

## Philosophical Hazards in the Neuroscientific Study of Religion

Daniel D. De Haan

To deny the truth of our own experience in the scientific study of ourselves is not only unsatisfactory; it is to render the scientific study of ourselves without a subject matter.

F. Varela, E. Thompson, E. Rosch, *The Embodied Mind*, 13

[G]iven that common sense recommends a world view with physical and psychological elements, the neuroscientific eliminativist about everyday psychology ... changes the subject when he speaks of states of the brain: he fails to bring a concept of *person* to bear when he says that commonsense psychology merits elimination.

Jennifer Hornsby, *Simple Mindedness*, 193<sup>1</sup>

### §1. Introduction

I am tasked with addressing philosophical hazards in the neuroscientific study of religion. As a philosopher concerned with the well-being of neuroscientists studying religion, I am inclined to begin with the philosophical hazards of philosophy. I am well aware of the extraordinary difficulties of both tasks, for the hazards are many and it is easy to miss the forest for the trees or the trees for the forest. Instead of focusing on one issue in great detail, I shall hang a number of warning signs around a forest of issues that identify various philosophical hazards which deserve particular caution when it comes to neuroscience and religion. Since I am aiming for breadth over depth, my brief remarks on each issue shall be synoptic, non-exhaustive, contentious, and suggestive for additional consideration and reflection. To redress such deficits, I have provided references for further reading.<sup>2</sup>

### § 2. Hazards of Identifying Hazards

I begin with a few hazards concerning the project of identifying hazards. First, most of the philosophical hazards I discuss are general problems for philosophy and neuroscience that are foundational with respect to the specific issues pertaining to scientifically studying religion. I start with these general hazards before touching on a few hazards specific to religion. Second, the dominant or “standard” stories in philosophy are not difficult to find; one can consult the online versions of the *Internet Encyclopedia of Philosophy* and *Stanford Encyclopedia of Philosophy (SEP)*, or the fine handbooks and companions published by University Presses. I aim to draw attention to hazards not often

---

<sup>1</sup> Francisco Varela, Evan Thompson, Eleanor Rosch, *The Embodied Mind: Cognitive Science and Human Experience* (MIT Press, 1993), 13; Jennifer Hornsby, *Simple Mindedness: In Defense of Naïve Naturalism in the Philosophy of Mind* (Harvard, 1997).

<sup>2</sup> The *Stanford Encyclopedia of Philosophy (SEP)* is an excellent resource for nearly any philosophical topic; its entries, written by philosophers or specialists, continue to grow and to be updated regularly. Most entries provide a general survey of a topic, the contentious issues surrounding different positions taken on some particular issue, as well as a representative bibliography on that topic. For limitations of space I shall not cite all the relevant entries from the *SEP*, but I recommend consulting the *SEP* for an expanded treatment of the topics I have raised here, especially for details on the dominant debates in analytic philosophy of mind. <http://plato.stanford.edu/>

flagged by these approaches. Many of these hazards challenge prevailing—sometimes merely fashionable—philosophical and scientific views, views that range from being false, to misleading, or merely require self-criticism or cautious charitable investigation of rival views. Sometimes it is important to remember that: default positions have their own uncritical assumptions; that the default view owes as much to historical contingency as it does to the cogency of its arguments; and, that there are persuasive *alternative ways* of seeing matters related to neuroscience and religion which cannot be ignored. Finally, due to limitations of space I am forced to make unqualified generalizations about various *hazards*, but the reader must recognize that there do not exist any nice and neat cookie-cutter shaped positions—*hazards* or otherwise—held by philosophers, scientists, and religious persons. Reality is much more complex than our abstractions, idealizations, and claims to knowledge sometimes suggest, including my own.

### § 3. *Explanandum*

The target of explanation for us is religion, the means by which it is to be investigated, studied, described, and explained is by neuroscience. What hazards might a philosopher draw attention to here? My central claim is this: The most fundamental philosophical hazard for the neuroscientist studying religion to avoid vigilantly is the failure to appreciate adequately the unified and integrated whole that is religious, namely, the human person. Human persons are religious, not their minds or brains. Neuroscience can only provide a *partial* explanation of religion insofar as the brain can only be but a *partial*—even if a *sine qua non*—factor in anything approximating a complete description and explanation of human persons, including those who are religious. Claims to explanatory comprehension that outstrip these limitations imperialistically exceed what is reasonable and true. The investigation of human persons and religion should be holistic and aim to illuminate, explain, and integrate the vast multilevel interconnections to be found from *subpersonal level* attributes like molecules, genes, organelles, cells, organic systems (e.g., cardiovascular, immune, endocrine, neuronal and glial systems), organs, organisms, to the ways these causally and constitutionally enable but neither eliminate nor provide explanatory substitutes for the complex *personal level* psychosomatic, rational, social, and historical factors that figure into the lives of human persons.<sup>3</sup>

This distinction between *personal level* and *subpersonal level* descriptions and explanations of human persons provides a helpful heuristic for demarcating and integrating the aforementioned diverse forms of enquiry, description, and explanation in the neuroscientific study of religious persons.<sup>4</sup> The expansive concepts of commonsense psychology—pain, sensation, perception, emotion, memory, imagination, enquiry, understanding, beliefs, desires, practical reason, and voluntary action—belong to the personal level insofar as the paradigmatic meanings and attribution of these concepts essentially involves mention of persons. A proper understanding of these personal level

---

<sup>3</sup> Mark Johnson and Michelle de Haan, *Developmental Cognitive Neuroscience: An Introduction* 4<sup>th</sup> ed. (Wiley-Blackwell, 2015); Carl Craver, *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience* (OUP, 2007). Charles Taylor, *Sources of the Self* (Harvard University Press, 1989); Robert Bellah, *Religion in Human Evolution* (Harvard University Press, 2011).

<sup>4</sup> This distinction has been used in a variety of ways by philosophers, psychologists, psychiatrists, cognitive scientists, neuroscientists, and others. I am roughly following a version of the distinction defended in Jennifer Hornsby, “Personal and sub-personal: A defence of Dennett’s early distinction,” *Philosophical Explorations* 3:1 (2000): 6-24.

psychological attributes must integrate their distinct manifestations from first, second, and third person perspectives. Understanding the many *ways* we accurately attribute the concepts of *belief* and *pain*, for instance, requires appreciating a host of behavioral, epistemic, social, linguistic, logical, and other criteria exhibited in the contrasts between *I believe X, you believe X, this person believes X* and *I am in pain, you are in pain, this person is in pain*.<sup>5</sup> Subpersonal level phenomena fall under the purview of psychological, cognitive, and neurological enquiries, descriptions, and explanations that address what factors underlie, constitute, enable, and cause phenomena at the personal level. Whereas the first and second person perspectives are essential and exclusive to personal level attributes, subpersonal level attributes are exclusively approached from a *distinctive form* of the third person perspective, namely, one that looks to the non-personal, non-conscious, non-rational, non-linguistic and principally biological factors that enable and make a difference to personal level attributes. The subpersonal level pertains to enquiry, descriptions, and explanations that are unavailable to the perspectives of first, second, and third person interactions with other persons *as* persons; it concerns what we can know about the human being's subpersonal psychology and biology by observations and inductive correlations achieved through invasive and noninvasive techniques, e.g., blood tests, X-ray, EMG, EEG, fMRI.

The personal/subpersonal distinction is not another version of the mental–physical dichotomy; most *personal level* descriptions—e.g., someone intentionally going to the store to buy milk—require putative mental *and* physical attributes. Indeed, the conceit behind the view I am urging recommends abandoning the mental–physical dichotomy in philosophy, psychology, and neuroscience, and adopting the integrated way of understanding human persons illuminated by the personal/subpersonal distinction. Significantly, the personal and subpersonal are not to be conceived as two independent things, they illuminate two ways of describing and explaining the complex and integrated features that constitute our embodied form of life as rational animals. Consider episodic or experiential memory; it admits of both personal level and subpersonal level accounts that are not in conflict or competition, but which can and should be integrated. Insofar as episodic memory shows up in enquiries, descriptions, or explanations that involve or require first-person, second-person, and third-person interactions—like narrating how someone met their spouse—then we are addressing the personal level features of episodic memory. So, the personal level pertains to both ordinary and clinical interactions and discourse concerning whether *I remember, you remember, or this person remembers* experiencing when, where, or what something or other happened. But insofar as we draw upon the resources of neurology, neuropsychology, or cognitive neuroscience—which concern medial temporal lobe damage and models of information processing, encoding, consolidation, and retrieval—we are looking to subpersonal level factors. Clinicians are especially adept at bringing together the resources of the personal and subpersonal level to provide a holistic account of, for instance, the symptoms of anterograde amnesia

---

<sup>5</sup> For a rich reflection on these differences, see P.M.S. Hacker's philosophical anthropology trilogy, *Human Nature: the Categorical Framework* (Blackwell, 2007); idem, *The Intellectual Powers: A Study of Human Nature* (Wiley-Blackwell, 2013); *The Passions: A Study of Human Nature* (Blackwell, 2018); Raymond Tallis's three volume philosophical anthropology, *The Hand: A Philosophical Inquiry into Human Being* (Edinburgh, 2003); idem, *The Knowing Animal: A Philosophical Inquiry into Knowledge and Truth* (Edinburgh, 2004); idem, *I am: A Philosophical Inquiry into First-person Being* (Edinburgh, 2005).

experienced by a patient and their family with imaging evidence of damage to the medial temporal lobe, and integrating both resources to find a holistic way to treat or cope with disorders.

The personal/subpersonal distinction is not a conceptual panacea; it is a heuristic or rough rule of thumb that helps us situate distinct but complementary forms of enquiry, description, and explanation about human persons. It shows one of several ways forward that allows us to reject the presumption that the only starting place for understanding ourselves is to adopt either dualism or reductive physicalism. Not only do these two metaphysical frameworks fail to exhaust the range of metaphysical views, they should not purport to be first principles of investigation; at best, they are among its possible conclusions. Prior to such investigations, there is no need to pigeon hole the human person into the conceptual straight jacket of either one of these restrictive frameworks.

#### **§ 4. *Philosophy, Science and Conceptual Frameworks***

From Socrates to the present, philosophers have been concerned with asking questions, understanding, and reflectively arriving at knowledge of the truth about the nature of reality. Philosophy draws attention to the fact that we all employ a worldview (*Weltanschauung*)—a conceptual framework that includes many concepts and categories—in our ordinary and scientific discourses about the world and ourselves. The collaborative human endeavor to know the truth about reality requires serious reflection on the overt and tacit ways in which we employ concepts and language to describe and explain the reality we experience. Unsurprisingly, there can be many subtle mistakes in the coherence and logical structure of our categories, arguments, and ways of thinking and speaking that can mislead our philosophical and scientific understandings of the world. Philosophy in particular challenges us to draw critical attention to the ways we use and understand such fundamental polysemous concepts as being, entity, identity, reality, phenomenon, nature, essence, kind, substance, property, relation, constitution, whole, part, datum, fact, concrete, abstract, modalities (necessity, possibility, contingency, and impossibility), state, event, action, passion, process, activity, agent, ability, disposition, power, capacity, manifestation, space, time, laws of nature, causation, supervenience, emergence, teleology, reduction, elimination, and many other concepts that are ubiquitously employed by us in a variety of ways to understand and arrive at knowledge of reality.

I introduce these polysemous concepts at the outset to draw attention to the fact that without many of them, reality would be unintelligible and the practice of science and its clinical applications impossible. Given the fundamental importance of these concepts in science there is a *desideratum* for scientists to cultivate and practice a disciplined use of concepts by reflecting on what they intend to mean when using such concepts, and know how to contrast the distinctive and connective meanings of these words and concepts that they do intend to employ, from those many other meanings that they do not intend to use. To fail to do so is to fall victim to the *Hazard of the Unreflective Conceptual Framework*. Disagreement over these concepts and their connections to reality abounds, but it is nevertheless important for all critical thinkers to dedicate themselves to the intellectual discipline of reflectively interrogating their conceptual framework. The aim is not to resolve all disputes about the proper meaning of these concepts, but to gain some critical perspective on the varieties of meanings used in the more specialized conceptual

frameworks employed in scientific, philosophical, religious, and other contexts. Without such conceptual discipline, one can never attain a critical perspective on how to interpret others or on how to anticipate and avoid unnecessary equivocations when stipulating technical uses of these concepts.

The interdisciplinary nature of cognitive neuroscience in particular demands greater attention to the *Hazard of the Unreflective Conceptual Framework* and vigilance about variant meanings and unintentional equivocations.

Cognitive neuroscience unavoidably traverses the boundaries between the neural cum physiological and the psychological, as well as the boundaries between the neural/physiological and the behavioral. The concepts in these three domains are categorically dissimilar. Their logical articulations are unlike, and the logical connections (of implication, exclusion, compatibility) between the different domains are exceedingly difficult to get into clear view. ... The different domains are not reducible one to another, and what explanations are appropriate to one domain may be inappropriate to another. How the description of phenomena in one domain bears on the description of phenomena in another is highly problematic.”<sup>6</sup>

There is an exigency for cognitive neuroscientists to attain a level of conceptual precision in their understanding of the psychological concepts they employ that matches the precision of their experiments. I am not alone in maintaining that the *Philosophical Foundations of Neuroscience (PFN)* and *History of Cognitive Neuroscience (HCN)*, co-authored by the neuroscientist Maxwell Bennett and philosopher Peter Hacker, provide one of the best places to begin an education in conceptual clarity in the domains of philosophy, psychology, cognitive science, and neuroscience.<sup>7</sup> Obviously neuroscientists do not need to become philosophers to do excellent neuroscience, but they do need to employ a rigorous conceptual framework in order to rigorously operationalize their definitions and interpret their experimental findings. And the detailed studies of experimental work in *HCN* and *PFN* provide an excellent resource and formation in how to achieve a conceptually rigorous framework. Many of their readers, myself included, will find much to disagree with in their work, but that is hardly grounds for dismissing the fundamental aims of the project and its usefulness for scientists. To those who eschew careful study of *PFN* and *HCN*, I ask: How could a meticulous critical reflection on the fundamental conceptual framework employed in any field of research be a disservice to that discipline?

Even the neuroscientist Joseph LeDoux has come to realize the importance of heeding the *Hazard of the Unreflective Conceptual Framework*. After decades of neuroscientific research on “fear,” LeDoux now believes it was a mistake for him to conceptualize this work as research on fear—a conscious emotion. What he and others are really studying are the nonconscious neural systems for threat detection. LeDoux entreats neuroscientists to aim for greater conceptual clarity.

[I]t seems obvious that scientists should be as precise as possible to short-circuit the opportunity for misunderstanding. If we can avoid confusion by simply changing the terms,

---

<sup>6</sup> Parashkev Nachev and Peter Hacker, “The Neural Antecedents to Voluntary Action: A Conceptual Analysis,” *Cognitive Neuroscience* 5 (2014): 194.

<sup>7</sup> M.R. Bennett and P.M.S. Hacker, *Philosophical Foundations of Neuroscience* (Blackwell, 2003); idem, *History of Cognitive Neuroscience* (Wiley-Blackwell, 2008).

why would we not do it?... By using different terms for conscious feelings and the nonconscious events that can, in some organisms, contribute to feelings in the presence of threats, much of the ambiguity and confusion about the neural mechanisms that detect and control responses to threats, and neural states that may result, is avoided.<sup>8</sup>

Scientific discoveries often lead us to introduce new, extended, or metaphorical uses of meanings and terms that are already part of our conceptual framework. This invites alternative ways of understanding reality, but it also opens us up to the possibility of misunderstandings. To avoid equivocations, stipulating novel technical uses of pre-existent terminology requires careful reflection about what is meant by these innovative extensions. For instance, cognitive psychology and neuroscience often draw upon the *personal level* terminology of commonsense psychology, like memory, emotion, and cognition, to develop technical terminology that pertains to the *subpersonal level* mechanisms of cognitive psychology and cognitive neuroscience, e.g., Alan Baddeley’s model of working memory, Endel Tulving’s episodic memory system, Michael Gazzaniga’s interpreter module, Giulio Tononi’s integrated information theory of consciousness. In fact, many of these psychological models promiscuously conflate and fail to distinguish between personal and subpersonal level features. LeDoux’s own self-critique evidences the massive confusions that result from uncritically employing the term “fear” which is a personal level concept of a conscious emotion, to characterize a subpersonal level cognitive neural mechanism that undergirds *not* the emotion of fear, but nonconscious “threat-detection”.

In short, the central contention of *PFN* enjoins us to ask important questions: If we do not know *what* we are talking about, then what progress can really be made? What are rigorous experimental protocols without rigorous reflection on the fundamental concepts operationalized to describe and explain the empirical phenomenon of interest? Raising objections to the exigency for conceptual rigor is as unscientific as objecting to the need for experimenters to concern themselves with experimental protocols and meticulously acquired data. Science requires a conceptually clear understanding of the basic concepts it employs, of the hypotheses it formulates, and of the ways its interpretations of data are situated within the conceptual framework of some theory.

### **§ 5. Psychological Discourse and Crypto-Cartesian Hazards**

Let us now turn our attention to the philosophical hazards bound up with a certain conception of the nature of psychological discourse, that is, the everyday attribution and predication of psychological concepts to ourselves, other humans, and nonhuman animals. The first is the *hazard of theoretical models of psychological discourse*. When a mother says her child has fallen in love with someone, is this attribution of “love” to her child always based on some psychological theory she has about a theoretical postulate called “love”? Or is the mother’s attribution of “love” relying on some everyday commonsense understanding of psychological concepts like love, joy, anger, desire, belief, and intention? Most contemporary philosophers and many scientists assume that commonsense psychological concepts (e.g., sensation, memory, belief, desire, intention) are non-observable *theoretical* entities that we postulate and attribute to others in order to predict and explain their observable behavior. If this “folk psychology” is a theory, then like any theory it is open to theoretical refutation or replacement by a more sophisticated cognitive

---

<sup>8</sup> Joseph LeDoux, “Coming to terms with fear” *PNAS* 111, 8 (2014): 2871–2878, 2876.

or neuroscientific theory of psychology, just as the folk physics and folk chemistry of astrology and alchemy have been replaced. Jerry Fodor, Daniel Dennett, and many others have argued for various reasons that folk psychology is a non-eliminable theory. In contrast Paul and Patricia Churchland contend that cognitive neuroscience will soon supersede the predictive and explanatory power of folk psychology, allowing us to eliminate the theoretical postulates of “belief,” “desire,” and so on. The widespread endorsement of various theoretical models of commonsense psychology has generated the problem of other minds as well as the seemingly irresolvable debates about the experiments in developmental psychology, comparative psychology, and evolutionary psychology about “theory of mind” or “mindreading,” that is, the ability of human children, adults and other animals to theorize, albeit tacitly via some theory or simulation, about the nonobservable internal mental attributes of other animals.

A small contingent of philosophers and psychologists reject the assumption that all psychological discourse is theoretical. It is true that there exist specialized forms of psychological discourse that are theoretical (e.g., the declarative and nondeclarative memory systems of Milner, Squire, and Kandel; Tulving’s episodic memory system; Fodor’s language of thought; the computational and representational theory of mind; propositional attitude belief-desire psychology, which is the mainstay of analytic philosophy of mind), but ordinary psychological discourse is not theoretical. The mother does not employ a theory of love or beliefs in order to ascribe love or beliefs to her child. The root problem underlying the *hazard of theoretical models of psychological discourse* can be identified with a host of more fundamental hazardous assumptions inherited from modern philosophers like Locke, Hume, and especially Descartes. While contemporary naturalists reject Cartesian mind-body substance dualism, many naturalists nevertheless embrace a Crypto-Cartesian conceptual framework. Crypto-Cartesianism denotes a *mélange* of philosophical hazards. The first is the twin *hazards of the mind-body and mental-physical dichotomies*. Adopting any version of these hazardous false steps—such as the division between the mind and body or between the brain and behavior—sets up an austere physical world against the disembodied ratiocinations of the mind or brain. By placing meaning, intelligibility, and reason outside of nature, these views encourage an unjustified separation of behavior from psychology. Indeed, it has been argued that Crypto-Cartesian approaches to cognitive neuroscience generate two problematic dualisms! First, the mind–mind dualism between the conscious mind and nonconscious computational mind, and a second mind–brain dualism that concerns the problem of how the brain can be an information processing computational mind.<sup>9</sup> Contrary to Crypto-Cartesianism, we do not need the mental-physical dichotomy as there are more accurate alternative ways of understanding and distinguishing animal behaviors; ways that acknowledge a plurality of descriptive and explanatory strategies for getting at the whole.<sup>10</sup>

Rather than taking for granted the *hazards of the mind-body and mental-physical dichotomies*, we should begin our investigations with the whole animal, which is a dynamic

---

<sup>9</sup> Valera, et al. *The Embodied Mind*, 52–57.

<sup>10</sup> William Jaworski, *Philosophy of Mind: A Comprehensive Introduction* (Wiley-Blackwell, 2011); Daniel Hutto, *Folk Psychological Narratives* (MIT Press, 2007); Matthew Ratcliffe, *Rethinking Commonsense Psychology: A Critique of Folk Psychology, Theory of Mind and Simulation* (Palgrave, 2007); Alva Noë, *Out of Our Heads: Why You are Not Your Brain, and Other Lessons from the Biology of Consciousness* (New York: Hill and Wang, 2009); Alasdair Macintyre, *Dependent Rational Animals* (Open Court, 1999).

psychosomatic unity that has evolved to interact with the affordances of its environment in a variety of fascinating ways. Crypto-Cartesianism also assumes a number of these closely tied up with the hazards of the mental-physical dichotomy that have been uncritically accepted as the default framework for investigating psychological behavior.

- a) *Hazards of the Mindless Body*: The body and its physical attributes, like behavior, are observable, outward, public, objective entities known from the third person perspective; some physical behavior of animals is explained by *theoretically* postulating or inferring private internal mental, cognitive, or neurophysiological causal mechanisms
- b) *Hazards of the Private Mind*: The mind and mental attributes are essentially private, unobservable, inner, subjective entities; mind or self is directly accessible through conscious first person introspection, but we indirectly and theoretically access other minds by inductive or analogical inferences based on observable behavior

Accepting these two hazards as gospel truth explains in part the genesis of theory of mind debates and the irresolvable philosophical problem of other minds, as well as the confrontation between behaviorists and cognitivists in empirical psychology. While behaviorists and cognitivists endorse incompatible views about what count as genuine psychological explanations, they share a common commitment to the view that behavior is observable and the mind is unobservable. In short, they both endorse the *hazards of the private mind* and *mindless body*, but behaviorists reject empirical explanations that theoretically appeal to unobservable mental postulates, and cognitivists maintain such theoretical postulates are necessary for empirical explanations of behavior. And this is why, given the *hazard of the mindless body*, the only way empirical psychology can address nonobservable mental attributes, is as postulated theoretical entities that provide causal explanations for overt behavior.

A similar story provides the impetus for assimilating all psychological discourse, including commonsense or folk psychology, to the theoretical model of psychological attribution. But we need not accept these assumptions. And rejecting Crypto-Cartesianism opens up conceptual space to consider alternative pictures of humans and other animals. Let us look at two complementary alternatives to the hazards considered thus far, the first being the account of psychological attribution detailed in *PFN*'s presentation of the *mereological fallacy*.

## **§ 6. Mereological Fallacy in Neuroscience**

Mereology studies the relationships between parts and wholes. “The neuroscientists’ mistake of ascribing to the constituent *parts* of an animal attributes that logically apply only to the whole animal we call ‘the mereological fallacy’ in neuroscience.”<sup>11</sup> The two formidable works by Bennett and Hacker, *PFN* and *HCN*, perspicuously and magisterially present the precise nature of the mereological fallacy, objections to it, scrupulously well documented violations of it, and the multitudinous hazards that ensue from overlooking it in studies on sensation, perception, cognition, memory, cogitation, belief, imagination, emotion, volition, consciousness, and self-consciousness by philosophers, psychologists, cognitive scientists, and neuroscientists.

---

<sup>11</sup> *PFN* 73.



It is important to draw attention to a common misunderstanding of their work. It *is not* an attack on neuroscience; it is an attempt to challenge and clarify the philosophical problems that underlie the uncritically endorsed conceptual framework operationalized by many neuroscientists in their research. It is conceptual medicine for a conceptual illness; we might not enjoy the dose or the therapy, but it is intended help neuroscientists acquire and operationalize a rigorous conceptual understanding that keeps up with their rigorously acquired data. Let us summarize a few salient features of their monumental contribution to the conceptual framework of neuroscience.

First, contrary to the *hazards of the private mind* and *mindless body*, the meanings of our psychological concepts are bound up with the characteristic psychological behavior—including linguistic behavior—they signify. The bodily behavior of animals *is* psychological behavior; the idea of “mere bodily movements” that need to be inductively, analogically, or abductively correlated with mental or psychological attributes is an abstraction of the behaviorists and philosophers. Long before humans achieve a mastery of language, let alone an understanding of psychological concepts, humans, like other animals, perceive and interact with the psychological behaviors of other animals. It is upon this pre-linguistic, pre-conceptual, and so certainly pre-theoretical basis that humans can transition from a perceptual engagement with the psychological behavior of animals, to a linguistic understanding and engagement with the psychological behavior of animals. And it is on the latter basis, that is, the eventual flowering of a linguistically rich commonsense psychology, that humans, quite late in their evolutionary history, transition into something akin to what might be called a theoretical investigation of psychology.

It is, of course, common for humans to be thinking or intending something, to be in mild pain or experience emotions without overtly exhibiting these psychological attributes.<sup>12</sup> Additionally, the meaning of psychological attributes, like anger, are not exhaustively characterized by the description of expressive bodily behavior, including stereo-typed facial movements. The mere description of the bodily movement of a sudden articulation of the knee does not discriminate between a reflex and a voluntary movement. But such abstracted descriptions are unlike our ordinary psychological ascriptions which occur within a wider ecological context that includes observing the psychological behavior that led up to and followed a knee jerk or voluntary movement. It is in this ordinary context that we observe and interact with the psychological behavior of others; this is the ecological setting in which we learn what psychological attributes mean by recognizing certain patterns of psychological behavior of animals as *criteria* for the ascription of certain psychological attributes to animals. This is why psychological behaviors are also partially constitutive of the meanings of personal level psychological predicates. It is not possible to describe aptly animal vision, audition, motivation, or executive functions independent from the stereo-typed psychological behavior the animal bearing these attributes exhibits in its environment. This criterial evidence from psychological behavior for the ascription of psychological attributes is defeasible by countervailing evidence. Humans and other animals are obviously capable of lying, dissimulating, deceiving, pretending, faking, and acting as if they are in pain, believe X, desire Y, and so forth. “However, if the criteria for a person’s being in pain, believing or intending something *are* satisfied on an occasion and

---

<sup>12</sup> PFN 82, n. 35.

are *not* defeated by countervailing evidence in the circumstances, then we are warranted in asserting that he *is* in pain, *does* believe or intend.”<sup>13</sup>

Second, we ascribe psychological behaviors to the animal as a whole, not merely to its parts. There are no *seeings*, *pain-sufferings*, *thinkings*, *desirings*, *runnings*, and *eatings* walking around in the world as independent entities in their own right. There are only seeing, suffering, thinking, desiring, running, and eating animals. Again, the very intelligibility of these basic psychological concepts is bound up with the psychological behaviors we can perceive and engage and eventually linguistically understand. This is the pre-theoretical point of departure for our basic experience and understanding of psychological attributes. This is why the *observable psychological behavior* of humans and other animals provides the primary evidence for the ascription of psychological attributes to them.

Third, the foregoing account of ordinary non-inductive psychological identification and attribution grounds the “possibility of inductive (non-logical) identification [which] becomes available through inductively correlating subjects’ having certain psychological attributes with other phenomena — for example, neurophysiological events in their brain.” However, if this inductive evidence from, say, fMRI conflicts with the “normal criteria for the ascription of a psychological predicates, the criterial evidence overrides the inductive correlation.”<sup>14</sup> Said otherwise, inductively establishing the neural correlates of some animal’s psychological attributes presupposes and depends on our ordinary non-inductive knowledge of and ability to ascribe these psychological attributes to the animal.

Fourth, it is for these reasons, among others, that the brain cannot satisfy the criteria for being the literal subject of psychological attributes. “Though neural phenomena are well correlated with an animal’s or a person’s being in pain, the brain does not exhibit *pain-behavior*... The observed neural phenomena that are concomitants of a person’s suffering pain, for example, are not forms of pain-behavior. They are *inductively* correlated with being in pain. The correlation is an empirical discovery, which presupposes the concept of pain and its nexus with criterial, non-inductive evidence for the application of the concept of pain *to a living creature* (not to its brain).”<sup>15</sup>

Fifth, as is explained at length in *PFN*, the mereological fallacy is not raising a jejune objection against the possibility of novel, metaphorical, analogical, technical, or synecdochical extensions of our ordinary psychological concepts. Rather, it is objecting to the conceptual confusions that result from mistakenly, misleadingly, uncritically, or inconsistently claiming to be doing any one of the former and then instead providing no justification for taking such psychological attributions literally or to be insightful, heuristically helpful, or explanatory. When “concepts undergo such analogical extension, something new stands in need of explanation.”<sup>16</sup> Without further elaborations, these novel uses lead to confusions and misunderstandings.

---

<sup>13</sup> *PFN* 83. Hence, *PFN* does not defend a form of methodological, psychological, or logical behaviorism. This is because the grounds for psychological ascription are not limited to behavioral evidence; empirical psychology is neither restricted to providing explanations of “mere bodily movements” nor is it limited to the theoretical resources of classical and operant conditioning; and, psychological attributes and their meanings are not essentially exhausted by their behavioral manifestations or dispositions for behavior. See *PFN*, 82, n. 35, 117; 416–417.

<sup>14</sup> *PFN* 83.

<sup>15</sup> *PFN* 83.

<sup>16</sup> *PFN* 77.

Talk of the brain's perceiving, thinking, guessing or believing, or of one hemisphere of the brain's knowing things of which the other hemisphere is ignorant, is widespread among contemporary neuroscientists. This is sometimes defended as being no more than a trivial *façon de parler*. But that is quite mistaken. For the characteristic form of explanation in contemporary cognitive neuroscience consists in ascribing psychological attributes to the brain and its parts *in order to explain* the possession of psychological attributes and the exercise (and deficiencies in the exercise) of cognitive powers by human beings.<sup>17</sup>

In sum, Bennett and Hacker are not, as is sometimes suggested, neuroscience skeptics; rather, they are critical of philosophical confusions that lead to misinterpretations of the very important achievements of neuroscientists. Their aim is to help neuroscientists free themselves from the shackles of conceptual confusions they have inherited from the conceptual framework of bad philosophy, so that they can proceed to do their important scientific work more perspicuously.

The author of "The Glass Onion" presents a critical evaluation of Bennett's and Hacker's *PFN*, especially its account of the "mereological fallacy." The summary of the mereological fallacy I just presented reveals that I fundamentally disagree with what I regard to be an unfair misinterpretation of, argument against, and concluding dismissal of the mereological fallacy presented in "The Glass Onion." It is an unfair misinterpretation and argument against the mereological fallacy, because nothing is said for or against the fundamental claim of Bennett and Hacker that ordinary behavioral criteria are constitutive of the meanings of our psychological concepts. Context does matter. The context in which we behave psychologically and later learn how to speak about and conceptualize psychological phenomena is very different from the contexts in which psychologists and neuroscientists perform their experiments and sometimes casually and sometimes literally take psychological concepts rooted in practices from the context of ordinary life and ascribe them to different subjects, like parts of the brain, in very different contexts without explaining if there are any differences between the meaning of memory and its behavioral criteria in everyday life, from the meaning of "memory" later ascribed to the medial temporal lobe which does not have that behavioral criteria. This is the crux of the mereological fallacy, and it is entirely overlooked by the criticisms of "The Glass Onion."

Additionally, the cheeky analogy based on treating Bennett and Hacker as a single psychological agent, the co-author "BH," whose philosophical attributes are explained by one part of BH, namely, H, belies a more important disanalogy. We know independent of their co-authorship that Bennett and Hacker are both independent psychological agents, whereas this is precisely what is under dispute about the brain: Is the brain also a psychological agent, a bearer of personal level psychological attributes? Our legitimate enquiries into what parts of the brain are subpersonal explanatory factors for the personal level psychological attributes of humans *depends upon* our basic and primary recognition of the psychological attributes that belong to human persons. However, if theorists wish to argue that it is explanatorily illuminating to ascribe the personal level psychological attributes of humans to the brains of humans (rather than focusing on subpersonal explanatory factors), then they need to demonstrate what is genuinely explained by re-introducing the same psychological attributes at another level or what new meanings are

---

<sup>17</sup> *PFN* 3.

given to these psychological concepts taken from the context of everyday experience and applied to the brain within the decontextualized environment of the laboratory or clinic.

### **§ 7. Psychological and Ontological Identity of Human Persons**

The *personal/subpersonal* distinction introduced earlier in this chapter is not equivalent to the *mereological principle* identified by PFN, though they are complementary. The mereological fallacy pertains to ascriptions of concepts to some part, that, given the paradigmatic intelligibility or meaning of some concept, it can only be literally attributed to a whole, not its parts. The mereological principle applies to wholes and parts that are artifacts and machines just as much as to humans and other animals. The *personal/subpersonal* distinction, as I employ it, contends that the meaning of some psychological attributes, namely, personal level attributes, are intrinsically bound up with the meaning of a person, a rational animal. We cannot understand beliefs, desires, and intentions apart from persons that have beliefs, desires, and intentions. If the complete story about, say, belief-desire psychology, leaves persons out and claims to get by without believing and desiring persons being in the picture, then something has gone terribly awry. We have, at best, *changed the subject* from the psychology of the believing and desiring human person, to that of a problematic conception of person-less beliefs and desires. Similarly, the application of these personal level psychological attributes to the nervous system, is to change the subject in a still more radical way. In both cases the person disappears as if irrelevant to our basic understanding of paradigmatically personal level psychological attributes. This uncritical form of changing the subject is one instance of the *Hazard of the Disappearing Person*. Bringing together the arguments afforded by the *personal/subpersonal* distinction and the *mereological principle* secures the claim that persons are attributed to wholes, like to rational animals, and not to any of their parts.

This last point bears further consideration. The human person constitutes a dynamically unified and integrated whole amid a magnificent spectrum of abilities and their manifestations. This brings us to the *Hazards of Mind, Body, Self Reification*. Without any doubt, it is certainly true that I am *myself*. I am also an *embodied being* with a range of intellectual abilities, and these might be equated with possessing a *mind*. Nevertheless, conceding such points does not mean we must accept the frequent reification of these attributes which obfuscates rather than clarifies the identity or nature of what it is to be a human person. The unity of the human person is not the unity of a *self, mind, or body*, but the unity of a developing rational animal with a panoply of powers pertaining to the integrated bio-psycho-social aspects of human life. These powers are manifested or impeded through various interactions and reactions with other things in reality. This unity of the human person is addressed by philosophical anthropology, but is often ignored by most work in philosophy of mind due to that discipline's myopic approach to the body, mind, and self. Unfortunately, the dichotomizing conceptual framework of *philosophy of mind* remains far more influential today than the holistic and integrated approach of *philosophical anthropology*.

While philosophy of mind commences with a dichotomy between the mental and the physical, the mind and the body, philosophical anthropology has a different point of departure. It starts with the human person insofar as it constitutes stable organic unity amid a dynamic manifold of developmental processes; we humans change, but we do not change in every way all the time. These contentions are related to a view of personal identity

known as *animalism*.<sup>18</sup> Long ago Aristotle resolved the nature versus nurture debate by a simple distinction between *first nature* and *second nature*. The first nature grounds the dynamic developments of our second nature, and our second nature allows for the gradual transformation of our first nature without undermining the stable *identity* of the whole. We can assimilate Aristotle's insightful distinction between first and second nature into contemporary parlance by employing a distinction between the *ontological identity* and *psychological identity*, respectively, of the human person.<sup>19</sup>

This distinction is critical for addressing questions of human or personal identity. Contrary to the *Hazards of Cartesian and Lockean Personal Identity*—which equate human identity with the psychological continuity of consciousness or memory—animalists contend that the loss or disintegration of one's psychological constitution through any of the myriad forms of human impairment, disability, disease, illness, and so forth, *does not* entail the loss of the more fundamental biological and ontological identity of the human person. This is because animalists hold that the more fundamental ontological identity of a human person can only be terminated through the death of the human organism. Humans cannot lose their ontological identity of being a human person from dementia or any other neuropsychiatric diseases no matter how extensive their suffering, memory impairments, or transformations of their psychological identity.

The psychological identity of the human persons, what we might call their *personality*, is manifested and disclosed both through ordinary human self-understanding and social interaction, as well through the keen and disciplined observations, descriptions, and explanations of the novelist, poet, clinician, behavioral and cognitive neuroscientist, psychologist, philosopher, and most importantly, our close friends and family. Our psychological identity provides a window into the ontological identity of the human person as a living rational animal, constituted out of sundry organized social and biochemical capacities and mechanisms and their integrated interactions. These points suggest alternative ways to conceptualize and understand human persons that eschew the false dichotomy of the mental and physical. To understand human persons requires integrated multilevel explanations that are complementary rather than in competition. We need to address questions concerning how the subpersonal neurophysiological dimension of our ontological identity enables our personal level conscious psychological *cum* social life, as well as how our psychological capacities and operations harness and order the psychosomatic neurophysiological dimensions of the human person.

Let us conclude this catena of hazards by reiterating a point briefly introduced at the outset. What I have claimed thus far is arguably *compatible* with a wide range of philosophical accounts of the ontological composition of the human person such as interactionist substance or property dualism, including the whole ambit of views like emergentism, panpsychism, as well as anomalous monism, functionalism, and even some versions of reductive physicalism. I reject all of these post-Cartesian views, and defend a version of Aristotelian hylomorphic animalism, but that is beside the point for our present

---

<sup>18</sup> See Stephan Blatti, "Animalism" and Eric T. Olson, "Personal Identity" in the *SEP*; Allison Thornton, "Varieties of Animalism" *Philosophy Compass* 11/9 (2016): 515–526.

<sup>19</sup> N.B. There are personal and subpersonal level attributes that apply to both our ontological identity (e.g., being a rational animal with a range of personal level psychological capacities that have subpersonal level explanations) and psychological identity (e.g., the particular personal level psychological operations and conscious experiences *here and now* which have definite subpersonal level explanations *here and now*).

purposes.<sup>20</sup> Thus, we should avoid the *Hazard of Simplified and Hasty Mind-Body Positioning*, as most views in psychology and neuroscience are compatible with a range of positions on the mind-body problem. Furthermore, it is not the mind-body problem but the neurophysiological and psychological constitution of the human person's abilities and activities that are more pertinent to the neuroscientific study of religion.

### § 8. *Consciousness*

Consciousness is a contentious issue; what follows is an opinionated sketch of consciousness and its hazards. There are two inseparable and unified facets of consciousness that we can distinguish according to two inextricable modes of presence: (1) the *intentional presence* of objects I am conscious of, and (2) *subject presence* wherein I am a subject conscious of intentional objects. Our conscious psychological operations are *intentional* insofar as they are *about* or *directed towards* some object: *seeing color; hearing sound; touching tangibles; perceiving or registering a dog; remembering my first pet dog; imagining my favorite painting or song; having a somatic ache on my back or hedonic feeling, being in love or angry with, afraid of, hopeful or sorrowful for some person; inquiring about the nature of justice; understanding Euclid's definition of a triangle; reflectively judging Libet's account of the brain science of free will to be true or false because of reasons X,Y,Z; deliberating about how best to travel to Paris; deciding to take the train to Paris; and, taking the train to Paris.*

Indivisibly bound up with the intentional character of our psychological operations is *subject consciousness* or *awareness*. I am the subject that is consciously *seeing, hearing, touching, perceiving, remembering, imagining, having somatic pains, passions, emotions, moods* and other affective and conative states, I am also the subject that is *wondering, understanding, conceptualizing and theorizing, rationally examining the evidence and reasons for some claim, reflectively judging it to be true or false, probable or unlikely, and engaging in self-reflective deliberative practical reasoning and action* concerning what is good for me and others here and now and how are such goods ordered towards the flourishing of myself and other human persons. We are also self-conscious, which "is not a matter of being conscious of something called a 'a self' ... but rather of a person's capacity to think about, reflect on, report and be conscious of his own mental states, beliefs, desires and motives, his skills, tendencies, attitudes and character traits, as well as his past life and experiences."<sup>21</sup>

Through the deleterious influence of misleading philosophical ideas about consciousness, empirical studies on consciousness frequently limit their categorization of consciousness to a kind of *mental state*. Three are noteworthy: (1) *Qualitative* or *phenomenal states* are conscious states that have some qualitative properties or *qualia* that might be restricted sensory qualities and so might include Thomas Nagel's gloss on consciousness as "what it is like to be an X". Cognitive phenomenology has recently extended these ideas to include the phenomenal features of cognitive and doxastic states. (2) *Meta-mental states* are conscious states about other mental states. (3) *Access consciousness* concerns the availability of mental information or content to the subject

---

<sup>20</sup> Daniel De Haan, "Hylomorphic Animalism, Emergentism, and the Challenge of New Mechanisms in Neuroscience" *Scientia et Fides* 5.2 (2017): 9–38; idem, "The Interaction of Noetic and Psychosomatic Operations in a Thomist Hylomorphic Personalism" *Scientia et Fides* 6.2 (2018).

<sup>21</sup> PFN 252.

independent from whether or not this information is phenomenally conscious. According to Ned Block's well known division of this conceptual territory, "Phenomenal consciousness is experience; the phenomenally conscious aspect of a state is what it is like to be in that state. The mark of access-consciousness, by contrast, is availability for use in reasoning and rationally guiding speech and action."<sup>22</sup>

All of these highly contentious divisions and subdivisions of consciousness are widely disputed in the philosophical and scientific literature.<sup>23</sup> But we should also be skeptical of philosophical and scientific projects that aim to treat real features related to conscious experience as distinct *properties* or *states* of consciousness that can be philosophically understood and explored empirically as discrete isolated objects of investigation. This modularization of consciousness encourages the misleading idea that, say, visual experiences can be decomposed into their distinct phenomenal (blueness), meta-mental (being conscious of experiencing blueness), and access properties (the un/availability of visual information pertaining to the experience of blueness), each of which can be studied on its own.

These *Hazards of Crypto-Cartesian Consciousness* can be avoided, for the sketch of consciousness I expounded above shows that a detailed picture of consciousness can be provided without any recourse to the problematic notions of phenomenal, meta-mental, and access consciousness and the tendency to reify these abstractions into discrete concrete objects of study. This alternative approach does not amount to a refutation of the dominant models of consciousness, but it does show that neuroscience is not obliged to remain hypnotized by the *Hazards of Crypto-Cartesian Consciousness*. Finally, this sketch also suggests that nature of consciousness can be addressed independently from settling quixotic ontological questions about physicalism, mental causation, and the mind-body problem.

### **§ 9. *Hard Problem of Consciousness and the Hazard of Zombie Neuroscience***

David Chalmers is well known for distinguishing between the *easy* and *hard problems* of consciousness. "The easy problems of consciousness are those that seem directly susceptible to the standard methods of cognitive science, whereby a phenomenon is explained in terms of computational or neural mechanisms. The hard problems are those that seem to resist those methods."<sup>24</sup> Most easy problems of consciousness concern the systematic identification of the neural mechanisms and correlates of conscious. Chalmers outlines a list of easy problems of consciousness that includes:

- the ability to discriminate, categorize, and react to environmental stimuli;
- the integration of information by a cognitive system;
- the reportability of mental states;
- the ability of a system to access its own internal states;
- the focus of attention;

---

<sup>22</sup> Ned Block, "On a confusion about a function of consciousness" *Behavioral and Brain Sciences* (1995) 18, 227-287.

<sup>23</sup> *PFN*, 237–351; Noë, *Out of Our Heads*; Dan Zahavi, *Self and Other* (OUP, 2014).

<sup>24</sup> David Chalmers, "Facing Up to the Problem of Consciousness" *Journal of Consciousness Studies*, 2 (1995): 200-19. See E.J. Lowe, "There are no Easy Problems of Consciousness" *Journal of Consciousness Studies*, 2 (1995): 266-71.

- the deliberate control of behavior;
- the difference between wakefulness and sleep.

The hard problem of consciousness—a version of the aforementioned problem of mind-mind dualism—concerns explaining conscious experience itself, including the range of phenomena identified with subject and object consciousness. Chalmers and others contend that the hard problem of consciousness has not and perhaps cannot be solved by the methods of neuroscience. Why? Aside from problems with the many failures of past attempts, one way Chalmers motivates this contention is by considering the possibility of zombies. Unlike the zombies of science fiction, a philosophical zombie is “someone or something physically identical to me (or to any other conscious being), but lacking conscious experiences altogether.”<sup>25</sup> The hard problem points out that the efforts of neuroscientists will continue to provide powerful explanations of zombie humans—via subpersonal cognitive processing minds—but will not come one iota closer to explaining what is distinctive of real humans, namely, consciousness. Chalmers’ argues that the conceivability of zombies entails their possibility and this undermines certain reductive and nonreductive physicalist views on the mind-body problem and shows that we must take consciousness seriously.

Despite my agreement with this final point, I think we should be skeptical about arguments that move from conceivability to possibility; furthermore, the conceit of Chalmers’s argument is deeply committed to the *Hazards Crypto-Cartesianism* I urged us to reject, and he also fails to distinguish, and so often conflates, personal with subpersonal level attributes. Aside from Chalmers’ purposes, however, I think philosophical zombies introduce a more modest point, call it the *Hazard of Zombie Neuroscience*, which is similar to the *Hazard of the Disappearing Person*. This hazard highlights difficulties with any purported claims to provide comprehensive explanations of human experience by psychology, cognitive science, or neuroscience that completely overlook or omit the fact that humans are conscious agents. Such alleged explanations of human persons could only be true of human zombies, not of human persons. The lesson to be learned is that no explanation of human persons can purport to be comprehensive if it has not adequately addressed human persons as being everyday conscious agents with the aforementioned spectrum of personal level capacities for conscious operations.

### § 10. *Causal and Explanatory Pluralism*

The debates on mental causation and free will in human and divine agents often share a strikingly similar blinkered theoretical commitment to causal monism. Causal monism is the view that causes are in competition to be the exclusive occupant of the one and only slot for the sufficient cause of some phenomenon. The *hazard of causal monism*, is closely linked with an exaggerated view of *parsimony* and the *hazard of extreme reductionism*. We should abandon such constrained understandings of causality and reductionism and acknowledge the wide range of causal explanations required for accounting for the multitude of factors that contributed to any phenomenon. Explaining a car accident, for instance, might require examining a range of causal factors like the poor choices of a sleepy driver, the way exhaustion affects attention and sensorimotor reaction times, poor visibility and wet pavement due to heavy rainfall, deterioration of automobile

---

<sup>25</sup> David Chalmers, *The Conscious Mind*, (OUP, 1996), 94.



tires, improperly designed banked curve, a deer running across the road, and so forth.<sup>26</sup> Against causal monism and forms of explanatory reductionism exponents of the new mechanist philosophy of biology, neuroscience, and psychology defend a form of explanatory and causal pluralism. Carl Craver argues that multilevel causal and constitutional explanations provide neuroscience with a mosaic unity.

This mosaic view of the unity of neuroscience is broader in scope than reduction because it covers both the integration of fields in research at a given level and in research that crosses levels. This mosaic view also provides a more accurate and elaborate view of interlevel interfield integration. Where reductionists understand the unity of science in terms of stepwise reduction to lowest levels, the mosaic view treats the unity of science as the collaborative accumulation of constraints at multiple levels. Whereas reduction focuses on relations of identity, supervenience, and ontological reductive links, the mechanistic mosaic view emphasizes the importance of explanatory relevance as the bridge between levels. Finally, whereas reduction models emphasize the importance of explanatory reduction to fundamental levels, the mosaic view can be pluralistic about levels, recognizing the genuine importance of higher-level causes and explanations. The mosaic unity of science is constructed during the process of collaboration by different fields in the search for multilevel mechanisms.<sup>27</sup>

Acknowledging the exigency to take seriously the hierarchy of integrated levels of causal and explanatory significance relevant for understanding human persons opens up clearer distinctions between complementary explanations from the personal and subpersonal levels. Indeed, the personal/subpersonal distinction sets a clearer research agenda for addressing the way the neurophysiology of subpersonal levels enable, influence, and are integrated into our capacities for conscious operations at the personal level.

Causal and explanatory pluralism not only elucidates the actual practices of scientists and provides a way to side step muddled philosophical debates about mental causation or determinism, compatibilism, and libertarianism in free will, but also contributes to theological reflection on the co-operative nature of divine and human action. By distinguishing sharply between what God does *qua* Creator from what humans do *qua* creatures, and seeing that there is no creaturely causation without a Creator, and that the Creator does not act *qua* creature, a significant advance can be made beyond the similarly tired debates about theological determinism.<sup>28</sup>

### **§ 11. Hazards of the Computational and Representational Theories of Mind**

Cognitive scientists typically define cognition as information processing that involves some form of computations—classical or connectionist—over representations that are causal, or intentional, or even phenomenal. The hazards here are many.

The concept of “representations” is employed ambiguously within and between different disciplines to sometimes mean mere causal covariation or *re-presentation* of patterns of information in one medium, e.g., photons bombarding a retina, to equivalent or

---

<sup>26</sup> Jaworski, *Philosophy of Mind*, 290–296.

<sup>27</sup> Craver, *Explaining the Brain*, 271.

<sup>28</sup> Brian Shanely, “Beyond libertarianism and compatibilism: Thomas Aquinas on created freedom” in Richard Velkley (ed.), *Freedom and the Human Person* (Catholic University of America Press, 2007), 70–89.

computationally processed patterns of information in another medium, e.g., neural spike trains in the visual cortex. Similarly, cognitive scientists often say a neuron or a node in a connectionist network *represents* the stimulus that it selectively responds to. But these meanings of “representations” as causal mediators are wholly different from, though they are frequently confused with, different accounts of “mental representations.” One account of mental representations understands representations to consist in “intentional” mental acts; typically as mental acts with propositional content. Our beliefs, thoughts, and other forms of “mental representations” are said to be “intentional” in the sense discussed before; they are *about* or *directed towards* something. Representations in the causal sense do not entail mental representations in the intentional sense. Confusions arise because while many theorists unreflectively conflate the two senses of representation, other theorists contentiously debate whether mental representations in the intentional sense are identical to, reducible to, or emergent from causal representations in the brain.

Debates also persist over incompatible theories of computation and representation; disagreements concern whether computational and representational theories of mind (*CRTM*) are helpful models, metaphorical heuristics, or literally describing what is occurring in the brain. Furthermore, proponents of *CRTM* take different stands on the modularity of cognitive systems, which is a distinct issue from whether the brain is modular. There are different theories of cognitive modularity, but most theorists would agree that modules are cognitive systems for processing domain specific information, e.g., perceptual or motor information. Fodor suggests that “the most important aspect of modularity [is] something that [he] call[s] “informational encapsulation.”<sup>29</sup> Modules are encapsulated because they only process domain specific information and are incapable of getting access to other kinds of information in a system. Informational encapsulation is often linked with the idea of “cognitive impenetrability” wherein a cognitive module is incapable of being influenced by information that falls outside of its input source or domain. Defenders of modularity are divided between theories of modest modularity à la Fodor (rejecting that central systems are modular because they do not have information encapsulation) or massive modularity à la Carruthers (maintaining the whole mind is modular based on arguments from evolutionary psychology). Proponents of *neural reuse* are skeptical of *CRTM* and reject the neural localization strategies of modularity theories on the grounds of neuroimaging meta-data analyses and alternative explanations of evolutionary psychology.<sup>30</sup> But even if some version of *CRTM* proves to be the most fruitful way of *modeling* neurophysiological cum psychological behavior of animals, this does not mean the brain is actually computationally processing representations.

Challenges to *CRTM*, some more *radical* than others, have come from alternative approaches to cognition that are often bundled together as the *4E's of Cognition*: Cognition is embodied, embedded, enactive, and extended. Not all proponents of these alternative approaches accept each of these 4Es; some accept modified versions of the *CRTM*, while others completely reject *CRTM* along with the claim that cognition is information processing. Advocates of *radical embodied cognitive science*, inspired by James Gibson and dynamical systems theory, reject *CRTM's* conceptual framework for interpreting data and defend an alternative that does not appeal to theoretical postulates like representations and classical computational processes. The *hazards of cognitive science* concern a

---

<sup>29</sup> Jerry Fodor, *The Modularity of Mind* (MIT Press, 1983), 36.

<sup>30</sup> Michael Anderson, *After Phrenology: Neural Reuse and the Interactive Brain* (MIT Press, 2014).

cautionary call to be vigilant and circumspect about these many debates and unresolved questions about the brain and whether and to what degree *CRTM* or any other accounts of cognition as information processing illuminate the neuroscience, psychology, and the subpersonal nature of the human person.<sup>31</sup>

### § 12. *Naturalism: Ontological, Methodological, & Epistemological*

Naturalism has been understood in many different and often incompatible ways—three are noteworthy. *Ontological naturalism* maintains that the whole of reality is compromised by nature or whatever is natural, where the natural is frequently taken to be whatever is constituted by or simply is identical to physical or material entities. This view excludes or denies that there are any supernatural or spiritual realities (God, spirits, immaterial souls or minds) intertwined with or distinct from nature. Insofar as ontological naturalism denies the existence of any reality that transcends nature such as the God of Abrahamic religions, it is a form of atheism. It does not—and cannot—prove God does not exist; rather, the nonexistence of God is entailed by the fundamental assumptions of ontological naturalism. On this front, it is important to point out that “One of the more persistent and inexcusable rhetorical conceits that corrupt the current popular debates over belief in God is the claim that they constitute an argument between faith and reason or between religion and science. They constitute in fact, only a contest between different pictures of the world; theism and [ontological] naturalism... each of which involves a number of basic metaphysical convictions...”<sup>32</sup> Challenges to ontological naturalism come from many directions: it is an unmotivated bald assertion, it cannot provide a naturalistic account, via evolution or otherwise, of the reality and knowledge of mathematical, logical, philosophical, moral, and aesthetic truths, and most fundamentally, for the human capacity to employ and know principles of rationality within the “space of reasons.”<sup>33</sup>

*Methodological naturalism* maintains that knowledge of natural phenomena should be described and explained by natural realities. Methodological naturalism adopts an epistemological stance concerning the way we can know nature, and so it does not directly entail any ontological commitments about the reality of non-natural or supernatural entities. Many methodological naturalists are ontological naturalists; however, commitment to the latter is orthogonal to the former. Methodological naturalism’s epistemological contention about the ways of investigating natural reality can be held *either* to exclude *or* to permit non-scientific descriptions and explanations of natural realities by supernatural or non-natural realities. Methodological naturalism typically

---

<sup>31</sup> Daniel Hutto and Erik Myin, *Radicalizing Enactivism: Basic Minds without Content* (MIT Press, 2012); Anthony Chemero, *Radical Embodied Cognitive Science* (MIT Press, 2011); Paul Churchland, *Matter and Consciousness* 3<sup>rd</sup> ed. (MIT Press, 2013); William Bechtel, *Mental Mechanisms Philosophical Perspectives on Cognitive Neuroscience* (Routledge, 2008); William M. Ramsey, *Representation Reconsidered* (CUP 2007); Bennett and Hacker, *PFN*.

<sup>32</sup> David Bentley Hart, *The Experience of God: Being, Consciousness, Bliss* (Yale University Press, 2013) 77.

<sup>33</sup> Thomas Nagel, *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False* (OUP, 2012); Daniel Robinson, ed. *Scientism: The New Orthodoxy* (Bloomsbury, 2014); Robert Koons and George Bealer, *The Waning of Materialism* (Oxford University Press, 2010); J.J.C. Smart and J.J. Haldane *Atheism and Theism* 2<sup>nd</sup> ed. (Blackwell, 2003); Alvin Plantinga, *Where the Conflict Really Lies: Science, Religion, and Naturalism* (OUP, 2013).

commends the scientific method as an ideal methodology for the investigation of natural phenomena; indeed, it privileges scientific knowledge over other forms of knowledge.

An *exclusive* methodological naturalism rules out the possibility of describing or explaining natural realities by appealing to supernatural entities. Significantly, this is a methodological constraint, not evidence of absence; for it leaves out as a matter of principle any consideration of what might not be deemed natural. Just as a metal detector on the beach is incapable of proving sand does not exist, and particle physics cannot show cells do not exist, so also methodological naturalism cannot prove there is nothing supernatural simply by assuming constraints to look only at what is natural.

A *permissive* methodological naturalism might hold that while science and other naturalistic disciplines only have recourse to naturalistic descriptions and explanations, the view is quietist or agnostic about the possibility of supplementing or complementing naturalistic explanations of natural phenomena with non-natural explanations. Alternatively, a *permissive* methodological naturalism might maintain that natural phenomena require both natural and supernatural explanations that are complementary, where naturalistic explanations address *proximate causes* of nature, which must be completed by supernatural explanations of *ultimate causes*. Such an account might hold that while empirical science is exclusively concerned with providing naturalistic explanations, philosophy or theology can provide both natural and supernatural explanations.

The *hazard of scientism* is a more extreme and often less sophisticated version of methodological naturalism that mistakenly identifies science and the scientific method with the claim that the only true, meaningful, or reliable knowledge about reality comes from empirical investigations of science. This mistake is problematic for many reasons. Aside from being self-defeating, scientism—an epistemological position—cannot itself be verified through empirical investigations. Scientism also rules out as non-scientific the non-empirical disciplines of mathematics and logic, which scientific methods presuppose and require.

The *hazard of naturalized epistemology* concerns the project and assumption that the best explanations for why some human claims to know that “X is Y,” comes from psychology and neuroscience. Extreme versions of naturalized epistemology contend that the best explanation for why a human person believes what they do, will not come from evaluating the personal level evidence and reasons that support their beliefs, but from examining the subpersonal neurophysiological and psychological mechanisms that lead to their formation of these beliefs. Like historicism, naturalized epistemology endeavors to explain away or debunk the person’s own reasons for belief by looking to something else that explains the origins of their belief, whether it be historical circumstances of race, class, and gender, or the psychological mechanisms selected through evolution. Naturalized epistemology arguably commits a subtle version of the genetic fallacy, and should be rejected for this reason and many others, not least is its self-defeating assault on our capacity to know the truth based on evidence and reason.

In sum: Naturalism matters for the neuroscientific study of religion insofar as some forms of naturalism rule out from the beginning the very possibility that some phenomena might also admit of supernatural explanations, while others methodologically ignore such possibilities, and still others leave open their possibility without investigating or by integrating them with naturalistic accounts. This is not the place to take a stand on such

contentious issues or the best way to proceed, however, it is critical to be aware of these forms of naturalism when approaching the neuroscientific study of religion.

In the final sections I will turn to a few philosophical hazards that are specific to the neuroscientific study of religion, beginning with a few noteworthy research programs.

### § 13. *Cognitive Science of Religion & Neuro-theology*

The neuroscientific study of religion should not be equated with either *cognitive science of religion (CSR)* or *neuro-theology*. While the first is merely the convergence of neuroscience investigations with religion, the latter two are ideologically driven interdisciplinary research programs that draw on neuroscience and cognitive science, developmental psychology, and evolutionary psychology to provide naturalistic causal explanations—often either vindicating or debunking—of the beliefs and practices of religion. Numerous *hazards* lurk around the assumptions of these research programs; I mention three *hazards of naturalized religion*, focusing on the scientifically more legitimate *CSR*. First *CSR* depends on scientific research programs that are problematic or very contentious. Evolutionary psychology is rightly criticized by evolutionary theorists and empirical psychologists alike for its slim empirical evidence, naïve correlations, and gratuitous explanations or suspicious “just so stories.” If evolutionary psychology is a house of cards, then the explanations of *CSR* rest on uncertain foundations.<sup>34</sup> Second, *CSR* is committed to the *massive modularity* view of cognition, which is also contentious, and is one among many views in cognitive science. And insofar as it takes these nonconscious subpersonal information processing modules to provide a complete explanation of the lives of religious persons, it completely leaves out the conscious personal level practices, beliefs, and experiences of religious persons. This zombie conception of religious persons runs afoul by ignoring the *Hazard of the Disappearing Person*. Whatever role subpersonal level cognitive information processing mechanisms play in the formation of the personal level doxastic convictions of religious belief or faith, we need to be careful to distinguish and not conflate the latter personal level beliefs or faith of the religious person from the subpersonal level information processing mechanisms postulated by cognitive psychology and cognitive neuroscience. Furthermore, *CSR* uncritically endorses some of the most problematic interpretations of massive modularity, and supposes that religion is developed unreflectively from our “core knowledge” and the default cognitive modules for intuitive psychology and biology. Such commitments have led many proponents of *CSR* to produce their own just so stories about how subpersonal level cognitive mechanisms for agency detection explain why religious people attribute agency to non-natural entities. Third, many proponents of *CSR* are committed to naturalized epistemology and the quest to provide cognitive explanations for why people are religious independent from whatever those person’s overt justifications are for their religious beliefs and practices. Neither the neuroscience nor neurology of religion require such assumptions to do solid scientific and clinical research.

Indeed, in “Religion and Neurology,” the first of his 1901-1902 Gifford lectures on *The Varieties of Religion Experience*, William James presents a view that is fundamentally at odds with naturalized epistemology and many facets of the *CSR*. James not only anticipated such problematic research programs, but also had the good sense to reject them

---

<sup>34</sup> Richardson, *Evolutionary Psychology as Maladapted Psychology* (MIT Press, 2007); Hilary Rose and Steven Rose, ed. *Alas Poor Darwin: Arguments Against Evolutionary Psychology* (Crown, 2000).

and distinguish such reductive explanations from the fruitful complementary explanations of neurology which do not replace personal level explanations of human persons. James also recognized the inconsistency hidden in the assumptions of naturalized epistemology applied to religion: for scientists also have brains, and perhaps the purported truths of evolutionary psychology or *CSR* can also be causally explained more adequately by eschewing their reasons and evidence and looking instead to the conditions of their brains.

In the natural sciences and industrial arts it never occurs to any one to try to refute opinions by showing up their author's neurotic constitution. Opinions here are invariably tested by logic and by experiment, no matter what may be their author's neurological type. It should be no otherwise with religious opinions. ... Immediate luminousness, in short, philosophical reasonableness, and moral helpfulness are the only available criteria.<sup>35</sup>

Like the research agenda set by William James, the neuroscientific study of religion is less concerned with the natural history of religion and efforts to vindicate or debunk religion through naturalistic explanations, and is more interested in understanding and potentially helping religious persons affected by neurological deficits. What philosophical hazards need to be identified here?

While not an overtly philosophical hazard, it is worth drawing attention to the importance of scientific research being ecologically sensitive. The beliefs, practices, and experiences of religious persons take place in environments and social settings that are obviously significantly different from the clinic or the alienating tunnel of a fMRI machine. These different contexts are not immaterial to the neuroscientific study of religion.

#### **§ 14. *Philosophical Hazards in the Neuroscientific Study of Religion***

Religious practices, beliefs, and experiences are personal level proprietary; there is nothing at the subpersonal level that is directly indicative of being religious for the same reason that there is nothing at the subpersonal level that manifests any other personal level attributes like beliefs, desires, intentions, reasons for action, anger, fear, frustration, despair, deep or tentatively held convictions. Of course, like any practices, beliefs, and experiences of human persons, religious ones admit of subpersonal level explanations. Indeed, it is unlikely to presume there will be a uniquely religious subpersonal level neuroscientific explanations any more than we would anticipate there to be *sui generis* "scientific" or "football" "classical music" subpersonal level neuroscientific explanations of a human person's practices, beliefs, or experiences vis-à-vis science, football, or classical music. As William James pointed out, "Religious melancholy, whatever peculiarities it may have qua religious, is at any rate melancholy. Religious happiness is happiness. Religious trance is trance."<sup>36</sup> The religious significance of the neuroscientific study of religious persons is entirely dependent upon the religious significance of personal level attributes; neuroscience's subpersonal level investigations do not introduce anything new with respect to the religious character of the religious person's practices, beliefs, or experiences.

This brings us to the *Hazard of Numinous Neural Localization*, that is, the effort to employ the techniques of neuroscience, especially brain imaging or neuropsychology, to

---

<sup>35</sup> William James, *The Varieties of Religious Experience*, Centenary ed. (Routledge, 2002), 19

<sup>36</sup> *Ibid.*, 24

locate the religious bits of the brain. Unfortunately, this hazard is sometimes embraced as a methodological agenda, as is illustrated by the aforementioned program of *neuro-theology*, which aims to correlate religious experiences with uniquely religious bits of the brain. This hazard should not be conflated with the perfectly reasonable neuroscientific research that has revealed stereo-typed lesions in certain parts of the brain correlated with specific forms of purported religious experience; indeed, such research expands the program of study of William James and others. And many of the papers in this volume contribute to the neuroscientific, neurological, and psychiatric study of religious persons in this valuable way.

### **§ 15. Conclusion**

I have argued that the neuroscientific investigation of religious persons do not address the personal level features that are essential to the practices, beliefs, and experiences of religious persons. Understanding these personal level phenomena is necessarily prior to and a prerequisite for coordinating the neuroscientific investigation of the subpersonal level neural systems that underlie, enable, and sometimes disable our abilities to practice religion or spirituality. These distinct forms of investigation are not in competition. What subpersonal level psychological and neuroscientific investigations provide are complementary explanations for what is affecting the person as a whole, which sometimes can confirm, challenge, or problematize personal level assessments. The challenge here is avoiding the philosophical hazards neuroscientific investigations of religion encounter on the way.