Examining naturalism in neuroscience

1. What do you take naturalism to mean? How does it influence the discourse in your field, particularly its conception(s) of human nature?

I understand naturalism to be a set of philosophical beliefs that aims to describe reality solely in physical terms, by appealing only to what is part of the natural world. Naturalism (and associated concepts, such as materialism and physicalism) asserts that only the physical is real, and that reality can only be meaningfully understood using the methods of science. Any appeal to a spiritual reality is rejected, and by extension, so too is the notion that religion constitutes a valid body of knowledge that shapes the trajectory of humanity's development.

Because naturalism rejects spiritual reality and entities (the eternal soul, a Creator, spiritual forces, etc.), it also rejects any appeal to values in order to explain phenomena around us. For example, prevailing theories of evolution emphasize that human beings appeared as a result of billions of years of genetic mutations that happened haphazardly as a matter of chance. The notion that the emergence of the human being may be an explicitly intended purpose of evolution, guided by values embedded within reality, remains anathema in most academic circles, because it would require the admission of values that cannot be explicitly proven using the methods of science. If evolution has a purpose guided by certain values, who determined them? Such questions lie plainly outside the realm of biology, so the tendency is to reject them entirely.

The resulting conception of human nature embedded within the natural sciences is that of a remarkably complex animal whose capacities are the product of a sophisticated brain refined through a long process of evolution. Although there are different branches of naturalism and not all of them necessarily lead to

neurological reductionism, within the context of the natural sciences, naturalism leads us to believe that the most effective way to generate knowledge about human nature is to study processes in the brain using the methods available to us through neuroscience. Human capacities such as learning, memory, emotions, and ultimately consciousness are reduced to patterns of activity among neurons in the brain. Though a research program examining the neural correlates of mental states is entirely valid, the insistence that human beings are exclusively physical bears significant consequences and, I believe, constitutes an assumption in itself, rather than a truth about human nature.

I should state at the outset that as a neuroscientist, I hold deep fondness and appreciation for my field, and I stand in awe of what my colleagues, past and present, have accomplished. The remarkable body of knowledge generated by neuroscientists about how thoughts, emotions, habits, and behaviors are scaffolded by the brain is an impressive one. It bears fruit in clinical settings, in technologies such as artificial intelligence, and at the level of policy in fields such as criminal justice and education. The thoughts I offer below about the limitations of a reductive materialist framework are offered not in a spirit of ruthless criticism, but instead with the hope of offering a perspective from which we can reframe insights and make progress on questions about human nature with which we are still grappling. Every scientific model by nature has a limited range of validity, and what I hope to do is to begin identifying the range of validity of naturalism in the context of neuroscience.

2. Why has naturalism become so widespread, particularly in certain intellectual circles in the West? What is so attractive about it?

Naturalism has become widespread in large part because many intellectual circles have seen it as a sign of a more mature humanity that is gradually

shedding its need for religion. Because historically religion has unfortunately been associated with unscientific rituals and superstition, the scientific world view and naturalism gained ascendancy in response. Often associated with the Enlightenment and some threads of thought within modernity, naturalism was celebrated as the ushering in of reason, and departure from dogma. As Western discourse becomes increasingly polarized around critical topics that have significant implications for human well-being, such as vaccine science, climate change, and the validity of evolutionary theory, naturalism grows deeper roots as people try to ward off misinformation and defend the rightful place that science holds in addressing a wide range of humanity's problems.

Within the field of neuroscience, I believe naturalism has become the dominant paradigm of the field because of the tremendous potential neuroscience seems to hold to address psychiatric disorders and mental health conditions. Viewing psychiatric illnesses as illnesses of the brain has served to remove the stigma associated with certain conditions by reframing them not as a failing of the individual, but simply as a problem in the brain (Corrigan & Watson, 2004). In short, if mental states are brain states, then aberrant mental states must simply be aberrant brain states. And if mental health disorders are biological states similar to other illnesses, then science should be able to treat them. Behavioral and systems neuroscience have developed a sophisticated level of rigor for studying psychiatric disorders in humans and in animal models and seems to offer a promising avenue for treating and preventing such conditions. Indeed, reversing drug addictions (Chen et al., 2013), alleviating conditions such as depression (Tye et al., 2013) and anxiety (Kim et al., 2013), manipulating memories (Ramirez et al., 2013; Ramirez et al., 2015; Redondo et al., 2014), decreasing aggression (Lin et al., 2011), curbing obsessive compulsions (Ahmari et al., 2013), and enhancing cognitive function (Roy et al., 2016)are now routine practices in animals using the tools of modern systems neuroscience.

Such impressive outcomes are less prevalent in humans because these interventions in animals are highly invasive and we lack the appropriate non-invasive tools to target neural components with the same precision in humans as we are able to in rodents. The assumption, however, is that we will eventually develop non-invasive tools for humans and experience a mental health revolution, to the extent that some neuroscientists even suggest that psychotherapy may be obsolete in the future. It is not difficult to see how such a vision is incredibly attractive to neuroscientists and clinicians alike, and I cannot help but admit that it is a very well-intentioned research program carried out by people who genuinely care about human well-being. I believe such a research program is an important one to advance and may indeed generate fruitful therapeutic strategies for some conditions, but it is somewhat short-sighted in that it neglects to consider that mental health is dependent on a confluence of factors related not only to the genetic and neurological composition of an individual, but also to adverse childhood experiences, oppression, undesirable aspects of culture, isolation, consumerism, and the harmful effects of social media - among other considerations (Sterling & Platt, 2022). Targeting a circuit in the brain that is responsible for generating a response to one's environment addresses the symptoms rather than the root cause of many conditions. In short, psychiatric illness cannot be adequately addressed until the very structures of society are addressed. However, I understand why it seems more straightforward to reconstruct a faulty circuit in the brain than to reconstruct a civilization built on faulty assumptions.

Another reason naturalism has become so prevalent in neuroscience is because of the development of powerful technologies such as artificial intelligence, which have shown remarkable propensity to mimic and even exceed human cognitive capacities. A rich set of technologies such as search engines, smart

phones, computers, pattern recognition algorithms, and self-driving cars have made many routine aspects of life more efficient and less labor intensive. Because these technologies are dependent on deep learning algorithms modeled on the basis of findings and theories in neuroscience, computational theory of mind (one of many expressions of naturalism) has gained traction as a powerful way to conceive of human cognitive capacity.

In addition to the fruits that neuroscience bears, there are also some broader conceptual reasons why naturalism is attractive to people. For instance, entertaining the idea of a soul that plays an important causal role in human consciousness, cognition, and behavior raises the question of mental causation – how an immaterial entity interacts with the material brain and body. This line of thought often attracts the derogatory label 'Cartesian dualism', on the basis of which one's position is often dismissed outright. The same question can be asked about how a non-physical God guides evolution by intervening in a physical universe. Naturalism appeals to people because it does away with questions of this nature entirely, and may in some ways be a cognitively simpler route to take.

Additionally, naturalism is attractive to some people because it avoids some misunderstandings that can arise in response to a teleological (purposeful) approach to evolution and human history. For instance, if we believe that evolution is teleological and was designed to give rise to the human being, exalted above the animal, does this mean we are superior and reserve the right to exploit the natural world? Moreover, if we adopt a teleological view of human history that has a certain end in mind, who determines what that end is, and how do we ensure that such a vision does not become a tool for some populations to oppress others by imposing their values? In many ways, people avoid teleological ideas in an attempt to preserve equality among people of different cultures/nationalities, as well as between the human being and the

natural world. Ironically, equality is in itself a normative value that cannot be explained or described within a naturalistic framework.

3. What contributions and/or difficulties does naturalism bring to the thinking around human nature?

Human nature is as much material as it is spiritual, and scientific approaches that attempt to understand the material aspect of our nature are valid and important. I imagine they will continue to constitute an important body of knowledge for humanity and generate countless fruits that will improve human welfare and well-being. I believe what will change is that we will develop more refined boundaries for the range of phenomena we expect science to explain, and the types of problems we expect it to solve, and give greater significance to spiritual solutions to spiritual problems.

One difficulty that naturalism brings to the thinking around human nature is that it is an oversimplified conception of reality and assumes that only one thing can be true at a time. However, insights from far less complex systems in the natural world suggest that even subatomic particles cannot be studied at such a level of simplicity. For example, the wave-particle duality of electrons implies that even electrons have to be conceived of in more than one way in order to be adequately understood. Moreover, the Heisenberg Uncertainty Principle suggests certain limitations of the human mind and our inability to grasp all of reality at once. It should not be difficult for us to extend these same ideas to the study of the human mind. If a single electron has multiple facets, then why should something as infinitely complex as consciousness be reducible to a single facet?

Another difficulty with naturalism in the context of neuroscience is that it makes some assumptions about causality, and at times might suggest causation when only correlation can be claimed. For instance, studies that use approaches to record brain activity (using for instance MRI or EEG) during a cognitive process or emotional state can identify the neural correlate of a particular process or behavior – e.g., neurons in brain region A are active when a person experiences happiness. The implication that is usually drawn here is that brain region A controls the subjective experience of happiness. However, linear correlations are correlations in either direction, and statements about the direction of causality are difficult to make – activity in brain region A may cause happiness (bottom-up causation), or happiness may cause activity in brain region A (top-down causation). Moreover, the cause of happiness may actually be something entirely outside the individual – a kind remark from a friend, an hour spent in nature, learning something new. In all of these cases, a neural correlate of a happiness state may exist, but it is not necessarily the root cause of one's happiness. Even studies that actually do use causal approaches must exercise caution in interpretation – stimulating a brain region and showing an increase in happiness would only suggest what a brain region *can be driven to do* if artificially stimulated by scientists in a certain way, but not necessarily that it is the cause of a person's happiness under standard conditions. Developing a more sophisticated discourse about causality and the interpretation of experiments will be a critical step for neuroscience in overcoming materialistic frameworks in the coming years.

Another difficulty with naturalism is that it ironically undermines the practice of science. Inherent to naturalism is the idea that values are not real, yet the practice of science relies on a belief that truth does exist, even if it is not fully accessible to us. How can scientists determine which models of reality are worth adopting and which we ought to discard if we do not believe that some

statements about reality are inherently better or more valuable than others? How do bodies of knowledge advance if we do not have some ideal of progress in mind? It seems, then, that naturalism undermines the worldview it so assiduously tries to uphold, because we must admit several foundational beliefs about truth and progress in order to carry out science effectively.

Finally, naturalism also undermines the most remarkable capacities of the human being that are needed to construct a new civilization, such as willpower, creativity, imagination, and spiritual insight. A prominent belief among many neuroscientists and cognitive scientists is that willpower and agency are illusions because all of our behaviors are ultimately controlled by the brain, not by an immaterial 'me' that is separate from my brain. Moreover, behaviorist frameworks that reduce the human being to a series of stimulus-response relationships with the environment leave no space to explain creativity and imagination, and the capacity to bring into the world of existence things that we have never seen before. Yet, a human being who lacks the willpower to act on his or her environment to bring about change, who cannot recognize and respond to challenges, who lacks the ability to imagine a better world and work laboriously towards its realization, cannot meaningfully contribute to the construction of a new civilization. Conceptions of human nature that describe a new kind of social actor are needed, and naturalism falls short in describing this social actor. In this sense, it may be useful to adopt a pragmatic approach by choosing the set of assumptions about reality and human nature that have the most practical utility in constructing a better world.

4. What scholar(s) has or have offered you insight into the relationship between human nature and naturalism? What points have they raised?

The work of philosopher Thomas Nagel offers brilliant insights into the relationship between subjectivity and objectivity that move beyond a naturalistic worldview (see his two books, *Mind and Cosmos* and *The View From Nowhere*). Though he is an atheist himself, Nagel expresses a healthy skepticism about how much can be explained using the language of science alone. He asserts that the current conception of objectivity employed by the natural sciences is insufficient to explain human consciousness. Nagel points out that physical conceptions of objectivity used to study the natural world require us to divorce our minds and subjective perspectives from reality in order to observe it objectively. Although this may have some merit in the practice of science, it poses problems when we want to study human subjectivity itself – because we must at some point admit that our thoughts and experiences are also objectively real things that exist, and that cannot be divorced from reality. Nagel's expanded conception of objectivity offers insight into the problems associated with trying to explain everything using the language of science. Moreover, Nagel insists that current theories of evolution which avoid teleological language fail to explain the emergence of the mind. By conceiving of the mind as a mere byproduct of evolution that evolved to help us survive more effectively, we undermine its capacity to know reality, and, in turn, undermine science itself. We must instead believe that reality is knowable, and that the human mind is an inherent and intended feature of evolution that was designed for the explicit purpose of knowing reality.

Similarly, I find that Karl Popper makes helpful contributions. Although Popper is mostly known for his work on falsification in philosophy of science, he also contributes some meaningful insights into the relationship between subjectivity and objectivity. In his essay "Three Worlds", Popper expands on dualism by instead insisting on pluralism. He describes three facets of reality: the physical world (World 1), the world of subjective psychological states (World 2), and the

world of knowledge and truths (World 3). Popper suggests that World 2 is an intermediary between World 3 and World 1 – that is, knowledge has effects on the physical world through the workings of the human mind. However, World 3 objects exist independently of World 1 and World 2, and cannot be reduced to them. By momentarily separating subjectivity and objectivity from each other, Popper illustrates that products of human thought such as scientific theories or moral values are best evaluated not on the basis of whether we subjectively like them, but how well they objectively align with objects in World 3. Although reality is far too complex to divide it into literal worlds, Popper's thinking offers valuable insight by examining the relationship between subjectivity and objectivity.

Finally, I have recently been delving into the work of Alfred Mele, who takes on the task of re-establishing a foundation for free will and agency. In his book *Effective Intentions,* Mele carefully examines famous experiments from neuroscience and cognitive science that purport to 'disprove' free will by suggesting that neural correlates for an action exist in the brain before a person is conscious of them. He shows how errors at the level of experimental design and inference have led to erroneous conclusions, and considers instead more tempered interpretations of the data. I find this work to be a rigorous and promising demonstration of how findings from neuroscience have ontological and epistemological commitments underlying them.

5. Are there any insights from religion that could illumine our understanding of naturalism and human nature?

Bahá'í writings that use the analogy of light to describe the relationship between the material and spiritual are useful in this context. 'Abdu'l-Bahá, one of the central figures of the Bahá'í Faith, writes: "Material civilization is like unto the

lamp, while spiritual civilization is the light in that lamp. If the material and spiritual civilization become united, then we will have the light and the lamp together, and the outcome will be perfect." Conceiving of the material world as a lamp and the spiritual world as light offers a helpful insight in conceiving the relationship between mind and brain. In this analogy, a lamp is the physical structure that is required in order for light to be made manifest. Its component parts must be arranged in an appropriate manner in order for light to shine, the same way that the component parts of a brain must also be appropriately arranged in order for capacities of the mind to be expressed. In this sense, a lamp is a conduit for light to be made manifest, and the brain is a conduit for the mind to be made manifest. Moreover, one can even establish causal relationships between the two. Damaging a lamp will prevent light from shining through it, the same way that damage to the brain will prevent capacities of the mind from being expressed. Yet, it would be a mistake to suggest that light is reducible to the lamp – no amount of description of the wiring and physical structures that composes a lamp will ever explain the nature of light itself. One needs a different kind of language in order to understand the nature and properties of light, the same way that we also need language of a different kind to describe the mind. By making this claim, we risk no danger of falling into unscientific thinking, because we retain the idea that the mind is connected to the brain, in the same way that light is to a lamp. All that we suggest is that multiple levels of description will confer greater insight.

In addition, I find that insights from religion can provide context for considering whether certain lines of research or technologies merit pursuit. For example, proponents of artificial intelligence suggest that we will eventually develop brain-machine interfaces that will allow us to rapidly upload large amounts of information into the human brain (e.g., entire languages and fields of study), such that the need for people to attend school will be dramatically minimized. Bahá'í

conceptions of education, however, suggest that education is akin to a process of kindling a light within a student, or mining the gems within them, rather than simply filling a receptacle with facts and information. In light of these insights, neuroscientists might temper their desire to replace education with technology. Similarly, another serious research agenda of some neuroscientists is to determine a way to completely simulate the human brain in the form of a computer, such that any individual's mind could theoretically be 'uploaded' into the hardware of a computer, and immortalized indefinitely, even once the individual's body dies. Such research programs are born partially of a fear of annihilation and mortality. However, insights from religion teach us that the life of the soul is eternal and continues long beyond our time on this material plane in realms not currently visible to us. Bahá'í writings liken the relationship between this life and the next to the existence of a fetus in the womb of the mother. Much as the fetus is unable to conceive of the world just inches away, we are unable to conceive of the spiritual realm immeasurably near to us. In this sense, attempting to immortalize ourselves in this plane is akin to a fetus attempting desperately to stay in the womb, unaware of the beauty and magnificence that awaits it on the other side. Empowered with these insights, neuroscientists can turn their attention to other lines of research that may bear greater fruit, confident in the knowledge that our immortality is already guaranteed.

Tara Raam is a Postdoctoral Research Fellow at the University of California, Los Angeles. Her work examines emergent properties that arise across multiple brains interacting simultaneously, in particular as social groups respond to external challenges together.

References

Ahmari, S. E., Spellman, T., Douglass, N. L., Kheirbek, M. A., Simpson, H. B., Deisseroth, K., . . . Hen, R. (2013). Repeated cortico-striatal stimulation generates persistent OCD-like behavior. *Science, 340*(6137), 1234-1239. doi:10.1126/science.1234733

Chen, B. T., Yau, H. J., Hatch, C., Kusumoto-Yoshida, I., Cho, S. L., Hopf, F. W., & Bonci, A. (2013). Rescuing cocaine-induced prefrontal cortex hypoactivity prevents compulsive cocaine seeking. *Nature, 496*(7445), 359-362. doi:10.1038/nature12024

Corrigan, P. W., & Watson, A. C. (2004). At issue: Stop the stigma: call mental illness a brain disease. *Schizophr Bull, 30*(3), 477-479. doi:10.1093/oxfordjournals.schbul.a007095

Kim, S. Y., Adhikari, A., Lee, S. Y., Marshel, J. H., Kim, C. K., Mallory, C. S., . . . Deisseroth, K. (2013). Diverging neural pathways assemble a behavioural state from separable features in anxiety. *Nature, 496*(7444), 219-223. doi:10.1038/nature12018

Lin, D., Boyle, M. P., Dollar, P., Lee, H., Lein, E. S., Perona, P., & Anderson, D. J. (2011). Functional identification of an aggression locus in the mouse hypothalamus. *Nature, 470*(7333), 221-226. doi:10.1038/nature09736

Ramirez, S., Liu, X., Lin, P. A., Suh, J., Pignatelli, M., Redondo, R. L., . . . Tonegawa, S. (2013). Creating a false memory in the hippocampus. *Science, 341*(6144), 387-391. doi:10.1126/science.1239073

Ramirez, S., Liu, X., MacDonald, C. J., Moffa, A., Zhou, J., Redondo, R. L., & Tonegawa, S. (2015). Activating positive memory engrams suppresses depression-like behaviour. *Nature, 522*(7556), 335-339. doi:10.1038/nature14514

Redondo, R. L., Kim, J., Arons, A. L., Ramirez, S., Liu, X., & Tonegawa, S. (2014). Bidirectional switch of the valence associated with a hippocampal contextual memory engram. *Nature, 513*(7518), 426-430. doi:10.1038/nature13725

Roy, D. S., Arons, A., Mitchell, T. I., Pignatelli, M., Ryan, T. J., & Tonegawa, S. (2016). Memory retrieval by activating engram cells in mouse models of early Alzheimer's disease. *Nature, 531*(7595), 508-512. doi:10.1038/nature17172

Sterling, P., & Platt, M. L. (2022). Why Deaths of Despair Are Increasing in the US and Not Other Industrial Nations-Insights From Neuroscience and Anthropology. *JAMA Psychiatry, 79*(4), 368-374. doi:10.1001/jamapsychiatry.2021.4209

Tye, K. M., Mirzabekov, J. J., Warden, M. R., Ferenczi, E. A., Tsai, H. C., Finkelstein, J., . . . Deisseroth, K. (2013). Dopamine neurons modulate neural encoding and expression of depression-related behaviour. *Nature, 493*(7433), 537-541. doi:10.1038/nature11740