeted stimulus manipulations, I have provided a comprehensive exploration of this dissociation as well as a guide for creating optimal paradigms that can experimentally isolate subjective ratings from perceptual performance (**Maniscalco, Castaneda-Graham, Odegaard, Morales, Rajananda & Peters, *PsyArXiv*, 2020**). Moreover, relying on computer simulations, I demonstrate the importance of perceptual capacity matching between conscious and unconscious conditions, in particular when comparing neural activity recorded with EEG (**Morales, Chiang & Lau, *Neuroscience of Consciousness*, 2015**). Finally, I have led theoretical articles with philosophers and cognitive neuroscientists arguing for the importance of controlling for perceptual capacity in neuroimaging research on consciousness (**Morales, Mouradi, Sergent, Block, Taschereau-Dumouchel, Rosenthal, Grimaldi & Lau, *Neuroscience of Consciousness*, 2017**; **Morales, Odegaard & Maniscalco, *PsyArXiv*, 2019**).

Summary

As a whole, my research program studies subjectivity and how it affects what we perceive and what we take ourselves to perceive. Studying perception, metacognition and consciousness requires me to develop empirical and conceptual tools that can overcome well-known challenges in the study of subjectivity. In doing so, my research creates results, tools, and concepts that researchers working in other areas can use, thus facilitating a more direct and rigorous engagement with the subjective features of the mind. Importantly, I aim to address foundational issues in the study of the mind and of subjectivity. To this end, my research crosses disciplinary boundaries and makes use of behavioral, neuroscientific, and philosophical methods. Having these complementary lines of work allows me—and indeed requires me—to integrate research and ideas from disparate areas (such as memory, emotions, pain, action, developmental and social psychology, to name a few) that take different routes to study related problems (for example, approaches that use animal models, computer simulations, modeling and those that study the brain and behavior of infants and neuroatypical individuals). I am excited to continue this research program, and I look forward to developing it even further through close collaboration with colleagues old and new.