Uncovering the Face of Nature: Autopoietic Enactivism and a Systems View of Mind, Life, and Language

Jacob Pruett

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Monika Kaup, Chair

Henry Staten

Jesse Oak Taylor

Program Authorized to Offer Degree:

English

University of Washington

Abstract

Uncovering the Face of Nature: Autopoietic Enactivism and a Systems View of Mind, Life, and Language

Jacob Pruett

Chair of the Supervisory Committee:

Monika Kaup

Department of English

The study of life and the study of mind present us with a causality that is not easily explained. How can we reconcile the non-material purposes and intentions that define life and mind with the materialist explanations on which scientific practice depends? This dissertation project broaches this problem using the autopoietic enactivist (AE) approach to cognition. AE reads "cognition" into the foundations of life, and in doing so avers a profound continuity between mind, life, and the environment—as well as between human beings and all other kinds of so-called "lower" life forms. Using the work of Evan Thompson (*Mind in Life*, 2007) and his AE approach, as well as Terrence Deacon (*Incomplete Nature* 2013) and his unique contribution to contemporary systems theory which he calls "emergent dynamics," this project attempts to reconcile AE's observer-dependent phenomenological and cybernetic underpinnings with the observer-independent methods of neuroscience, evolutionary biology, and thermodynamics. Furthermore, it argues that AE's focus on system-environment couplings—and the embodied cognitive processes therein—provide contemporary literary theory with a robust framework for moving beyond the trappings of language and ideology inaugurated by structuralist linguistics. This project draws a coherent line between materialist science, systems theory, and the field of literary studies, and does so in a way that combines disparate—and seemingly incommensurable—scales of being and knowing, everything from the phenomenology of simple bacteria to contemporary approaches in literary formalism.

Acknowledgements

I would like to thank Dr. Henry Staten for introducing me to the profound and thought-provoking theories—as well as the two primary texts—that form the heart of this research project. I would especially like to thank Dr. Monika Kaup, who read through *heaps* of material in order to help get this project over the finish line: I am grateful for your steady encouragement, incisive feedback, and all the great novels you have introduced me to over these past five years. I would like to thank my good friend and interlocutor Jack Grace—our countless hours conversing at the gym, and at the crag, helped sharpen my understanding of a system and its relationship to philosophy. Finally, I would like to thank my wife, Jinna. Your love and friendship stand behind every hard-fought word of this project. I could not have done any of this without you.


Πάντες ἄνθρωποι τοῦ εἰδέναι ὀρέγονται φύσει

Table of Contents

Introduction:

The Face of Nature

This project began in the spirit of ethics. In 2016 I finished a master's thesis on Emmanuel Levinas, a Lithuanian-born French philosopher who wrote his main body of work during the postwar period, and who spent the majority of his professional life as a principal of a Jewish high school in Paris. As translator Annete Aronowicz tells us in the collection *Nine Talmudic Readings*, Levinas describes his project as one of translating Hebrew into Greek: "Hebrew" representing the ethical, theological, and mystical traditions of Judaism; "Greek" representing the secular and philosophical language that "Jews have in common with other inhabitants of the Western world" (Aronowicz xxxvii). All his work can be located at the nexus of these two traditions.

In the years following my master's degree, ethics would come to occupy a lot of my thinking, particularly the phenomenological approach taken by Levinas in his first major work *Totality and Infinity* (1961). Rather than denote a system of moral prescriptions—a catalogue of *oughts* meant to govern behavior—Levinas describes ethics as a metaphysical structure of intersubjectivity. Ethics is revealed in the affective and prelinguistic *experiences* in which a consciousness, existing within its own existential lifeworld, becomes subject to another field of vision. This is the face-to-face encounter with the

Other.[1] The appearance and appeal of the face calls the self out beyond its finite circle of care and into an infinite horizon of responsibility.

Prior to Levinas I had never read any work of phenomenology, and I was struck by the strangeness of the method. Because he is attempting to describe an experience, and not outline a system of moral precepts, the reader is compelled to adopt a new kind of reading strategy. You cannot simply wrest from Levinas a neat package of ideas to be catalogued, carried away, and applied. Levinas calls ethics "an optics" (Levinas, *Totality and Infinity* 23).[2] It is meant to defamiliarize the world, alter our perceptions of others, and make us feel something strange and new. These phenomenological dimensions often make it feel more literary than philosophical.

Levinas begins his phenomenological meditations from a place of autonomy. He describes a consciousness at home in the world and engaged in acts and activities that further one's self-possession and self-determination. Yet at all points his analysis seeks to trouble and defamiliarize the reflexive tendency to view the world as a function of self. In short, *Totality and Infinity* attempts to read the world from the standpoint of the Other, and not, as has characterized the majority of Western philosophy, from the standpoint of the same (*T&I* 43). This is the radical move at the heart of his phenomenological project.

Totality and Infinity is foremost "a defense of subjectivity…founded in the idea of infinity" (*T&I* 26). Infinity is a thought that overflows itself. Infinity is a concept that

---

[1] Capital "O" other denotes another person. Lower case "o" other denotes alterity—a radical otherness that cannot be incorporated into, or appropriated by, one's cognition.

[2] Henceforward cited as *T&I*

cannot be reduced to its predicate. It cannot be understood on the basis of identity, substance, or presence. It cannot be made to equal itself. Infinity never simply is what it is. The more one thinks about the infinite the more one produces of it and the deeper and less resolved one's relationship to it becomes. Levinas defines infinity in structural terms as *the more within the less* (*T&I* 50). Infinity, so conceived, is a rupture.

The infinite is most concretely revealed in the everyday encounters we have with another person that faces me. In the face of another person, one is presented with a trace of infinity. He says, "The way in which the other presents himself, exceeding *the idea of the other in me*, we here name face" (*T&I* 51). The face, so conceived, is not a physical face, as this would reduce the ethical experience to an ontological object. Rather, the face is the phenomenon through which the infinite instantiates itself, revealing to the self the radical otherness (alterity) of the other—that which exceeds all one's attempts to cognize it, to circumscribe it in categories and signifiers or otherwise reduce it to an object of one's own cognition. The face refuses to become wholly subjected to another's gaze.

How does the face accomplish this? Levinas says, "The face is a living presence; it is expression. The life of expression consists in undoing the form in which the existent, exposed as a theme, is thereby dissimulated. The face speaks. The manifestation of the face is already discourse" (*T&I* 66). By this reckoning the face is an incommensurable act of expression which at all times undoes and overflows the forms in which its living presence is concealed. Thus, the face is not reducible to an object of pure *theoria* (contemplation) because at every moment it is exceeding itself as mere object. In this description, speech and discourse refer not so much to an identifiable set of linguistic procedures or abilities, but to a fundamental capacity to undo and exceed the ossifying effects of another's gaze. If the face is speech, then its primary mode is interrogative: it questions one's sense of freedom and autonomy.

These are all beautiful sentiments. They represent a fervent desire to respect the Other as other, by refusing to let their otherness become subject to the will and intention of the same. But I soon encountered a troubling moment in Levinas' philosophy, one that would propel me into a decidedly ecological and environmental frame of mind, and which has motivated much of this dissertation—in spirit if not always in content. In an essay entitled "Philosophy and the Idea of Infinity," Levinas makes a flippant remark about animal being. He says:

> A face…differs from an animal's head in which a being, in its brutish dumbness, is not yet in touch with itself. In a face the expressed attends to its expression, expresses its very expression, always remains master of the meaning it delivers. A 'pure act' in its own way, it resists identification, does not enter into the already known, brings aid to itself, as Plato puts it, speaks. The epiphany of a face is wholly language. (55)

Simply put, the Other for Levinas is a uniquely human other. This was an ugly revelation: the idea that only *people* can be bearers of a face, that only humans have subjectivity. He differentiates human subjectivity from animal being by granting a special status to speech and language, but as in *Totality and Infinity* this capacity for language seems to boil down to the ability to "attend," in an ongoing manner, to one's own processes of objectification by the same. Language is not understood as an abstract and conventional system of reference and meaning making. Instead, language seems to stand for a more fundamental capacity for agency and autonomy—an interiority that challenges and exceeds objectification. If this is so, then why does Levinas exclude animal being from its ambit?

I was then, and am now, utterly moved by Levinas' phenomenology of the face and by the ethical relation therein. The face *speaks* precisely because we intuit an agency, a mentality, and a capacity for suffering that together manifests a primordial injunction "Thou shalt not kill." But Levinas' commitment to linguistically determined notions of subjectivity ends up betraying the spirit of his own ethics with respect to other organisms. He falls into an all too easy anthropocentric fallacy: that subjectivity only belongs to a *speaking* subject. Despite all the beautiful and penetrating insights his phenomenology unfolds, Levinas' overvaluation of spoken language ends up foreclosing his ability to perceive the faces of nonhuman animals.

Levinas' failure is part and parcel of a deeper anthropocentric impulse. Failure to understand one's ethical relationship to non-human animals and other so-called "primitive" or less-developed organisms stems from the superiority with which we view our own cognitive capacities. There are many reasons for such anthropocentrism, but I think Levinas' own cognitive dissonance is representative of a much more pervasive prejudice: that because animals do not use spoken language that they are somehow devoid of interiority and mentality; or that because they lack "higher" forms of cognition—such as linguistic or other symbolic modes of thought—that whatever interiority they do possess must necessarily pale in comparison to our own such that we do not stand in an ethical relationship to them.

This strikes me as an interpretive problem on two fronts. First, we lack a clear and consistent understanding of how our cognitive capacities emerge from, and participate in, the most fundamental processes of the natural world; second, we fail to perceive all the commonalities that exist between our own cognitive processes and those of other organisms. Together these failures make it all too easy to situate human being as somehow separate from, and superior to, the other inhabitants of the natural world, and once *this* idea takes hold, it becomes all too easy to efface and deface nonhuman animals and by extension the larger ecological collectives they inhabit.

Autopoietic Enactivism: A First Glance

This is not an ethical project per se. In the chapters that follow, I will not pursue any explicit political agenda regarding animal rights or environmentalism. My ambitions are more theoretical and fundamental. This dissertation attempts to intervene in these anthropocentric prejudices surrounding nonhuman subjectivity by attending to the concept of *cognition* in its place.

In the past twenty years, there has been a spate of research and renewed interest in perception, cognition, and consciousness across the suite of disciplines that constitute cognitive science (i.e., psychology, neuroscience, philosophy, linguistics, computer science, and anthropology). One of the most dominant and far-reaching paradigms in this new cognitive science is the 4E approach, which looks to the Embodied, Enacted, Embedded, and Extended aspects of mental life. It departs from earlier theories by downplaying, or rejecting outright, the idea that cognition is a skull-bound process that operates by way of abstract mental representations and computations, and thus the sole province of the brain and nervous system. Instead, 4E cognition looks to the various ways that thought and perception are determined by bodily affordances, sensorimotor "know-how," and environmental contexts and constraints. This work has revised many long-held scientific and philosophic assumptions about the nature of cognition, for as soon as we start to recognize the *embodied, embedded, enacted, and extended* nature of mental life, then it becomes untenable to treat the world as some separable, objective, and universally applicable set of conditions that the brain passively registers in its neurons.

This dissertation will focus on the enactivist account of cognition; more specifically Evan Thompson's autopoietic-enactivist approach developed in his 2007 book *Mind in Life*. Of all the 4E approaches, Thompson's enactivism attempts the most far-reaching analysis and integration of mental and biological processes—everything from the most basic mechanisms of bacterial chemotaxis to the most complex interactions of the Earth system and its myriad ecologies. My hope is that by scrutinizing and coming to terms with this autopoietic-enactivist account of "cognition," I can contribute to a growing body of scholarship in the environmental humanities that is attempting to counter and qualify the hegemony of analytical frameworks that conflate spoken language and subjectivity in our theorizing of nonhuman life and its worth.

Thompson, along with psychologist Eleanor Rosch and biologist Francisco Varela, first articulated enactivist theory in their 1991 book *The Embodied Mind*.[3] As Thompson

---

[3] (For clarity) This project has two primary intertexts: *Mind in Life* by Evan Thompson and *Incomplete Nature* by Terrence Deacon. Parenthetical citations that refer to "Thompson" or "Deacon" will refer implicitly to these two primary works.

tells us in the opening pages of *Mind in Life*, "…the enactive approach aimed to build bridges between embodied dynamicist accounts of the mind and phenomenological accounts of human subjectivity and experience" (Thompson 13).[4] In *Mind in Life*, Thompson updates and expands their initial methodology in a number of useful, multidisciplinary ways: for example, by tracking the development of neurophenomenology and its growing import in the field of cognitive science; by placing enactivism in the context of evolutionary and cell theories; and by closely examining important figures in phenomenology who prefigured core enactivist ideas. But most importantly, *Mind in Life* explicitly and systematically extends the notion of enaction developed in *The Embodied Mind* down into the most basic biological and organismic scales via a sustained analysis of autopoiesis and developmental systems theory.[5]

As a general point of departure Thompson breaks down the field of cognitive science into three principal approaches: cognitivism/computationalism, connectionism, and embodied dynamicism. Thompson situates enactivist theory as a subset of the last

[4] In *The Embodied Mind*, Varela and co. situate the notion of "enaction" between two disciplinary poles. The first is philosophical hermeneutics (à la Martin Heidegger and Hans-Georg Gadamer) in which the act of interpretation is "understood as the *enactment* or *bringing forth* of meaning from a background of understanding" (Varela et al. 149). Here they emphasize that in all these hermeneutical approaches, "knowledge depends on being in a world that is inseparable from our bodies, our language, and our social history—in short, from our *embodiment* (Varela et al. 149). The second is empirical. They look to the self-organizing properties of simple systems—specifically cellular automata—which are "structurally coupled" to a particular environmental milieu and its history of changes and transformations. They say, "By enriching our account [of cellular automata] to include this dimension of structural coupling, we can begin to appreciate the capacity of a complex system to enact a world" (Varela et al. 151). We will turn to the issue of structural coupling shortly.

[5] Without doubt, the notion of enaction pursued throughout *The Embodied Mind* is informed by the theory of autopoiesis (Varela being one of its two original architects), but it did not play an explicit role. In contrast, *Mind in Life* grounds its analysis of cognition in terms of autopoiesis from the very beginning. Therefore, when I invoke the notion of "autopoietic enactivism" I am referring foremost to Thompson's elaboration of the theory in *Mind in Life*.

major approach to emerge in cognitive science: embodied dynamicism. This approach rejects the cognitivist view that mental life is radically "skull bound," separate from the world and its goings on, and merely a mode of "abstract representation: symbolic or subsymbolic representations in the mind-brain [that] stand for states of affairs in some restricted outside domain that has been specified in advance and independently of the cognitive system" (Thompson 10). The key phrases here are "specified in advance" and "independently of the cognitive system." The embodied dynamicist approach critiques the core cognitivist position that thinking primarily involves mental representations that are produced in an abstract internal syntax, one that can be separated from the embodied and environmentally embedded context of cognition. Thompson continues, "Like connectionism, embodied dynamicism focuses on self-organizing dynamic systems rather than physical symbol systems…but maintains in addition that cognitive processes emerge from the nonlinear and circular causality of continuous sensorimotor interactions involving the brain, body, and environment" (Thompson 10-11). Like connectionism, the embodied dynamicist approach looks to the circular causality of dynamic systems, in which a system's outputs are fed back into the system as inputs thereby superseding cognitivism's linear input-output model. But unlike connectionism, which focuses entirely on the internal working of the nervous system and the brain primarily via the computer modeling of neural networks, the embodied dynamicist account extends these circular causal loops beyond the body and out into the world. In short, the embodied dynamicist intervention redraws the boundaries of the nervous system to include its environmental context.

Thompson aligns enactivism wholly with the embodied dynamicist approach in cognitive science but with the added dimension of phenomenology, which seeks to reintegrate subjective, first-personal *experience* into neuroscientific theory and practice. An objective world clearly exists. Things impinge on cognitive systems from without, and we can choose to study this world in a mind-independent way. But when the object of study is cognition itself, this mind-independent, third-personal stance fails to capture the inherently first-personal, subjective experience of mental life. Thus, for enactivism, what is decisive is the way the cognitive system actively engages with this world as a function of its own phenomenology (interiority), and in so doing, brings forth its essential qualities.

This process of *bringing forth a world* is where enactivism, phenomenology, and literary studies exhibit their most significant points of overlap—connections that will be explored in greater depth in chapters four, five, and six. But for the purposes of this introduction, I want to focus briefly on a tripartite breakdown of embodied cognitive processes that Thompson lays out for us in *Mind in Life*, and which offers us three distinct scales for viewing the enactivist problematic. He says, "Our mental lives involve three permanent and intertwined modes of bodily activity—self-regulation, sensorimotor coupling, and intersubjective interaction" (Thompson 243).

The following details each of these scales in turn. *Self-regulation* refers to the basal scale of cognition, which Thompson grounds in the discourses of cybernetics, second-order cybernetics, and autopoiesis. Cybernetics studies control and feedback mechanisms in both natural and technological systems. These control mechanisms are understood via their causal circularity: closed loops or circuits in which outputs are fed back into the system as inputs. There are two basic types of circular (feedback) processes. Negative feedback loops are those which constrain or dampen a process, in effect stabilizing the system. As Bruce Clarke notes, "Cybernetics' name for this form of system regulation is homeostasis—correcting discrepancies so that operations stay the same on average or within a range of a desired level" (Clarke 38). In contrast, positive feedback loops enhance or amplify changes in a system. The first implementation of cybernetic design occurred in the late 1940s with the invention of the homeostat. Initially concerned with engineering questions, cybernetics focused on control mechanisms as they related to machines. But early cyberneticists recognized that "…certain aspects of physiological systems ('in the animal') and technological systems ('in the machine') could be considered formally equivalent, insofar as both natural and designed systems could exhibit 'control' in the form of self-regulation produced through circuits or closed loops of negative feedback" (Clarke 38).

Many of the connections I hope to make between mental life, perception, and behavior on the one hand, and ecology and natural processes on the other, can be expressed in what Clarke and others call a "second-order cybernetic" idiom. First-order cybernetic concepts involve the aforementioned ideas of control, feedback, and homeostasis. Second-order cybernetics is predicated on *recursion* (formal self-reference) and what follows from this: things like identity, autonomy, and ultimately, *cognition*. In effect, second-order cybernetics is a cybernetics of cybernetics (Clarke 84).

Autopoiesis is an example of second-order cybernetic logic. In 1971, biologist and philosopher Francisco Varela and neuroscientist Humberto Maturana coined the term *autopoiesis* (literally, self-making) to refer to the set of circular, self-referential system processes that form the basis of biological life, of which the paradigm case in an individual cell. Thompson says, “The concept of autopoietic organization arose from an attempt to abstract from the molecular processes of the cell the basic form or pattern that remains invariant through any kind of structural change, as long as the cell holds together as a distinct entity” (Thompson 98). Autopoietic theory abstracts the basic patterns of cellular self-organization and re-production—patterns that persist despite ongoing material and energetic flux. A cell is a metabolic network that creates and maintains itself by establishing a semi-permeable membrane or boundary. The permeability of this membrane ensures that cells are energetically open to molecular/chemical inputs from the environment. These inputs, in turn, help sustain the metabolic processes responsible for the maintenance of the membrane, thus ensuring the continued efficacy of the metabolic network.

In these ongoing, interdependent processes of self-organization and reproduction, the metabolic network and the cell boundary together create what in second-order cybernetic terms is called *operational closure*, meaning that the cell sequesters its organizational processes from the chaotic, energetic and thermodynamic “noise” of the environment. Life is characterized by the order it creates and maintains in relation to the thermodynamic disorder of its environment—that ineluctable march towards disorder (entropy) described by the second law of thermodynamics. We must always keep in mind that this operational closure is made possible by the system’s being structurally open to energetic inputs from the environment. This dual (and somewhat counterintuitive) relationship of being operationally closed but structurally open is the pith of second-order cybernetic description.

Together operational closure, structural openness, and the membrane boundary distinguish an *inside* (the systemic unity, or the invariant pattern) that emerges and persists relative to an *outside* (an environmental and chemical milieu). Operational closure emerges as a function of recursion (*formal self-reference*), in which certain features of the environment become salient to the cell *on the basis of* its own organizational unity. In order to maintain itself, a cell must actively engage with and select information (sensory stimuli) from its environment. In effect, it must selectively orient itself towards those features of the environment that will ensure its continued self-making. As systems theorists Bruce Clarke and Mark Hansen tell us, "…to maintain their autopoiesis, (self-referential) systems must remain operationally (or organizationally) closed to information from the environment. On that basis, they can construct their interactions with their environment *as* information" (Clarke and Hansen 9). In other words, it is only on the basis of this sequestered organizational unity that the environment and its features become salient as such. This "on the basis of" implied by the word "recursion" represents the most challenging philosophical sticking point, as it confronts the observer with a causality that, at bottom, is not easily explicable according to normal scientific methods. Clarifying autopoiesis' and autopoietic enactivism's explanation of this process will be the subject of chapters two, three, and four.

One final piece of terminology is needed here: this relationship between a recursively organized cognitive system and its environment is what Maturana and Varela refer to as *structural coupling*. They say, "We speak of structural coupling whenever there is a history of recurrent interactions leading to the structural congruence between two (or more) systems" (*The Tree of Knowledge* 75). Structural coupling denotes the irreducibility of the cognitive system and its environment. It is the totality of system-environment interactions and transformations that give to each their essential features or properties. In this schema, system and environment are wholly interdependent phenomena. Structural coupling represents the systems theoretical (or if you like, second-order cybernetic) foundation for enactivism's broader phenomenological claim that self and world are co-created or *enacted*.

In addition to defining the formal organization of biological life, autopoiesis also implies a behavioral characteristic, one that is coextensive with the organizational, and which issues from this immanent generation of systemic identity and ongoing self-production. This behavioral characteristic is *autonomy*. In their 1992 book *The Tree of Knowledge*, Maturana and Varela assert, "We are *not* proposing that living beings are the only autonomous entities…We *are* proposing that the mechanism that makes living beings autonomous systems is autopoiesis. This characterizes them as autonomous systems" (Maturana and Varela 48). By describing autopoiesis as the *mechanism* of autonomy they mean that autopoiesis is the process by which the property of self-determination manifests.

Given the above descriptions (*recursion*, *operational closure*, *structurally open*, *structural coupling*, and *autonomy*) proponents of autopoiesis argue that "cognition" should be defined as a system's basic information-processing procedures, in which sensation guides action which in turn shapes sensation. In other words, as soon as there exists this ongoing process of formal self-reference, the basic conditions for cognition have been met. *According to autopoietic enactivism, an autopoietic system is inherently a cognitive system, and therefore life is an inherently cognitive phenomenon.* Maturana articulates this foundational cognitive claim in his 1970 essay "The Biology of Cognition":

> A cognitive system is a system whose organization defines a domain of interactions in which it can act with relevance to the maintenance of itself, and the process of cognition is the actual (inductive) acting or behaving in this domain. *Living systems are cognitive systems, and living as a process is a process of cognition*. This statement is valid for all organisms, with and without a nervous system. (Maturana 13)

Here we see an unequivocal unification of life and cognition. A cognitive system can be described as one whose endogenously determined boundaries establish a domain of interactions, and cognitive processes are the behaviors that constitute that domain. Maturana maintains that all processes enacted by living systems are cognitive processes. The fact that a cell can undergo structural and energetic change "without the loss of its identity" (13) makes it a cognitive system. Just because the interactions between system and environment are chemical or physical in nature does not make it any less cognitive. Thompson's autopoietic enactivism is an extension of this foundational synthesis first outlined by Maturana and subsequently taken up in Maturana and Varela's collective venture in "Autopoiesis: The Organization of the Living."

For enactivism, the power of autopoietic theory lies in its ability to abstract the processes, patterns, and inside/outside relations that constitute the organization of *all* living things, everything from ancient prokaryotes to the most cognitively complex mammals, and to define this basic autonomous and informationally-sensitive dynamic as inherently cognitive. When defined in this way, cognition gets baked into the description of biological life from the very start and from the ground up. Thompson's simple and illuminating example is the fact that a bacterium can "sense the concentration of sucrose in its immediate environment and…move itself accordingly" (Thompson 125). In reference to this example enactivist theorist Daniel Hutto, says, "There seems no good reason to rule this out as an instance of cognition or mentality, albeit basic, other than attachment to the idea that true cognition or mentality must involve symbols and concepts" (Hutto 55). I agree. If we can let go of such cognitivist presuppositions, then the myriad upshots of enactivist theory can begin to emerge and do meaningful explanatory work in the world. These cybernetic and systems-theoretical abstractions that define an autopoietic unity represent the pith of Thompson's ambitions to unify mind and life. It represents the theory's greatest affordance, but also, its biggest weakness. As such, I will dedicate three chapters (2,3,4) to exploring the "autopoietic" side of Thompson's enactivist theory.

The next level of embodied description Thompson outlines for us is, in many ways, a recapitulation of the first. Like the phenomenon of structural coupling, *sensorimotor coupling* signifies this autopoietic entity's embeddedness in its environment but scaled up to the level of nervous systems with distinct sense modalities. On this scale, enactivism explicates the various sensorimotor patterns and processes that constitute perception, and by extension, cognition. This is the basic position of Alva Noë, Susan Hurley, and J.K. O'Regan, known as sensorimotor theory of perception, or as Thompson calls it the "dynamic sensorimotor approach" (256). This is primarily a theory of perception, and as such focuses on the distinct sense modalities present in the human nervous system.

In contrast to the sweeping scope of Thompson's *Mind in Life*, Noë zeroes in on these myriad and distinct sensorimotor loops—on how one's body and its capacity for movement and change fundamentally shape perception. Rather than functioning as a kind of representational "mapping" procedure, in which an invariant outside is isomorphically reproduced in the brain, Noë argues that perception operates more like touch: grasping, groping, and enacting the world through perpetual acts of contact (Noë 17). This kind of pragmatic and quotidian engagement with the world becomes folded into what he calls "sensorimotor knowledge." Noë says, "Crucially, the knowledge in question is practical knowledge; it is know-how. To perceive you must be in possession of *sensorimotor bodily skill*" (Noë 11). The sensorimotor is the scale around which many enactivist works tend to gravitate, and consequently, where the enactivist account has been most salient for neurophenomenology and cognitive narratology. I will return to these narratological dimensions in the last chapter of the dissertation in order to demonstrate how enactivism is already being mobilized in humanities—and specifically literary contexts. For now, I simply want to flag that in much of the 4E literature I have encountered, theorists gravitate towards Noë and the sensorimotor scale (a) because it avoids the sweeping (and sometimes problematic) scope of Thompson's autopoietic enactivism and (b) because it offers a more concrete repertoire of analytical tools for understanding perception and its objects (e.g., narratives). Furthermore, given the overwhelming focus on Noë's sensorimotor scale from within literary studies, I believe my project can speak to an outstanding gap in literary theory and ecocriticism with respect to the far-reaching environmental and ecological implications at work in Thompson's autopoietic enactivism.

The final scale in Thompson's tripartite breakdown is *intersubjective interaction*, what Thompson calls "the cognition and affectively charged experience of self and other" (Thompson 243). This scale acknowledges our inherently social and ethical nature, under which we can include language in its most fully developed forms. This is the realm of semiotic systems: expression, discourse, and artistic creation, and therefore the realm in which Levinas' ethics—and by extension most poststructuralist approaches to subjectivity—implicitly operate.

Given the complexity of these ideas, as well as the interdisciplinary scope of the chapters to follow, this extended introduction to Thompson's tripartite breakdown of embodied cognition is essential. My project will spend time investigating mental and biological processes at each of these scales, but the controlling idea of this dissertation can be found in the foundational level of "self-regulation" described by autopoiesis. This is the idea *that living systems and cognitive systems are equivalent.* In the words of Humberto Maturana: "Living systems are cognitive systems, and living as a process is a process of cognition" (Maturana 13). The language of *process* here is paramount. Cognition and living are activities—ongoing acts and actions which constitute the organismic and mental "self" and its world. This idea can liberate us from anthropocentric thinking by showing us how knowledge and all related mental phenomena are not locked away behind the wall of the

Cartesian divide—in which inscrutable and immaterial minds stand in opposition to material and self-evident bodies—but are part of the very weave of life itself. I will affirm this equivalence between life and mind at every scale of living organization—from the bacterial to the human—and I will trace this equivalence in every biological, neuroscientific, and philosophical theory to follow.

By tracing a continuous line between these three levels of embodied experience that Thompson outlines for us—self-regulation, sensorimotor coupling, and intersubjectivity—I believe that his autopoietic enactivism offers the environmental humanities a theory of mind that can obviate many of the prejudices that emerge when we focus our gaze solely on the level of intersubjective interaction and its determination by spoken language. Autopoietic enactivism offers the environmental humanities a powerful framework for putting aside such prejudices and uncovering the cognition that operates at every scale of living organization. I argue that autopoietic enactivism can help us expand our ethical gaze by training us to see the interiority and purposiveness which the term "cognition" implies. In other words, that—countering Levinas's anthropocentric exceptionalism—we can learn to see "the face" that exists within every living being, no matter how small.

Enactivism, Literary Studies, and the Environmental Humanities:

"Intersubjective interaction" facilitated by humans' innate faculties for language and speech form the core of social constructionist logic. This is the idea that reality, be it ontological or epistemic, is a product of social and cultural interactions. Social constructionism came to prominence with the so-called "linguistic turn" that dominated the humanities from the postwar period to the mid-1990s and was instigated chiefly by two strands of theory coming from the continent, particularly France. These were structuralism, and shortly thereafter, poststructuralism/deconstruction. Despite the explicit overlap that occurs between post/structuralism and enactivism at this intersubjective scale of description, it is clear to me that there exists an enormous *disconnect* between these two approaches to mental life. Perhaps the biggest point of contention can be located in the idea of autonomy. For autopoietic enactivism, autonomy is an essential property of a cognitive/living system and one which can explain the purposive and teleological (end-directed) organization we observe therein. In contrast for post/structuralism autonomy is largely antithetical to its basic understanding of language and subject formation. Let us explore why this is the case.

Despite its disciplinary variations, structuralism abstracted two essential features of semiotic systems: (1) signs are arbitrary—there is no necessary connection between a symbol and that which it represents, and (2) meaning is a product of differential relations—it maintains no positive or univocal value. Together these led to the recognition that language does not refer to the world or to experience, but instead to its own systemic arrangement which stands apart from it. In general terms, structuralists supposed they could occupy a context-independent space from which to analyze linguistic and cultural systems. The pith of poststructuralist critique is that no such position exists. In the words of deconstruction, we are *always already* enmeshed in the systems and structures we purport to study.

One of the chief consequences of the linguistic turn was that most phenomena became subject and subordinated to the structures and processes of language and signification. Whereas linguists constrain the study of language to issues of phonology, morphology, syntax, or concrete historical development; proponents of structuralism and poststructuralism tended to use "language" as a kind of shorthand for anything that could signify and bear meaning, even extending this term to the psyche, and to various forms of cultural production. In this extended sense of language *meaning* becomes a fraught and highly unstable issue—a site of conflict, repression, and ideological determination. In poststructuralism, what is often at stake is not so much the particular meaning of this or that object or utterance, but the positionality and power that is revealed in the construction of that meaning.

To be sure, many theorists we could group under the headings of structuralism or poststructuralism are also interested in pre-linguistic, perceptual, phenomenological, libidinal, and affective *experiences*.

For example, Heidegger's existential analytic of Dasein, its being-towards-death, and the striving for authenticity that results from it; Levinas' descriptions of the face of the Other, its rupture of a totalizing gaze, and the affective sense of responsibility that flows from it. These are structures of being and feeling that have not yet been crystalized in, or reduced to, an object of speech, categories, or signifiers; moreover, both utilize the phenomenological methods on which Thompson's enactivism relies. We could also include in this list psychoanalyst Jacques Lacan, who saw psychoanalysis as a way to sublimate instinctual drives such as eros and aggression, as well as feminist theorist Luce Irigaray who explored a phenomenology of the female body outside the bounds of phallocentrism. From a purely rhetorical standpoint, it is tempting to include Jacques Derrida's work which is notable for its often embodied, reproductive, and sexual conceits. His famous 1966 lecture "Structure, Sign, and Play" which introduced his Deconstruction to the American academic scene—and which I will explicate at length in chapter four—is full of embodied imagery such "conception," "gestation," "labor," and "childbearing". Even Michel Foucault, whose historiographic analyses of the body is almost *entirely* determined by discursive matrices, requires, in principle, repositories of semiotically unmediated embodied experience from which these networks of epistemic power can emerge—here I am thinking specifically of the chapter "Docile Bodies" in *Discipline and Punish* where he describes an emerging "'micro-physics' of power" (Foucault 139)—i.e., the myriad ways in which the modern episteme exercised itself in habits and regimented behaviors in the school, the barracks, the hospital etc.

Despite these varied expressions of materiality and the body, structuralism, poststructuralism, and deconstruction were united by the idea of what Catherine Belsey calls the "social fact" of the signifier (Belsey 39). Their commitment to the abstract and virtual systematizing of structuralism meant that the social fact predominated, and so any forays into embodied experience, no matter how inspired, found itself hounded by "discourse." This inability to access the material body represents one of the biggest failings of poststructuralist discourse: (to use Derrida's famous phrase) the body was *always already* one more sign captured by a discursive matrix. In sum, this post/structuralist paradigm allowed for deep insights into the nature of psychology, sociology, and epistemology, but ultimately it made "language" do too much.

Perhaps the most pervasive and lasting impact of poststructuralist theory involves the act of reading itself, exemplified in the dichotomy symptomatic/surface reading. "Symptomatic" reading was first proposed by Louis Althusser in *Reading Capital* and was later taken up by Frederic Jameson in *The Political Unconscious*. In brief, this hermeneutic approach assumes that texts are constituted around blind spots, lacunae, or unconscious elements that are constitutive of the conscious meaning that appears on the text's surface. Poststructuralism, despite the diversity of its investments and interventions—be they Marxist, feminist, psychoanalytic, linguistic etc.—is united by an overarching skepticism of common-sense interpretation of texts and their surface elements. Sharon Marcus and Stephen Best's article "Surface Reading: An Introduction" (2009) was their response (primarily to Jameson) but also to a whole host of reading practices that had accrued during the linguistic turn. Eve Sedgwick's oft quoted dichotomy of paranoid/reparative reading can also be read as part and parcel of this larger debate.

By the 1990s, people were beginning to tire of critical theory and its general pretense that self, subjectivity, and culture could be thoroughly analyzed and explained on these linguistic grounds. In other words, they tired of language and ideology as an explanatory framework. This reaction to the linguistic turn resulted in at least three distinct branches of literary inquiry.

The first is a broader genealogy that can be traced from trauma theory up to contemporary theories of affect. Trauma studies took off in the 1990s for a number of reasons, but a decisive one was that it concretized an important aspect of poststructuralism: the constitutive aporias/lacks it posited at the center of signification.

Trauma begins in an inciting event, one which is so powerful and disturbing that it exceeds the psyche's capacity to process. In effect, one lacks the interpretive framework—i.e., the experiences, social constructions, ideas—to properly cognize and understand the event. Due to its power, this trauma inevitably manifests itself in a slew of negative psychological states and neuroses, and it is only until much later that one becomes consciously aware of the trauma *as trauma*. In other words, trauma manifests between two poles: an event, and an interpretive relation towards that event. These are the boundary conditions in which "trauma" manifests, and within which trauma theorists and medical professionals continue to wrestle.

This episode in literary studies had some important consequences. For one, many theorists were delighted to discover that they could point to trauma as an empirically verifiable expression of poststructuralist theory—already a move beyond linguistic solipsism. More importantly, trauma theory created a clear embodied undertow. Interpretive (epistemological) approaches, many argued, risk ceding too much to the ambiguities of signification and social construction and thereby stripping the concept of the very real emotional-psychological *experience* the term seeks to address. In trauma theory there emerged a clear desire to situate the body and its lived experiences as a non-linguistic repository of sociopolitical contexts and meanings. This embodied undertow can be clearly traced into contemporary theories of affect. It is also important we note that "trauma" continues to do a lot of critical and interpretive work, both as a diagnostic (i.e., medical) and interpretive tool, and it is also no longer confined solely to individuals but often used to speak to the history and collective experience of communities affected by things as diverse as structural racism and natural disasters.

The second branch is marked by a turn towards identity: if difference proliferates ad infinitum—as deconstruction argued—then one way of responding is to double down on the signifiers that mark such difference. This position recognizes the socially constructed nature of identity, but rather than linger on its esoteric epistemological implications, identity interpretations try to fix these signifiers in a well-defined sociological milieu, and in doing so, foreground the concrete political and material realities that issue from them: for example, issues of race and gender (and their intersection), indigenous and decolonial approaches, and disabilities studies. These identity-based approaches strike me the most prominent of our contemporary analytical frameworks, important work to be sure.[6]

The third and final reaction to the linguistic turn—and the most important for the present work—was also a return to materiality, but by way of nature and the environment. While environmentalism and environmental writing had been prominent since the early postwar period, an environmental approach to literary studies did not gain momentum

[6] It is important to point out that *all* of these identity-based frameworks have had their embodied proponents—theorists who attempt to reconcile the social construction of identity with its embodied underpinnings, and no doubt there are many ways in which 4E approaches can be put into productive conversation with these frameworks. The one that comes most powerfully to mind is disability studies. If one's body is disabled, or one's brain neurodivergent, then it stands to reason that the ways one will couple to the world cannot be explained without recourse to this identity. Nevertheless, I do not think that this intuition obtains in the autopoietic enactivist approach I adopt (à la Thompson) and the sensorimotor approach I adopt (à la Noë), both of whom intervene in cognition in the most general sensorimotor, and systems-based terms. Both Thompson and Noë purport to explain *universal* features of system-environment couplings, that is, applicable in all cognitive agents, regardless of social, political, ethnic, or gender constructions, including neurodivergence and disability. I will not do too much to engage such an objection here, but I am confident that over the course of this dissertation, that any misgivings the reader may have with these statements will be allayed.

until the mid-eighties and early nineties. In 1992 the Association for the Study of Literature and Environment (ASLE) was formed, and by "1993…ecological literary study had emerged as a recognizable critical school" (Glotfelty xvii).

As the hegemony of structuralist and poststructuralist paradigms has waned and given way to more identity-based and eco-critical investments in contemporary literary studies, alternative semiotic theories such as that of Charles Sanders Peirce and Maurice Merleau-Ponty are becoming increasingly commonplace. In chapter four I will examine Peirce's semiotics as the most promising alternative to structuralism. It is also worth noting that Merleau-Ponty's work forms the basis of much of Thompson's own phenomenology.

One notable work of ecocriticism that engages with the phenomenology of Merleau-Ponty is Louise Westling's 2014 book *Logos of the Living World: Merleau-Ponty, Animals, and Language* which attempts to dismantle many of the anthropocentric tendencies inherited from these structuralist paradigms. In her introduction she says:

> The increasing body of archaeology and scientific studies of animal behavior and cognition shows that humans cannot be considered separate from other living creatures with whom we coevolved and shared ancestry, genetic makeup, and morphology. Thus, as Balkan German ethologist Jakob von Uexküll explained with his Umwelt theory, we overlap with the others in many ways and can understand them to some degree, as they can us, because of a long shared past and similar physical qualities, abilities, and habits. (Westling 4).

Westling uses Merleau-Ponty's oeuvre as theoretical backbone, and pairs it with an eclectic mixture of primary literary texts—everything from ancient works such as *The Epic of Gilgamesh* and Euripides' *Bakkhai*, to the modernist poetry of W.H. Auden, as well as contemporary novels such as *Life of Pi*—and throughout the book she argues for a conception of language and logos rooted in embodiment.

A good example of contemporary work that engages with Peirce's semiotics is Eduardo Kohn's 2013 book *How Forests Think: Toward and Anthropology Beyond the Human*. Like Westling's work, Kohn tries to confront anthropocentric views of language via phenomenological and embodied methodologies. He addresses his provocative and somewhat counterintuitive title as follows:

> Is such an exploration [an anthropology beyond the human] possible? Or do the all-too-human contexts in which we live bar us from such an endeavor? Are we forever trapped inside our linguistically and culturally mediated ways of thinking? My answer is no: a more complete understanding of representation, which can account for the ways in which that exceptionally human kind of semiosis grows out of and is constantly in interplay with other kinds of more widely distributed representational modalities, can show us a more productive and analytically robust way out of this persistent dualism. (Kohn 41)

These recent works are representative of a larger embodied and phenomenological shift occurring in literary studies and ecocriticism. They also speak directly to my overall framing for this project: to understand how "cognition"—as a systems-based and embodied phenomenon—can help us rethink anthropocentric notions of "subjectivity."

But unlike Kohn whose work is a mixture of ethnography and autotheory; and Westling whose work is largely ecocritical and literary; I have attempted to stick closer to the scientific and philsophical source material rather than extend it into a specific literary or ethnographic domain.

Underwriting this new "ecocriticism" in literary studies, is a whole host of theoretical and philosophical attempts to come to terms with the real in non-constructivist terms, a philosophical movement that became known as speculative realism. Speculative realism is defined by its ambition to study the world not—in what Quentin Meillasoux has called "correlationist" terms—as the correlation of thought and being, but in terms of networks of interaction and relations between objects and actors.

Bruno Latour's Actor Network Theory (ANT), for example, adopts a network model devoid of organizational levels or hierarchies and instead traces connections of distributed agency, or *actants*. Like poststructuralism, ANT is inherently anti-essentialist, but by focusing on relationships and networks, Latour's approach attempts to straddle the line between material and semiotic modes of being.

Graham Harman's object-oriented ontology is another important instance of speculative realism based around a rereading of Heidegger's tool-analysis in *Being and Time*, specifically the concept *Zuhandenheit*, or readiness-to-hand. For Heidegger, objects in the world become ready-to-hand when Dasein encounters them and puts them to some immediate and situated use (e.g., a hammer). Readiness-to-hand constitutes Dasein's ordinary and everyday relationship with the world—our default setting so to speak—because we habitually encounter people, places, and things via interpretive fore-structures that are grounded in immediate and pragmatic uses.[7] In contrast, Harman argues that "readiness-to-hand (*Zuhandenheit*) refers to objects insofar as they withdraw from human view into a dark subterranean reality that never becomes present to practical action any more than it does to theoretical awareness" (Harman 1), and that these objects not only withdraw from human being, "but from each other as well" (Harman 2). This means that "contrary to the dominant assumption of philosophy since Kant, the true chasm in ontology lies not between humans and the world, but between objects and relations" (Harman 2). In other words, the site of ontology is not to be found in the relationship between humans and their world, but between objects themselves—a radically non-correlationist claim.

While I share this realist impulse and am piqued by the general attempt to move beyond correlational forms of knowing, I am skeptical that this type of discourse can be sustained in any lasting or meaningful way. For one, these analyses can tend towards excessive abstraction. In his book *The Ecological Thought*, Timothy Morton offers a good critique of this more generalized speculative realist (i.e., non-correlationist) impulse by way of "posthumanism." They say:

> Posthumanism (a current trend in the humanities) too glibly combines (1) a deconstruction of humanness—and animal-ness, and life form-ness—into sets of machine-like algorithmic processes; and (2) decidedly nonreductionist, holistic,

---

[7] Heidegger distinguishes the ready-to-hand from the "present-at-hand," which is the contemplation of a thing once it has been removed—insofar as it is possible—from all such immediate contexts. Accordingly, the present-at-hand represents a very abstract and derivative mode of being because it takes a lot of mental gymnastics to contemplate a thing removed from its myriad fore-structures.

> quasi-mystical systems theory…Posthumanism seems suspiciously keen to delete the paradigm of humanness like a bad draft… (Morton 113)

Though Morton does not name it outright, by invoking computational and machine processes they seem to be aligning posthumanism with a much more general skepticism of what makes organisms unique, namely their purposive, teleological, and autonomous character. Such skepticism strikes me as a holdover from early poststructuralist tendencies that equate autonomy with the larger enlightenment-humanist project. I think Morton rightly impugns the posthumanist impulse that would (a) reduce or redistribute these essential characteristics of organismic life into a set of algorithmic processes or (b) swing too hard in the opposite direction by adopting a facile approach to systems theory and its understanding of irreducible wholeness. Ironically, I believe that one of the clearest exemplars of this quasi-mystical posthumanist tendency is none other than Morton and their concepts such as *the mesh* or *dark ecology*, which offer boundless energy and pathos but little, I believe, in the way of concrete analytical tools.

An important early work aimed at critiquing the advent of posthumanist discourse was Katherine Hayles' *How We Became Posthuman*. Her breakdown of the term "posthuman" runs as follows. First, she argues that the posthuman view "privileges informational pattern over material instantiation, so that embodiment in a biological substrate is seen as an accident of history rather than an inevitability of life" (Hayles 2). Second, it views consciousness as an epiphenomenon—an effect with no causal efficacy of its own. Third, it views the body as just another "prosthesis," something we learn to manipulate but which is not constitutive of the process of manipulation as such (3). Most importantly for Hayles, posthumanism "configures human being so that it can be seamlessly articulated with intelligent machines" (3). In Hayles analysis, the posthuman human evokes equal parts pleasure and terror because it challenges the hegemony of the liberal humanist subject that has long occupied the center of Western philosophy and epistemology. One of her main theses is that "the erasure of embodiment is a feature common to both the liberal humanist and the cybernetic posthuman" (4), and thus one of her primary goals is "to keep disembodiment from being rewritten, once again, into prevailing concepts of subjectivity" (5). In these essential features, my project is in complete agreement with Hayles'. By ignoring or downplaying how agency, intention, and interpretation occupy and proceed from specific embodied contexts, posthumanist theories run the risk of reproducing male, European, and anthropocentric subjects in all their unethicality—a reproduction made all the more insidious because of how such theories appropriate the language and ethos of science.

Thus, while issues of embodiment, cognition, and mind/body dualism have occupied an increasingly important part of the humanities, at least since the publication of Hayles' work, my project takes a different tack. For most of this dissertation, I adopt and analyze the language of science, specifically neuroscience, biology, and thermodynamics. But I hope to obviate the above charge of cultural erasure by focusing on only the most granular accounts of agency, intention, and subjectivity. The bulk of this dissertation keeps to the level of ideal cognitive systems and to the phenomenological features that emerge from them.[8] Stated simply, I focus on bacterial bodies, *not* racial or gendered bodies. In my final two chapters I emerge from these systems theoretical depths into the realm of human subjectivity once again, but I am confident that my grounding in these fundamental theories of cognition will allow me to intervene in issues of inter/subjectivity, language, and interpretation in ways that precede the emergence of things like race, gender, class etc.

My project also differs from someone like Hayles' in that it centers heavily on the phenomenological dimensions entailed by these theories of cognition. In her attempt to explain the emergence of a novel cultural phenomenon (posthumanism), Hayles relies on literature as an important repository of cultural changes—as a means of revealing "the complex cultural, social, and representational issues tied up with conceptual shifts and technological innovations" (Hayles 24). In contrast, my focus on phenomenology means that I understand literature in the context of a more general type of cognitive and narrative *activity*. To the extent that I talk about literature at all, it is from the standpoint of process, form, and aesthetics—literature as embodied experience per se, not a story *about* embodiment.

These distinctions lead nicely into the disciplinary framing for this project. While these aforementioned non-correlational, ecological, systems, and network-based theories have exerted a powerful influence on contemporary literature—specifically ecocriticism—the appellation "ecocriticism" is ultimately too constraining for my present purposes.

---

[8] E.g., the theory of autopoiesis developed by Maturana and Varela (and extended by Thomspon), and the theory of autogenesis developed by Terrence Deacon in his 2011 book *Incomplete Nature*.

Ecocriticism is wedded explicitly to literature.[9] While my last chapter will engage with narrative modes of cognition, phenomenological approaches to narrative, and enactivism's place among them, this project engages in no literary readings. As previously stated, I am interested in understanding the deeper theoretical aspects of "cognition" and how these insights can contribute to humanities disciplines more generally and not just in terms of literary engagement.

In contrast to "ecocriticism," I situate this project in the context of the *environmental humanities*: a larger, interdisciplinary framing that believes "the humanities have a crucial role to play in understanding and in solving environmental problems," (Emmett and Nye 2) and that "reconceives the relationship between scientific and technical disciplines and the humanities, which are essential to understanding and resolving dilemmas that have been created by industrial society" (Emmett and Nye 4). Accordingly, ecocriticism occupies a subset of the larger interdisciplinary scope entailed by the appellation "environmental humanities," the origins of which began in the 1960s and 1970s with the emergence of the environmental crisis, and whose "critical agenda…emerged in response to a multi-pronged crisis of ecology, economy, politics, and epistemology" (Emmett and Nye 8).

Early works of the environmentalist movement, such as Rachel Carson's *Silent Spring*, and Barry Commoner's *The Closing Circle* exemplify this environmental humanist

[9] Cheryl Glotfelty makes its purview clear: "ecocriticism is the study of the relationship between literature and the physical environment" (xviii); and later, "Ecocriticism takes as its subject the interconnections between nature and culture, specifically the cultural artifacts of language and literature" (xix).

goal of bridging scientific facts and societal and cultural concerns. Carson's landmark book traces the destructive path of the synthetic pesticide DDT through forest, river, soil, sky, and city. Her exposition moves smoothly between the human, the biological, and the industrial and chemical. It was the first book to unabashedly challenge the rosy picture of postwar American abundance characterized by a pervasive faith in scientific progress and in humanity's capacity to control nature. At every point in her exposition Carson's remedy is clear: we must humble ourselves before nature and learn to see the larger ecological systems that we inhabit.[10]

Barry Commoner—who was quite literally "the face" of the first Earth Day in 1970 when he landed on the cover of Time magazine—is even more expressly committed to analyzing human and environmental systems. The very title of his book *The Closing Circle* is meant to invoke the systemic and circular feedback mechanisms of a cybernetic system (32) and lays the foundation for his famous "four laws of ecology." Like Carson, Commoner uses this systems-theoretical logic to move between different scales and networks of organization, but what I find so striking about Commoner's work is his sustained focus on

[10] As we think about this larger disciplinary framing that is the "environmental humanities," I want to point out that even though Carson is speaking to a general readership she does not shy away from technical discussions, particularly around biochemistry. For example, the chapter "Through a Narrow Window" begins by describing the process of energy production within the mitochondria of cells. She describes in minute and technical detail for her readers how "ATP is the universal currency of energy—found in all organisms from microbes to man" (Carson 202), how its production depends on an oxidation cycle she analogizes to a wheel that is always turning in the service of life's metabolic processes, and that "The crowbar to wreck the wheels of oxidation can be supplied by any of a number of chemicals commonly used as pesticides" (Carson 204). This is one of my favorite sections in the book, and also one of the best examples how Carson uses suggestive imagery and simple analogies alongside highly technical descriptions from the sciences as a means of moving between vast ecological scales, starting at the molecular level and then expanding its implication to all members of the biota.

the linear, input-output patterns of economic and social organization and how these come into conflict with nature's underlying circular dynamics. He concludes the book as follows: "Human beings have broken out of the circle of life, driven not by biological need, but by the social organization which they have devised to 'conquer' nature…The end result is the environmental crisis, a crisis of survival. Once more, to survive, we must close the circle" (Commoner 288). Throughout the book, his refrain is that because the environmental crisis is a product of social organization, its remedy will also be found in the social and political sphere.

I pause on these two works because, to my mind, they perfectly embody the goals of the environmental humanities. Even though Thompson, at times, ventures into rather esoteric philosophical and scientific waters, his overarching ambition to unite mental and physical phenomenon using a systems perspective makes his autopoietic enactivism part of this larger genealogy of the environmental humanities, and so belongs alongside thinkers such as Carson and Commoner. This project will focus primarily on Thompson's theory of autopoietic enactivism presented in *Mind in Life*, but I will also spend a significant portion of this dissertation engaging with Terrence Deacon's *Incomplete Nature*—another work of systems theory that, unlike Thompson, approaches issues of mind/matter dualism by cleaving more closely to scientific explanations (specifically thermodynamics), as opposed to Thompson's recourse to phenomenological ones.

This leads me to the other controlling idea at work in this dissertation: *that dynamical systems theories are narrowing the explanatory gap between living and physical processes, as well as between the opposing disciplinary and philosophical "cultures" of* *materialism and holism.* Both life and mind confront the observer with teleological (i.e., immaterial and end-directed) organization that cannot be neatly reduced to the analytical and reductionist frameworks that prevail in physics and chemistry. The humanities take this purposiveness and end-directedness for granted—as a fundamental axiom of mental life; while mainstream science has tended to eschew the problems posed by teleology by either (a) defining it out of existence or (b) burying it in some deeper level of explanation. Together Thompson's autopoietic enactivism and Deacon's emergent dynamics can help us to reenvision the gap between the hard and human sciences via the common denominator of a systems theoretical perspective.

The study of life and mind confront us with a causality that is not easily explained. My goal is to use these theories to present a multipronged, systems-based critique of traditional linguistic, social constructionist, and cognitive scientific frameworks that separate human subjectivity from the organismic and ecological systems of the natural world. I believe systems theory is uniquely capable of tracing the interconnections between physical, embodied, and mental processes, as well as accounting for the *emergent*, i.e., inherently non-reducible and holistic aspects of life and mind. It is also philosophically protean. Its conceptual apparatus of *parts*, *wholes*, *matter*, *form*, *organization*, *structure* and *change* feels equally at home in ancient works like Aristotle's *Metaphysics* as it does in the autopoietic theory of Maturana and Varela. Because of this ability to move between philosophical modes, I also believe it can obviate ideological dog whistles (particularly around the notion of *autonomy*) which, as we have already seen in our brief introduction to Levinas, is an easy target for the hermeneutics of suspicion and of all poststructuralist and postmodern philosophical theories skeptical of its enlightenment-humanist antecedents.

Conclusion and Chapter Overview:

Life and its activities—thinking, feeling, desiring, willing, intending, behaving—are phenomena that cannot be neatly reduced to a mechanistic worldview. An understanding of the brain, its structures, and its cellular and electrochemical mechanisms has yet to provide a thoroughgoing account as to why goal-directed behaviors occur, that *for the sake of which* Aristotle first enshrined in the word *telos*. Statements of mechanical and material fact can only go so far in explaining phenomena constituted by non-material purposes—those absent end-states toward which living things ineluctably grow and develop. Without question, subjective conscious experience depends on physical processes in the brain: there is complete scholarly consensus on this point. This means that if the brain dies, so too does conscious experience. I think that most rational people would, if pressed, concede these points. Yet even in the absence of a religious belief system there remains a powerful and pervasive tendency to make exceptions for human consciousness and to cling to dualistic thinking. These dualistic prejudices regarding human consciousness and our cognitive capacities strike me as one of the most deep-seated and pervasive features of anthropocentrism.

Until very recently scientific materialism had very little to say about the origins of mind intention, consciousness, and by extension the values and meanings that define a life. This schism between first-personal experience and third-personal knowledge brought about by scientific rationalism has had huge sociocultural consequences. Terrence Deacon frames his book *Incomplete Nature*, at least in part, as an attempt to suture this schism. In what seems to be the *only* overt attempt at cultural criticism Deacon makes in its nearly 550 pages, he says, "No wonder the all-pervasive success of the sciences in the last century has been paralleled by a rebirth of fundamentalist faith and a deep distrust of the secular determination of human values" (Deacon 12). Deacon published these words over a decade ago, and they feel more relevant than ever. Ultimately, I believe that the vision of autopoietic enactivism put forward by Thompson in *Mind in Life* offers a coherent and convincing explication of "cognition," one capable of overcoming the prejudices surrounding notions of subjectivity and language leftover from the last half century and their concomitant anthropocentric tendency to de-face and efface the natural world.

Chapter one will provide an overview of cognitive science as an interdisciplinary research program that began in the 1950s. It will outline its main methodological investments—computationalism and representationalism—and demonstrate its vexed, and often ambiguous relationship to these explanatory frameworks and their two guiding metaphors: the computer and the map. Through this exposition it will become clear (a) how enactivism came to exist in its current form, and (b) why it provides important qualifications and critiques of the orthodox positions at work in the field of cognitive science.

Chapter two begins with a recent historical overview of biological and systems theoretical attempts to explain the perennial philosophical and methodological problems posed by purposive and end-directed organization in living systems—what Capra and Luisi analogize as the back-and-forth swing of a pendulum between the mechanism and reductionism of the hard sciences and the holism of systems theory. The body of this chapter deals with a few key positions on this pendulum's oscillation: neo-Darwinism, Shannon information theory, organismic biology, and of course, autopoiesis. I identify several shortcomings in Maturana and Varela's original work, "Autopoiesis: The Organization of the Living" with respect to this problem of purposive/teleological organization. I end this chapter with a close-reading of what Thompson calls "the body-body problem," which is his attempt to redefine the explanatory gap that emerges in the analysis of living and cognitive systems. In the course of his explication Thompson partitions the body-body problem into two discrete "tasks," one ostensibly materialist (scientific) and the other ostensibly phenomenological (enactivist). I use this distinction as the organizational schema for chapters three and four.

Chapter three addresses the "scientific" i.e., materialist dimension to the body-body problem, that is, how to explain the mind via mind-independent descriptive means. I use Terrence Deacon's 2011 book *Incomplete Nature*, and the novel approach to systems theory he develops therein, *emergent dynamics*, as a way to (a) supplement what I perceive to be a lack in Maturana and Varela's original theorizing of autopoiesis regarding the notions of emergence and autonomy, and (b) reintroduce teleology into scientific-materialist descriptions. These two payoffs are interconnected. *Emergence* is the crux of the mind/life explanatory gap: how to precisely account for the transition from basic thermodynamic systems into the teleological organization of living and cognitive systems. Deacon provides us with novel and unambiguous concepts for understanding the nature of an emergent transition, as well as a precise, stepwise account how the autonomous character of cognitive systems could have emerged from underlying thermodynamic processes. This is an account which Thompson's autopoietic enactivism lacks, and which I argue contributes to its dubious reception in the field of cognitive science.

Chapter four will build off Deacon's novel methodological approach to emergence and teleology by situating Thompson's autopoietic enactivism as its phenomenological counterpart. I begin by describing the epistemological implications that follow from a cognitive system/environment coupling and then unpack Thompson's argument for why autopoiesis' cognitive limit case—a single-celled bacterium—not only evinces the rudiments of cognition, but the rudiments of self and interiority; in a word *sentience*. The goal in this chapter is (a) to show how enactivism can be understood as a philosophical counterpart to Deacon's emergent dynamics and (b) to throw a wrench in anthropocentrism of every stripe by demonstrating just how much human minds have in common with bacterial bodies.

Chapter five marks our ascent out of the cognitive scientific, biological, and thermodynamic depths and back into the realm of language. In the first half of the chapter, I demonstrate the systems thinking that was already at work in structuralist and poststructuralist theories in order to show how this contemporary approach to mind and life has, in many ways, also been a key feature of the discipline of literary studies for the past century. In the second half of the chapter, I argue that enactivism provides literary scholarship a way to ground the linguistically determined subject (and its concomitant hermeneutics of suspicion) in an embodied autonomy that underlies all language.

Chapter six examines the history of narratology, hermeneutics, reader-response theory and formalist aesthetics. In the first half of the chapter, I identify the increasingly disembodied, representational, and pictorial ways in which these theories conceived of narrative engagement and literary interpretation—epitomized in the notion of a literary "gestalt," and I argue that Alva Noë's dynamic sensorimotor account of perception—in which perceptual content is enacted as a function of the body and its capacity for movement—can help us reconceive the literary gestalt on embodied grounds. I end the chapter by discussing the Russian Formalist notion of defamiliarization and argue that enactivism updates this approach via neuroscientific means. As such, it offers literary studies an embodied and ecologically embedded formalist approach to interpretation.

Chapter One:

Cognitive Science and the Evolution of Enactivism

> But that there were natural causes to all these things I am willing to concede, for the resources of nature are infinite apparently. It was I who was not natural enough to enter into that order of things, and appreciate its niceties.
>
> —Samuel Beckett, *Molloy*

A first step towards seeing the face of nature is to reaffirm a deeper sense of our own. Levinas' ethics, realized in the encounter with another cognitive being that faces me, is a pervasive feature of everyday life. Every day we are confronted by the experiences and lifeworlds of others and given a choice to face them—to respond. Whether it is something highly mediated, like the ongoing violence between Israel and Palestine, or highly intimate, like the needs of a child or a spouse—we are not closed, psychologically-immune systems. We are perpetually called beyond our selves and asked to recognize the interiority, autonomy, and suffering of others.

This capacity to "see" and be affected by the faces of others—in a word, *ethics*—is subject to ongoing processes of concealing and uncovering. I know personally, for example, that being in a loving relationship has enabled me to extend my circle of care—my ethical vision—in ways I would never have been able to otherwise. I uncover and cultivate this sensibility in my most intimate and immediate interpersonal sphere and then try to bring it to fruition in an ever-expanding circle. Conversely, I can sense that my role as a consumer, for example, my position as a tiny node in a sprawling and byzantine global food system built on profit, exploitation, and factory farming functions as a mechanism of concealment. These systems obscure the full consequences of the actions and behaviors I take to feed myself, and the result is that my ethical vision becomes hazier.

These phenomenological processes of concealing and uncovering are at work in all aspects of one's life. The presuppositions and acquired structures of meaning that I bring to bear on an object determine my cognition of that object: this is a phenomenological axiom. And just as I am a consumer in a massive food system, so am I a consumer and inheritor of intellectual history, one that shapes my capacity to see the objects and faces of those before my gaze. With this in mind, this chapter confronts a decisive development in intellectual history, one that, I believe, has contributed to a profound sense of alienation from ourselves, and by extension, our fellows. It is an alienation made all the more illegible because of the equally profound advances this development has enabled.

This chapter follows the rise of cognitive science in the 1950s and the resulting "cognitive revolution," whose defining feature was the mind-as-computer metaphor. This paradigm shift has shaped our collective understanding of mental life. On the one hand, it laid a new foundation for rigorous, materialist explorations and explanations of the brain, but it also bracketed (or concealed) from consideration the deeper philosophical—and I would argue, moral-psychological—problems regarding inner experience, intention, agency, and "self." After all, if the brain is little more than a computer inside our heads (a piece of hardware), and if mind is little more than programming (software), then where does that leave our humanity? How can we hope to encounter the faces of other cognitive beings if we have been alienated from our own?

Here I think it is worth reiterating Terrence Deacon's words in the introduction: "No wonder the all-pervasive success of the sciences in the last century has been paralleled by a rebirth of fundamentalist faith and a deep distrust of the secular determination of human values" (Deacon 12). To be sure, the cognitive revolution and its technological advances represent a massive intellectual achievement; a huge step forward in our ability to observe, name, diagnose, heal, and explain the brain. But as this chapter will demonstrate, it also helped *explain away* many of the defining, life-affirming features of mental life—namely, qualitative first-personal experience and a deeper, existential and spiritual sense of our capacity for self-determination. This shift in our collective thinking is not the end result of orchestrated and malign forces but the slow and steady accretion of practices, perspectives, institutional structures, and disciplinary framings.

Therefore, this chapter will lay an historical and conceptual foundation for those to follow. We will examine how this incipient research program of the 1950s explained inner experience by recourse to the burgeoning insights of computer science, often burying the problematic of interiority in ever-deeper layers of description. This is a procedure that will echo throughout the remaining chapters, at different scales of description and in different disciplinary contexts.

This chapter begins with an historical overview of the cognitive turn, its key stages (cognitivism, connectionism, embodied dynamicism) and its underlying philosophical rationales (computationalism, representationalism, functionalism, cybernetics), before examining enactivism's emergence from among them. As such, this chapter is less concerned with individual texts and more concerned with larger disciplinary and historical trends. Throughout the chapter I identify these larger historical and philosophical inflection points in which novel trends emerged. Accordingly, much of the historical information is pulled from my primary sources—Thompson and Deacon—as well as Mark Cain's book *The Philosophy of Cognitive Science*. As I transition into a closer exposition of enactivism and 4E theory from out of the discourse of embodied dynamicism, I turn to key works within the 4E field—particularly Alva Noë's *Action in Perception*. I conclude this chapter by showing how enactivism and dynamical systems theory form an essential part of neurophenomenology—an interdisciplinary research program which actively tries to bridge this experiential/materialist divide.

Ultimately, I argue for a version of enactivism that functions as a sustained qualification—rather than an outright repudiation—of neuroscientific orthodoxy and its representationalist underpinnings. Even though enactivism emerged foremost as a critique of representationalism, I will argue that we need to leave the door open to the role of representation in mental life, provided that these processes are understood within the larger phenomenological implications outlined by enactivism.

The Cognitive Revolution:

The origins of cognitive science can be traced back to the late 1950s and early 1960s, during an intellectual turning point in the sciences of mind, now referred to as "the cognitive revolution" (Cain 19). At the time, the discipline that dominated scientific inquiry into mental life was psychology, and the school of thought that dominated psychology was behaviorism. Behaviorism was a reductionist theory whose appeal for psychologists lay in its falsifiability. Unlike psychoanalysis, it explicated mental phenomena in a way that did not hinge on subjective mental states, introspection, and the intractability of meaning-making, focusing instead on those external aspects which could be finely controlled and measured. Behaviorist methodology was built on the mechanisms of input (stimulus), behavioral output (response), and reinforcement. It eschewed internal mental states in its calculus because this was viewed as unreliable, untestable, and thus unscientific. Yet by the late 1950s and early 1960s it became increasingly difficult to ignore the constitutive role of mind—that is, the internal workings of organisms and their nervous systems—in determining things like behavior and learning.

Noam Chomsky's 1959 review of B.F.Skinner's *Verbal Behavior* is viewed as a watershed moment in the shift away from behaviorist dogma and towards a new science of cognition (Cain 20). The pith of Chomsky's critique revolved around the complex process of language acquisition. For Chomsky, stimulus, response, and reinforcement were blunt instruments woefully incapable of accounting for the integrated, intuitive, and rapid manner in which human beings acquire a natural language. Chomsky argued that the cognitive abilities required to learn a language could not be reduced to a mechanical process initiated from without. Mind, i.e., some innate and internal process was simply the most expedient and common-sense explanation.

Another key figure in the emergence of cognitive science was Donald Broadbent. Broadbent, together with Anne Treisman, are often cited for their experiments on selective attention in which participants were given headphones with different sounds (words, numbers, etc.) issuing into each ear. Participants were then asked to attend to the sounds coming into one ear or the other and report what they heard. What Broadbent and Treisman found was that participants could not recall the sounds from the unattended ear. These experiments became the foundation for Broadbent's "filter theory" which provided strong experimental evidence for the existence of internal mechanisms that selectively engage with and filter out the overwhelming number of stimuli our brains receive in a given act of information processing. During this period cognitive psychologists became increasingly occupied with discovering and speculating on the range of underlying information bottlenecks and filters that enabled selective attention, and ultimately, higher-order cognition. These early studies were particular interested in whether stimuli were selected due to the physical features of the sounds or on the basis of meaning; whether selection mechanisms were activated earlier or later in these information processing stages; to what extent two or more stimuli could be processed at the same time, and in the requirements for "stimulus processing to become highly rapid, automatic, and effortless…" (Kornblum and Meyer 179). To this end, Broadbent employed "black box diagrams" that depicted salient aspects of this filtering process using boxes connected by arrows. These were called "black" boxes because while the filter could be reliably inferred, as could the general sequence of events in the brain, the causal mechanisms that ultimately determined this process remained unknown (Passingham 2).[11]

Increasing disillusionment with behaviorism as an explanatory model, coupled with new experiments such as Broadbent's and Treisman's were key players in the cognitive revolution. But perhaps the most decisive development in this new science of mind was its early marriage to computer science. As Thompson notes, "At the center of this revolution was the computer model of mind…according to this classical model, cognition is information processing after the fashion of a digital computer" (Thompson 4). What computer science afforded early cognitivists was the idea that the aforementioned "black boxes" could be understood as computations.

Computation as we now understand it was given its rigorous, formal definition thanks to the work of Alan Turing. He demonstrated that every mathematical calculation that could be completely defined in a finite number of steps could also, in principle, be performed by a machine. A "Turing machine" was imagined as a small box situated above an infinite reel of paper on which symbols were inscribed (either zeroes or ones). The machine begins in an initial state and depending on the symbol encountered is prompted to perform a certain operation; either move one square to the left, one square to the right, or remain in place. This step-by-step set of instructions for achieving a particular outcome or consequence is what we now know as an *algorithm*. Turing's imagined machine, programmed with a given algorithm, scans the symbols on the tape and

---

[11] These black box diagrams are not to be confused with the "black box" metaphor of the mind adopted by behaviorism, which denotes the unobservable nature of mental life.

performs the corresponding operations until the final state has been reached and its operation ceases. An algorithm's essential function is to transform information from one form into another. Turing machines were a thought experiment that provided a concrete and intuitive way of demonstrating how computation could be rendered in purely formal terms, i.e., as a set of fully describable mechanical steps. Terrence Deacon describes this mechanical underpinning as follows: "In this respect, the workings of my desktop computer, or any other computing machine, are just the electronic equivalents of levers, pendulums, springs, and gears interacting to enable changes in one part of the mechanism to produce changes in another part" (Deacon 96). In modern computing devices, these levers and gears take the form of logic gates and digital circuits.

The breakthrough for modern computing was the recognition that *any* completely definable process or procedure can be translated into binary code and reduced to an algorithmic form capable of being computed. Turing's insights revolutionized computing, but importantly for cognitive science, it provided a robust scientific framework; and perhaps just as important, a new idiom for integrating symbolic representation back into mental life. Cognition understood as computation could be imagined in mechanistic terms that avoided (or so it seemed) putting the ghost back into the machine. The effects of these changes were felt across disciplines. Chomsky's generative approach to grammar for example is essentially computationalist. It argues that language processing functions on similar grounds to computation (iteration and recursion) and that language is more easily explainable (i.e., its outstanding paradoxes and problems) if one approaches it not as an external phenomenon aimed at communication, but as an internal process intimately tied to cognition. The core of the generative position is that surface structures—the variations in syntax we see deployed in natural languages—are expressions of deeper syntactical structures, all of which resolve themselves into a hypothesized "universal grammar," an innate biological and mental faculty that constrains the grammatical constructions of all natural languages.

Why the excision of mental representations in the first place? Representationalism has been a perennial position in philosophy of mind. As one might expect, it holds that thought occurs through mental representations: images, concepts, or symbolic structures that in some way *stand for* objects of experience. Recourse is often made to the ideal of pictorial images, but the basic position can be extended to conceptual schema or symbolic structures (i.e., the kind of algorithmic logic of computationalism), as well as to abstract mental states such as beliefs, perceptions, and intentions. To this end, philosophers distinguish between categorical and noncategorical representations. A category establishes a class of belonging based on what *is* the case. While mental states are often speculative—what *might be* the case (Papineau 97).

One useful way to begin envisioning the notion of mental representations is through the analogy of a map. A map stands for some object in the world. It can scale up this slice of the world, scale it down, or highlight particular features (e.g., roads vs topography) but in all these instances what defines a map are its isomorphic properties, that is, some kind of formal correspondence between groups or sets (OED). Maps preserve a set of relations found in the world in a convenient, portable, interpretable manner. What defines the map, and the representation as such, is its content-bearing property.

Representational content is determined by truth conditions. David Papineau says, "What all representations have in common is 'truth conditions.' Any representation will portray the world as being a certain way. It will draw a line in logical space, dividing the possibilities into those that verify it and those that do not" (Papineau 95). Truth conditions provide the boundary conditions of a representation. Content emerges within those boundaries marked out by the truth conditions by distinguishing between valid and invalid cases. Papineau continues, "It is not always easy to articulate what is being claimed by a perception or by a map or other pictorial means of representation. But this does not mean that these states lack truth conditions…" (Papineau 95). In the case of a map, if one is lost on asphalt thoroughfares then the representation of roads becomes the salient truth condition. If one is trying to determine one's height relative to other features, then a topographical representation makes sense. Thus, representationalism in the context of cognitive science holds that our myriad mental states, perceptions, and cognitions are defined by these mental representations, the content they carry, and the truth conditions that determine and demarcate that content. Even though enactivism emerged as an embodied critique of many of the representationalist underpinnings at work in early cognitive science, we need to leave the door open to representation because as Terrence Deacon will convincingly argue, representation captures the essence of purposiveness and intentionality. Representations are relational, they reach beyond their material instantiation and connect material objects with intentional actors. They are the essence of telos, and the intersubjectivity that defines the ethical relation. By excising representation, we risk excising our hold on ethics.

Let us consider a few specific examples of a representational theory of thought. Doing so will bring these foregoing ideas more sharply into focus. During the Enlightenment, Immanuel Kant created an elaborate architectonic of human understanding and experience which relied on a representationalist view of mental life. I highlight Kant because his views remained highly influential up until the early twentieth century when behaviorism became ascendent.

In the first Critique, he lays out his so called "Copernican" revolution in philosophy by inverting the terms by which epistemology had traditionally been ordered. He says, "Thus far it has been assumed that all our cognition must conform to objects. On that presupposition, however, all our attempts to establish something about them a priori, by means of concepts through which our cognition would be expanded, have come to nothing" (Kant 21). In this passage we can see the basic outline of the empiricist/rationalist debate that dominated Enlightenment epistemology: If we assume that thought must conform to the objects of experience and that these objects determine our cognitive content, then how, rationalists wondered, is it possible that we can have a priori knowledge of the world?[12] He continues:

---

[12] The basic question at the heart of Kant's first Critique is "How are synthetic judgements possible a priori?" (Kant 59). In Kant's vocabulary, a judgement is a propositional thought. An *analytic judgement* is a proposition that is contained in the meaning of the subject e.g., all bachelors are unmarried men. A *synthetic judgement* is a proposition that adds something not contained in subject, e.g., the ocean is blue. In other words, it would seem to require experience. Therefore, how is it possible to assert new knowledge of the world prior to experience? For clearly this is the case, as Kant's examples physics and Euclidean geometry attest to: they are disciplines enable one to assert new, truthful propositions in pure intuition

> If our intuition had to conform to the character of its objects, then I do not see how we could know anything a priori about that character. But I can quite readily conceive of this possibility if the object (as object of the senses) conforms to the character of our power of intuition. However if these intuitions are to become cognitions, I cannot remain with them but must refer them, as presentations (*Vorstellungen*), to something or other as their object, and must determine this object by means of them. (Kant 22)

Kant makes decisive distinctions between *sensations* (raw unshaped feelings or perturbations of the subject), *intuitions* (unmediated, conscious perceptions of objects), and *cognitions* (intuitions which have been shaped by the categories of the understanding). But what is ultimately at stake in this passage is the idea that objects of experience do not determine our cognitions—although clearly these objects must be present and impinge on us if our thoughts are to have any determinable content at all. Rather, it is our pure a priori intuitions of space and time, our receptivity to external objects, and our innate categories of understanding that determine our cognition of those objects. In essence, these intuitions and categories are nothing other than mental representations (the word *Vorstellung* is traditionally translated as *representation*) they *stand for* an external reality.

(i.e., without experience of the world). Therefore, the first Critique assumes the truth of the question "How are synthetic judgements possible a priori?" and then works backward, explaining how this strange phenomenon of human cognition can, in fact, be the case.

The genius of Kant's transcendental idealism was to synthesize the two dominant threads of Enlightenment epistemology, exhaustively synthesizing core aspects of empiricism and experiential knowledge, and the rationalist belief in innate ideas. Perhaps the most famous line of the first Critique: "…thoughts without content are empty; intuitions without concepts are blind" (Kant 107) demonstrates in uncharacteristically plain language for Kant the synthesis of these two epistemological positions. Kant tries to strike a middle ground between the extremes of skeptical empiricism and subjectivist rationalism, and his (unnecessarily) strong claim that experience is, in the final instance, *wholly* removed from things in themselves (noumena) places this representational aspect of mind front and center. Kant is a touchstone in any genealogy of representationalism: not only given his outsized influence of Enlightenment epistemology (and by extension, the philosophical assumptions operative in contemporary neuroscience—see page 19); but because he is widely cited as the ur-expression of modern correlationalism by speculative realists (Sutherland 109) and so plays a central role in their attempts to uncouple continental philosophy from these anthropocentric roots; for if the being of things can only be known in their relation to thought, then there is an inherent privilege afforded to those who can think such thoughts, i.e., humans. Speculative realism and object-oriented ontologies name and confront this anthropocentrism, but as stated in my introduction, their focus on unmediated object-object relations and/or attention to large networks of interaction (e.g., Latour's ANT) effectively lose sight of the face-to-face relation that characterizes ethics.

Another important representationalist theory of mind, broadly conceived, can be found in the structuralist linguistics of Ferdinand de Saussure. Saussure's definition of language, speech, and his understanding of the sociological process he called the "speech circuit" are all predicated on a two-sided psychological conception of the linguistic sign, composed of a signified (a concept) and a signifier (a sound-image). Both aspects of the linguistic sign refer to mental representations, but it is the signifier in particular which points to mental representations as possessing a distinct image-like quality. He says:

> The linguistic sign unites, not a thing and a name, but a concept and a sound-image. The latter is not the material sound, a purely physical thing, but the psychological imprint of the sound, the impression that it makes on our senses…The psychological character of our sound-images becomes apparent when we observe our own speech. Without moving our lips or tongue, we can talk to ourselves or recite mentally a selection of verse. (Saussure 66)

Here the adjective "psychological" clearly refers to the inner-sense and interiority of mental experience, one which is not reducible to the materiality of phones. Saussure defined "language" as the overarching, and virtual set of grammatical rules and conventions that determine meaningful utterance. Saussure's focus was on the functioning of linguistic sign systems as something acquired and virtual (i.e., non-material—existing between multiple consciousnesses). But as the foregoing passage demonstrates, to arrive at these semiotic processes, he made use of representationalist assumptions. The raw sounds of human speech get *mapped onto* a representationalist schema of which the signifier is the first step. The signifier is a sound-cum-psychological imprint that functions as a mental representation. Even though the signifier is a product of social construction, for Saussure it gets assimilated into our psyche as a fixed psychological fact. Embedded in the speech circuit, signifiers soon develop into full-fledged representational content via their indelible connection to concepts. Saussure is another important point of reference because his binary notion of the linguistic sign shaped the course of twentieth century continental philosophy, and by extension many of the critical and theoretical insights at the heart of contemporary literary studies, and while he sought to understand language foremost as virtual and emergent *system*, what enabled Saussure to link this shared system to human thought and perception was a largely representationalist conception of the signifier and the way it embedded itself in our cognitive architecture.

Perhaps the most important representational theory of thought to emerge in this early period of the cognitive turn came from Jerry Fodor. In 1975 he published *The Language of Thought*, which revived the language of thought hypothesis from its deep Scholastic roots and transposed it onto the debates of contemporary linguistics, analytic philosophy, and computational theories of mind. In his retrospective of Fodor's life and work, Georges Rey gives the following summary of his position in *The Language of Thought*:

> Propositional attitudes, such as belief, preference and expectation, were to be treated as computational relations to sentences in a 'language of thought' (an 'LOT') that was encoded in the brain. In that book and in much subsequent work (e.g., Fodor, 1981b, 1987), he argued that only such an approach could capture and

> begin to explain what seems to be the lawful compositional structure, productivity and 'systematicity' of thought: possible thoughts have recursive logical structure that permits people to be able to think a potential infinitude of ever more complex thoughts… (Rey 326)

The language of thought hypothesis, both ancient and contemporary versions, proposes the existence of some kind of "mentalese"—an internal, symbolic language by which cognition organizes itself. As in natural language, this mental language operates according to compositional semantics: there are basic symbolic units that can be combined to form more complex representations and logical relationships. Like natural language, this mentalese operates by recursion—repetition of a grammatical structure by nesting one rule within another—a consequence of which is that there is no upper limit to the number of mental representations that are possible.[13] Daniel Hutto says, "Proponents of Fodor's 'language of thought' hypothesis are interested in mental representations with semantic content of essentially the same kind as that which natural language sentences possess (this follows given that whatever content the latter have is wholly derived from the former)" (Hutto 49). This parenthetical remark by Hutto is the pith of Fodor's approach, and of all approaches that sought to naturalize the semantic content of mental representations by appeals to natural language. The idea is that "an exact parallel exists between the content of a thought and language" and thus that "saying

[13] In linguistics recursion is a fundamental feature of grammar. Steven Pinker says, "These rules [the grammar] embed one instance of a symbol inside another instance of the same symbol (here, a sentence inside a sentence), a neat trick—logicians call it 'recursion'—for generating an infinite number of structures" (Pinker 93).

something really only requires finding a public means of expressing oneself" (Hutto 49). For Fodor, this symbolic language of thought serves as the raw symbolic material over which Turing-style computations occur. Thus, Fodor was one of the first theorists to attempt a synthesis of computational theories of mind and analytic philosophy. These computations resolve themselves into explicit propositional content, and ultimately propositional attitudes—the beliefs and intentions that define the contents of mental life.

Cain describes propositional attitudes as "a belief, a desire, an intention…[which] in virtue of being relations to propositions…have meaning or semantic properties. They are therefore akin to declarative sentences of a natural language. Just as such sentences are about particular things (or types of thing) or states of affairs and represent them as being a particular way, so do propositional attitudes" (Cain 9). The notion of a proposition is inherently linguistic and conceptual. Propositions operate by predication, ascribing properties to things and in the process subordinating terms and ideas to one another in grammatical fashion. Propositional attitudes describe a combination of two types of mental phenomena: (1) mental states (intentions, desires, fears, hopes etc.) that one holds towards a (2) proposition. According to this view, by describing both the mental state and the propositional content, one can effectively describe the contents of thought. In this way, propositional attitudes are the end result of complex arrangements of mental representations and logical (computational) operations. Given these foregoing descriptions, it is clear that the LOT hypothesis focuses exclusively on the highest-level cognitive processes—things like logical inference and abstract reasoning—and has little to say about other aspects of mental life such as basic perception, sensation, and motor control.

Yet for all their advances in theorizing cognition representationalist theories of mind pose one glaring problem, namely, they open themselves up to the homunculus fallacy. Homunculus is Latin for "little man." In the context of philosophy of mind, it refers to a tiny, fully formed conscious being that possesses all the basic cognitive faculties of a person. In effect, a homunculus is a way of smuggling in the essential properties of mind (i.e., a non-material, purposive, teleological cause) into an otherwise materialist, scientific description. It is a logical fallacy because it tacitly assumes the very quality (cognition) that the description is attempting to explain. In other words, it begs the question.

External representations (i.e., artifacts, symbols, speech etc. we encounter and manipulate in the world) do not intrinsically signify. They require conscious, cognitive beings such as ourselves to employ them, to put them to use, and thereby attribute to them meaningful content. Just as external representations require our conscious intentions to animate them with meaning, so too would internal mental representations require some kind of conscious agent or intention—enter the homunculus. Of course, this would then require *yet another* homunculus to ascribe meaning to the representations for the first homunculus, and so begins an infinite regress. One of the upshots to behaviorism was that it avoided the pitfalls of the homunculus fallacy. Unfortunately, it accomplished this by evacuating mental life of its interiority. Deacon says, "Probably one of the keys to the success of the new cognitive sciences…was that they found a way to incorporate an empirical study of this hidden dimension of psychology while still appearing to avoid the dreaded homunculus fallacy. The solution was to conceive of mental processes on the analogy of computations, or algorithms" (Deacon 94). But as we will see, the computationalist answer to the homunculus fallacy turned out to be short lived.

According to the mind-as-computer model, the *meanings* of mental representations are understood solely by their formal (syntactic) structure. In a computationalist framework "nonsymbolic sensory inputs are transduced and mapped onto symbolic representations of the task domain. These representations are then manipulated in a purely formal or syntactic fashion in order to arrive at a solution to the problem" (Thompson 5). Sensory neurons in our peripheral nervous system are stimulated from without (either physically or chemically) and are then transduced into the electro-chemical form that can travel between the peripheral and the central nervous systems and ultimately into the brain where these signals are arranged and processed according to their various problem-solving tasks. According to computationalism, this process retains the basic properties of the input just in a different form, one that can be manipulated via the brain whose computations constitute the mental representations, i.e., their "meaning."

In this way, software supersedes hardware, meaning that even though computation is essentially a mechanical process, what ultimately determines one's cognitive content is the internal state—the arrangement of symbols and the syntax—and not the physical system in which it manifests. This view was given philosophical legitimacy by Hillary Putnam and his theory of functionalism. According to this theory, a mental state, like a machine state, is defined by its formal arrangement and the functions it performs in relation to inputs, outputs, or other states, and thus can be instantiated in any number of physical ways. Cain says, "Putnam argued that mental states are Turing machine states. In developing functionalism along these lines, Putnam made clear its affinity with the idea that cognition involves computation and so revealed its value as the metaphysical theory of the mind needed by cognitive science" (Cain 25)[14].

But as one problem was remedied another soon emerged. In the cognitivist-computationalist paradigm *cognition* could only be legitimized in materialist terms as the effectuation of subpersonal routines or programs of token manipulation, "that the processes of perceiving and reasoning about the world are also constituted by algorithms 'instantiated' as physical processes…" (Deacon 494). In this way, cognitivists revived symbolic representation as scientifically valid, and by extension the importance of what goes on *inside* a person's head; but this legitimation of interiority had no way of explaining the purposive and intentional qualities of mind, nor the felt, qualitative and phenomenal experiences of mental life (i.e., qualia). As Deacon tells us:

> In all computational theories…representation is understood in terms of a rule-governed mapping of specific extrinsic properties or meanings to correspondingly

[14] Putnam would later adopt a position known as "semantic externalism," which, as the name suggests, maintains that meaning is constituted, at least in part, by features that are external to the speaker (Kallestrup 2), and this "…eventually led him to abandon functionalism. Although it identifies mental states by their function rather than by their material composition, functionalism is still committed, according to the later Putnam, to an internalist conception of these states. In other words, functional states are brain states, which are insensitive to the outer world and to the individual's concrete interactions with it" (Ben-Meneham).

> specific machine states. Although this may at first sound like an innocuous claim, it actually begs the fundamental question: how is representation constituted? Simply asserting the existence of mapping between properties in the word [sic] and properties of a mechanistic process presumed to take place in the brain implicitly frames the relationship in dualistic terms. There are two domains: the physical processes, and whatever constitutes the representational relationships between them. Everything depends on the nature on the nature of the correspondence… (Deacon 495)

According to computationalism the "extrinsic properties or meanings" (meaning a specific problem-solving task) are an assumed feature of the world that is being mapped onto the "machine states" of the brain (meaning the patterns of activation exhibited by neurons). But as we know from spoken language, what allows a representation to *signify* is the purpose and context in which it occurs. In the computationalist paradigm this teleological dimension is implicit in the outside domain that is being mapped onto the internal domain of mental life. Deacon continues, "the logic of computation is defined in a way that is agnostic to how (or whether) representation, inference, and interaction with the world are implemented. These issues are bracketed from consideration in order to define the notion of an algorithm…" (Deacon 496). Thus, no real explanation has been given about how these purposive processes represented in the algorithm (such as "solving a mathematical equation" or "organizing a list") came about in the first place (Deacon 496). In effect, the teleological dimension to mind has been buried in the description itself.

As Thompson notes, “This radical separation of cognitive processes from consciousness created a peculiar ‘explanatory gap’ in scientific theorizing about the mind” (Thompson 6). This revived explanatory gap is what David Chalmers calls “the hard problem of consciousness.” According to Chalmers, cognitivism buried the traditional mind-body problem of Descartes in a deeper level of explanation, recasting it in the form of a mind-mind problem. Deacon says, “So, while at one time it was possible to argue that the computational concept of the mind/brain relationship was the only viable game in town, it now appears that it never really addressed the problem at all, but only assumed that it didn’t matter” (Deacon 496). We will return to the explanatory gap at the end of this chapter because Thompson’s autopoietic enactivism will propose a clear means of bridging this gap.

The next big step in theories of cognitive science occurred during the 1980s with connectionism. I will pause briefly on this theory for one important reason—its recourse to cybernetics (Thompson 8). Whereas cognitivist models hinged on abstract reasoning and information processing envisioned on the model of a digital computer, connectionist models of cognition “emphasized perceptual pattern recognition” and invoked as their paradigm the model of the neural network (Thompson 9). Instead of abstracting a syntax that operates according to a logic of linear (or serial) input/output; connectionists focused on the behavior of neural connections in a network and how these connections could be taught to converge toward certain goals or end states. In other words, given certain inputs and training the network would learn how to effectuate a certain output without needing to rigorously define in advance the serial, step-by-step process that led to this output. All that is needed are the inputs/outputs and sufficient time for the network dynamics to emerge and learn from itself. While still computational in the sense that the problems or “cognitive tasks” are predefined, i.e., the outputs/inputs are fully defined and programmed into the network from without, the connectionist approach was seen as a much more accurate model of biological neural processes. As Cain notes, “[connectionism] portrays cognition as being based upon the activity of networks of simple units that bear considerable similarities to the networks of neurons we find in the brain” (Cain 43). Whereas cognitivists took inspiration from algorithmic processes of digital computers, connectionists looked directly to the cognitive architecture of the brain for theoretical and experimental inspiration.

Instead of symbols, connectionist networks are composed of connection points (nodes) and links. Connections between nodes “fall into two classes, namely, excitatory and inhibitory” (Cain 44). Here again we see how connectionist networks closely mimic the electrochemical processes by which neurons in the body communicate. Each node can be assigned a numerical value depending on the activity it receives and “the specific input–output behaviour of the network is determined by the nature and weight of the connections and the threshold values of the units” (Cain 45). Cain sums up these insights as follows:

> For the connectionist the mind-brain is a connectionist network or ensemble of such networks. But one might ask how this could be so. How could a system like the one described above support cognition? The crucial point is that patterns of activation at the input and output layers can have semantic properties – that is,

> they can represent. Another important feature of such networks is that they can learn by changing the profile of their input–output behaviour in the light of experience. (Cain 46)

In other words, instead of postulating some kind of inscrutable mentalese, representational content is carried in the overall *patterns of activation* that emerge in the changing dynamics of neural networks. Because these connections can be assigned numerical values they can be computed, thus preserving the basic computationalist paradigm of information-processing after the fashion of a digital computer, only in this case, instead of simple input/output procedures we have a complicated networked architecture. I linger on this connectionist paradigm because as Thompson points out, this modeling of neural networks revived important aspects of cybernetic systems theory.

The link between cybernetics and the neural networks of connectionism is the idea that a cognitive system is *not* maintained via rigid control—such as that suggested by traditional computation—but rather by flexibility. Variation in a system is acceptable and in most contexts even beneficial so long as it stays within a specified range. For example, artificial neural networks need to be "trained" to compute the proper output, and they do this through repeated cycles of trial and error. Bray says, "A computer neural network changes its connection strengths in response to a teaching input. A training set of inputs is presented to the network again and again, and the output compared with the correct answer. The difference is then used to modify the weights of the network…the process is continued until the error has dropped to an acceptably small value" (Bray 120). In other words, neural networks learn by error correction. It is the flexibility of the network and its capacity to induce small changes in its organization as it homes in on the desired output that enables training to occur. Furthermore, this training regime can exhibit the rudiments of circular organization (output=>input=>output) described by cybernetic control systems. Traditionally this circularity has not been endogenous: it had to be closely facilitated by the computer scientist. With current advances in machine learning, networks can now effectively train themselves and so embed these circular processes in their own computational architecture (provided of course that the scientist still define the overall input/output parameters and data set on which to train). It is the flexibility and circularity of connectionist neural networks that brings the computational aspects of cognition closer in line with the behavior of biological systems.

Coming to terms with "Perception":

Thus far, we have considered the role of mental content and mental representations in terms of thoughts: i.e., *conceptually* grounded mental states. Where does perception fall in this schema? And why, as this subsection asserts, do we need to "come to terms" with perception? Cognitive science seeks to update and redefine the concept of mind, and to that end, many different mentalistic expressions come under review. This section unpacks the significance of "perception" (and the related faculty of sensation), mentalistic expressions which cognitive science has rendered distinct from thought and cognition.

In our overview of the history of cognitive science thus far two basic positions have emerged: representational theories of mind and embodied theories of mind. As we've seen, the history of cognitive science is overwhelmingly a history of the former, which holds that mental processes are the computation of intrinsically content-bearing states. Computation was a key facet of this theory because it provided the metaphor and mechanism that allowed such representations back into the scientific fold. By analogizing the brain to a kind of digital computer, neuroscientists, armed with the best tools of their trade such as EEG and fMRI, could look at the patterns of activation (neural assemblies) that emerge in the brain and correlate these physical, measurable events with their associated experiences. These neural correlates of consciousness provide a predictive and generalizable framework assembled from demonstrable facts: in other words, a proper scientific theory.

In the context of analytic philosophy—which was very clearly neuroscience's philosophical counterpart during this cognitive turn—a *thought* is a composite of logical or propositional content with an associated mental state (belief, desire, intention etc.), known as a *propositional attitude*. According to this view, a propositional attitude is the natural terminus of a thought process. While there is much about our cognitive life that escapes conscious awareness, a thought by this definition is *not* one of them. It assumes that most of our cognitive processes flow towards, or work directly in the service of, these conscious and conceptually grounded mental states. Because propositional attitudes are built stepwise from smaller parts into more complex conceptual wholes they remain amenable to analytical methods. As Cain says, even when thoughts appear as if out of nowhere, with no discernable process informing them, "we can retrace our steps by deliberately seeking to justify our conclusion and so gain awareness of our thought process retrospectively" (Cain 10). Because it seeks to unify mental states with a definable content, propositional attitudes were another important framework for mediating between the disciplines of neuroscience and analytic philosophy. Cognitivists endorsed it because it helped reintegrate mental representations and interiority into a materialist explanation, and analytic philosophy endorsed it because by defining thoughts along linguistic lines they became amenable to logical analysis.

But perceptions don't seem to work this way. Cain says, "I open my eyes, orient them to the world and have a perceptual experience without having any awareness of executing an extended process the earlier elements of which justify the later elements" (Cain 10). In other words, in the realm of perception most of what goes on seems to be either (a) immediate and unconscious (and therefore not subject to logical analysis), or (b) phenomenal and qualitative (and therefore not subject to materialist explanation). Thus, many philosophers of mind have traditionally distinguished between propositional attitudes on the one hand; and perceptions, sensations, and emotions on the other precisely because of their phenomenal and qualitative character (Cain 10).

Despite these differences Cain also articulates a powerful common denominator between perceptions and propositional attitudes, one which will serve as a good point of departure for understanding perceptions in neuroscientific terms. He says:

> Even if, as many philosophers have thought, perceptual experiences are very different from propositional attitudes in not involving the deployment of concepts and in having an intrinsic qualitative character, they are like them in a key respect. For perceptual experiences are representational in the respect that they are

> typically about objects located in the external world and represent them as being a particular way. (Cain 10)

This quotation echoes what was said earlier by Papineau of representations in general.[15] It argues that perceptions and propositional attitudes are united in their role of mediating the same basic inside/outside structure via truth conditions, meaning a set of contexts or conditions in which the representation or perception is valid or not. This act of mediation Cain casts in the loaded language of "aboutness," that is, as an action of reference, intentionality, and interpretation.

Even though Cain is explicating these concepts primarily in the context of analytic philosophy this is the same basic picture we see adopted by mainstream neuroscience. In a 2015 article in the journal *Dædalus,* for the American Academy of Arts and Sciences, Thomas Albright, a neuroscientist working in the field of visual processing and attention, outlines several definitions at the core of contemporary neuroscience, and which further illuminate the disjunctions between first/third person descriptions. He says, "In modern neuroscience, our working definition of perception is captured well by the *Oxford English Dictionary*: 'The action of the mind by which it refers its sensations to an external object as their cause.'" (Albright 22). This definition, Albright says, is born from epistemological theories of the Enlightenment period—empiricists such as Berkeley, Hume, and Thomas Reid—and contains two essential parts.

---

[15] "Any representation will portray the world as being a certain way. It will draw a line in logical space, dividing the possibilities into those that verify it and those that do not" (Papineau 95).

The first is the distinction between perception and sensation. Perception is an inherently referential and active process, while sensation is a passive index of environmental change. He says, “Sensation is the immediate neurobiological consequence of stimulating *sensory transducers* such as photoreceptors, mechanoreceptors, and chemoreceptors. Sensory events are ubiquitous…but they are fleeting, discontinuous, and lacking semantic content” (Albright 23). Sensation is a neurobiological (i.e., wholly mechanistic) process that results from the world impinging on the sensorium. It is passive. It does not require any activity of mind, and the result is that we are constantly bombarded with, and awash in, sensations. This passivity leads us to the second part of this OED definition. Albright says, “the referent is viewed as the ‘cause’ of sensation” (Albright 23). In other words, when one perceives one attributes or *refers* one’s sensations to the world and not to one’s own sensorium or cognitive processes. While this claim may seem unremarkable—that one is not the cause of one’s own sensations, “This attribution to the source of sensation is profoundly important and meaningful for perception and behavior” because it structures experience as belonging to an objective world, and not to oneself. According to Albright, this attribution of an external cause is implicit in the act of perception, and it impresses upon the percipient that they are witness to a shared reality, not hopelessly lost in solipsism.

This intuition of an “outside” world is the first “referential” act of perception: it implicitly refers its sensations to an environmental cause. But there is another important aspect to perception’s referentiality. Albright says, “Perception enriches sensation by reference to other knowledge or experience” (Albright 23). In both cases, perception is a secondary process that depends on sensation, but in addition to simply referring sensations to the world outside oneself, perception also shapes those sensations into a meaningful framework by means of reference to other knowledge or experience. It creates meaningful relationships that allow the percipient to judge, navigate, and predict things about his or her environment.

Notice, too, that in Albright's definitions, sensation is referred to as an *event*, while perception is referred to as *action*, one that is closely allied to both knowledge and experience. Sensations are events because they require no faculty of mind or intentionality, they merely register a perturbation or change, whereas perceptions are akin to *experiences* because they rely on the intentional qualities of mind—to context, reference, and meaning. In other words, they refer to things. They refer to the percipient who implicitly understands that these sensations in some sense belong to them—that they are occurring *in me*; and they refer to the world because the source of these sensations are environmental causes.

Cain defines perception along lines similar to Albright: "a process where the external world stimulates an individual's sensory organs, resulting in a perceptual experience. In the case of vision, light reflected off external objects is focused onto the retina, a light-sensitive surface at the back of the eye that sets off a mental process which results in a visual experience" (Cain 10). Like Albright, Cain makes a distinction between mechanical stimulation from without, and the internal *experience* that follows from it. We see this division between mechanical stimulation and mental experience reified in how neuroscience describes the problematic of vision. Albright says:

> Technically this is two problems, one of which is optical and the other biological. The optical problem involves reflection of light off of environmental surfaces, refraction of that light by the crystalline lens of the eye, and, finally, projection of the light onto the back surface of the eye—the surface known as the retina, which is lined with neuronal tissue responsible for phototransduction—to form the retinal image. This image is merely a pattern of light that changes in intensity and wavelength over space and time. (Albright 23)

Thus, the first problem is one of physics: how does light reflect, refract, and project itself onto the retina. Albright continues:

> The biological problem, which is by far the more difficult of the two, thus involves reconstruction of the properties of the visual scene, given only the pattern of light present in the retinal image. This inverse problem of optics is an example of what is known formally as an ill-posed problem: it is a problem without a unique solution. Because of the dimensionality reduction that accompanies optical projection of the world onto the retina, there is simply not sufficient information present in the retinal image to uniquely identify its environmental causes. To put it bluntly, the number of visual scenes that could give rise to any specific retinal image is infinite. (Albright 23)

The reduction from a three-dimensional world to a two-dimensional image—and the indefinite number of visual scenes that could have caused that particular pattern of light to be projected onto the retina—means that in order to function as a source of accurate information about the world, that the retinal image must be *interpreted*. Noë, who spends almost the entirety of his book examining vision, echoes this basic problematic. He says, "There is an enormous discrepancy between the character of the input to vision…tiny, distorted, upside-down retinal images—and the high-resolution colorful world that we know in experience" (Noë 36).

As Albright's breakdown of the optical/biological dichotomy demonstrates, neuroscience makes stark distinctions between mechanical and biological processes—between the physical "events" of sensation and the biological "experiences" that constitute perception. Despite being couched in the language of "biology," this so-called biological problem is really one of *mind*, that is, subjective experience. The point at which the physical pattern of light needs to be *interpreted* by the perceptual faculties of the organism to produce an accurate representation of the world, is precisely the point at which the quality of mind/intentionality instantiates itself. In the case of Albright, this explanatory gap becomes coded as a "biological problem," but calling it biological does little to alter the fact that intentionality and mind has been assumed in the description.

The challenge, then, for materialist science, is how to account for this transition from the optical to the biological (i.e., the materialist to the purposive) in a way that can remain consistent with the exigencies of third-personal description—or at least not impede its forward march. As Albright goes on to tell us, neuroscience's recourse is to the language of *context*. He says:

> Vision is able to accomplish reliable disambiguation of the retinal image by virtue of context, which, broadly speaking, consists of other pieces of information that are either 1) co-present in the retinal image (spatial context); 2) learned based on

> the observer's prior experience with the world (temporal context); or 3) embedded in the computational machinery of the brain as a result of evolution in an environment that has consistent and well-defined properties (evolutionary context)…Using the available context as clues, the process of disambiguation–the process of perceiving–is best characterized as probabilistic inference about the cause of sensation. (Albright 24-25)

In an experimental setting, *spatial context* refers to all the non-target features of the retinal image, meaning all those stimuli which are not simple and well-defined, but which are still present in the visual image (26). Markers of this kind of spatial context are color, pattern, and motion present in the regions surrounding the target—otherwise known as the "surround" (26). Gestalt psychology, for example, described many different types of organizing principles which are examples of this spatial context described by Albright and which they dubbed *laws*. There is *the law of prägnanz*: when given a complex array of stimuli our brains will render them as simply as possible; *the law of similarity*: objects that share common features (shape, color, size, etc.) will be perceived as coherent groups rather separate; *the law of proximity*: things that are close will implicitly appear related; *the law of closure*: our brains tend to fill in what is lacking in a visual percept (e.g., the Kanizsa triangle). The list goes on. It was a research program that focused almost exclusively on these kinds of spatial contexts described by Albright and argued that perception is structured according to these irreducible perceptual patterns, or *Gestalten*.[16]

---

[16] In chapter six I will examine the role of gestalts in narrative theory and make the case for why these gestalts can be understood—not on visual grounds—but on embodied grounds. It is also worth pointing

Evolutionary context refers to those processes acquired and perfected on evolutionary timescales in response to environmental exigencies. A clear example of evolutionary context is the phenomenon of face pareidolia, which is "the illusory perception of meaningful shapes from random or ambiguous stimuli…" As Palmisano et. al. tell us, "…from an evolutionary perspective, among the multitude of random stimuli we perceive in everyday life, others' faces represent the most relevant stimuli for social relations, and our visual system seems to be particularly sensitive to facial configurations" (Palmisano et. al. 1). This hardwired ability to perceive faces—even where there are none—is precisely the kind of "computational machinery" referenced by Albright, machinery selected over countless generations due to the advantages it confers on social organisms.

For our enactivist purposes, *temporal context* is by far the most germane. It refers to how prior personal experience determines how one perceives a sensory stimulus. As Albright tells us, the effects of temporal context are "ubiquitous and fundamental to perceiving. They are rooted in the phenomena of *associative learning* and *memory retrieval*: as an observer learns the relationship between a sensory stimulus and the 'farther facts associated with the object of sensation,' the stimulus alone is capable of

---

out that Gestalt psychology is largely at odds with cognitivism. Whereas cognitivism is interested in the relationship between bottom-up and top-down processing, the pith of gestalt psychology involves irreducible perceptual wholes. As Houdé et al. say, "By focusing on the geometric and structural aspects of stimuli (→ MORPHOLOGY), gestaltists refuse the sensation/perception distinction stipulated by the associationists. They see an isomorphism between the structure of the stimulus and the corresponding percept, and from there, deduce an isomorphism with the underlying physiological mechanisms. Their position is necessarily globalistic and emphasizes the idea that subjects organize and structure (shape) the environment (contrary to the associationist idea of an environment that acts upon the subject)" (311).

eliciting retrieval of those 'farther facts,' which become incorporated into the percept" (Albright 32). Here "farther facts" is a term appropriated from the work of William James and refers precisely to the various facets of context that Albright is expounding for us.[17] Stated simply: *temporal contexts are learned associations*. They are a repository of sense impressions and prior experiences that inform and contextualize new percepts. This is what Albright means by "associative learning" which "results when stimuli appear together in time…The product of associative learning is that presentation of one stimulus elicits retrieval of its associate and all of the experience that that retrieval entails" (Albright 32). Temporal contexts can accrue over minutes or years and their assimilation allows one to perceive meaningful patterns where initially one encountered only perceptual noise.

Whereas spatial contexts are in a sense "baked into" the visual percepts, and evolutionary contexts are acquired through the adaptation to specific environments, temporal contexts depend on memory and learning, and as such, can be traced through a number of physiological processes in the brain. In neuroscientific terms, the process of learning can be reasonably be summed up by the axiom "neurons that fire together wire together," a phrase attributed to neuroscientist Donald Hebb (Armstrong 49).[18] Hebb's true postulate is as follows, "'When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change

[17] The quote, from James' *The Principles of Psychology, vol. 2* reads: "Perception thus differs from sensation by the consciousness of farther facts associated with the object of the sensation" (qtd. Albright 23).

[18] It's worth noting, however, that this is a common misattribution. As Keysers and Gazzola inform us, "Hebb himself never wrote 'Cells that fire together, wire together'. This mnemonic phrase was first introduced by Carla Shatz [12] in an article for the Scientific American aimed at a lay public" (Keysers and Gazzola 2).

takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased' ([11], p. 62)" (qtd. in Keysers and Gazzola 2). The key difference, according to Keysers and Gazzola has to do with antecedence: neurons do not need to fire simultaneously to reinforce their connection. All that is required is that one neuron repeatedly take part in the firing of another for the connection to become solidified. This principle is also known as "Hebb's rule" (Lazari et al. 1).

Hebb's rule, also called "Hebbian learning," would seem to extend many of the basic insights of behaviorism down into the cellular level. Behaviorism argued that all learning results from an individual's interaction with their environment, in which an environmental stimulus and a behavior become connected and reinforced over time. This is essentially the same logic at work in the Hebbian firing/wiring, but the changes that occur are metabolic and connective rather than behavioral. Habitual firing of neurons, and assemblies of neurons—as in response to a repeated stimulus—will create measurable and lasting changes in the circuitry of the brain and nervous system. Lazari et al. say that Hebb's rule, initially formulated as a computational principle, was soon "found to have a biological substrate in the synapse. Synapses can detect coincident activity of two neurons, i.e., detect when neurons 'fire together,' and effect plastic changes in the synaptic connections between them, i.e., make neurons 'wire together'" (Lazari et al. 1). In effect, Hebbian learning describes the precise neural mechanism by which learning takes place in the brain. In the 75 years since Donald Hebb advanced these views, these mechanisms have become refined and elaborated, but the basic dynamic remains unchanged: learning involves plasticity at the neural level, where repeated firing of a neuron influences the growth and development of nearby neurons, creating new connections, or strengthening existing ones.[19]

The "temporal context" described by Albright is precisely where the enactivist account of perception resides, but it does so by resisting the sensation/perception distinction on philosophical grounds. The enactivist approach is radically *nonlinear* in that it does not carve up the process into discrete domains in which a certain type of raw input (sensation) becomes a percept which then becomes a behavior (an output). Enactivism challenges us to understand these processes as coextensive and wholly interdependent.

Emergence of the Enactive Approach (in the Context of Cognitive Science):

Now that we have a brief history of cognitive science, its two major approaches (cognitivism and connectionism), and the philosophical positions that underpinned their success (computationalism, representationalism, and functionalism), at long last we arrive at the enactivist moment in the story.

As Thompson tells us, enactivism can be understood as a subset of the last major approach to emerge in cognitive science: embodied dynamicism. This approach rejects the computationalist assumption that mental life is radically "skull bound," separate from the world and its goings on, and merely a mode of "abstract representation: symbolic or

[19] As stated by Lazari et al., the primary "growth process" involved in the neuroplasticity of Hebb's rule occurs at the synapse—the site at which chemical or electrical signals pass from neuron to neuron. Recent research has also demonstrated that Hebbian learning and its attendant "growth processes" or "metabolic changes" also extend to the neuron's myelin sheathes—the insulation around the axons of neurons that facilitates the transmission of electrical signals across the cell membrane (Lazari et al. 7).

subsymbolic representations in the mind-brain [that] stand for states of affairs in some restricted outside domain that has been specified in advance and independently of the cognitive system" (Thompson 10). In other words, embodied dynamicism rejects the computationalist idea that representational content is produced in an abstract syntax or neural pattern that can be separated from the embodied and environmentally embedded context of cognition; it avers the constitutive role of the body in the creation of our mental life. Like connectionism and its recourse to cybernetics, the embodied dynamicist approach looks to the circular, self-organizing properties of systems. But unlike connectionism, it extends these circular causal loops beyond the nervous system and out into the body and its world. In *The Embodied Mind* Varela et al. articulate their foundational definition of enactivism as follows:

> We propose as a name the term *enactive* to emphasize the growing conviction that cognition is not the representation of a pregiven world by a pregiven mind but is rather the enactment of a world and a mind on the basis of a history of the variety of actions that a being in the world performs. The enactive approach takes seriously, then, the philosophical critique of the idea that the mind is a mirror of nature but goes further by addressing this issue from within the heartland of science. (Varela et al. 9)

"Enactivism," since its inception, was a term that sought to unify a burgeoning and disparate body of work united by a skepticism towards these "pregiven" features of mental life and how they embedded themselves in the institutional study of the mind. In their opening pages, Varela et al. discuss what they call a "fundamental circularity" (3).

They say, "we find ourselves in a circle: we are in a world that seems to be there before reflection begins, but that world is not separate from us" (3). This circularity defines the enactivist approach as an epistemology—it structures all that we can know about the mind and our experience. They continue, "From the standpoint of enactive cognitive science, this circularity is central; it is an epistemological necessity" (9). Epistemologically, neuroscience purports to stand outside this circle. It seeks to correlate changes in brain structure with changes in behavior and experience (i.e., neural correlates of consciousness), and asserts that this objective relationship between the two constitutes the cognitive domain. Enactivism points out that these descriptions are themselves a product of this cognitive domain. All scientific descriptions are bound up in this more fundamental cognitive "circle." This epistemological circularity is one of the key insights of twentieth century phenomenology and a defining feature of the enactivist approach in cognitive science. We need to emphasize that enactivism is not opposed to the materialism of mainstream neuroscience. Enactivist descriptions of sensorimotor (perceptual) processes recapitulate many of the empirical, physiological facts at the core of contemporary neuroscience, but they do so in a different idiom—one that remains open and amenable to phenomenological description, and thus to the larger epistemological implications therein. We need to emphasize this shared empirical foundation because we don't want to lose sight of the scientific underpinnings that inform many of its phenomenological descriptions.

One of the refrains of enactivist theory—and by extension most who adhere to the embodied dynamicist account—is that a *dogmatically* representationalist view needs to be rejected. Rather than scrap the representationalist paradigm root and branch, it needs to be qualified and integrated into a larger 4E picture. Alva Noë describes this revisionary impulse in the following way:

> No doubt perception depends on what takes place in the brain, and very likely there are internal representations in the brain (e.g., content-bearing internal states). What perception is, however, is not a process in the brain, but a kind of skillful activity on the part of the animal as a whole. The enactive view challenges neuroscience to devise new ways of understanding the neural basis of perception and consciousness. (Noë 2)

For many enactivists, *representations*, understood as intrinsic content-bearing states, can still play a role in cognition—indeed, there is a preponderance of neuroscientific evidence that demonstrates their importance—the key point is that it cannot be the *only* way of understanding thought, perception, and the like. Daniel Hutto puts the matter thus: “Enactivism of this stripe denies that *the most basic forms* of genuinely mental activity necessarily involve or are to be explained by, the manipulation of contentful representations” (Hutto 46; emphasis added). This is enactivism understood as a kind of sustained qualification rather than an outright repudiation of the larger neuroscientific orthodoxy. In this view, the decisive qualification that enactivism places on neuroscience is that representation cannot be understood as a *wholly* skull-bound activity, divorced from the active participation of the autonomous agent and his or her sensorimotor capacities, and furthermore, that the outside world is not disinterestedly available for cognition and waiting to be mapped, but instead emerges through various world-constituting acts.

I count myself firmly among this aforementioned qualificatory "stripe" of enactivism. In contrast to so-called "radical" enactivist approaches which reject all notions of mental representation, I believe that representation recapitulates a basic feature of biological life, and thus, a basic feature of cognition. I believe that we risk losing sight of the intentional and purposive (i.e., teleological) core of life and cognition by rejecting representationalism root and branch. This is a tension that we will explore in chapters two, three, and four, but particularly in chapter three via the work of Terrence Deacon.

At this point we need to disambiguate the relationship between "perception" and "cognition" for enactivist theory. For, as the title of Noë's book announces, it is "perception," *not* "cognition," that grounds his enactivist approach. Does this change in terminology signify a distinct facet of mental life? Or is Noë simply channeling a new term for the same process?

Some enactivists, like Noë, will argue for the primacy of "perception" because it grounds all subsequent (i.e., explicitly representational) cognitive activities in immediate environmental contexts and conditions, and it has been argued that such perceptual, embodied processes must by necessity constitute a large percentage of our conceptually explicit cognition.[20] For example, Capra and Luisi say:

> The mind's embodiment can easily be illustrated by our use of spatial relations, which are among our most basic concepts. As Lakoff and Johnson (1999, pp. 34-5) explain, when we perceive a cat 'in front of' a tree, this spatial relationship does not exist objectively in the world, but is a projection from our bodily experience. We have bodies with inherent fronts and backs, and we project this distinction onto other objects. Thus, 'our bodies define a set of fundamental spatial relations that we use not only in orienting ourselves, but in perceiving he relationship of one object to another.' (Capra and Luisi 272)

This strikes me as a reliable and wholly convincing inference regarding the nature of perceptual life and how representations can be said to emerge from these embodied depths. Noë will argue precisely for such a bottom-up determination of perceptual content.

Thompson, in keeping with the discourse of second-order cybernetics and Maturana and Varela's work on autopoiesis, adopts the language of *cognition* as a way to encompass the totality of an organism's mental activity. This difference also makes sense if we consider their respective scales of analysis. Noë is interested in the basal cognitive processes of human beings, hence his emphasis on "perception," while Thompson is

---

[20] I qualify this (i.e., "large percentage") because as we have seen from Albright, there are other ways in which perception gets coded into the human organism: spatial contexts (e.g., gestalts), and evolutionary contexts.

interested in the basal cognitive processes of organisms per se—of which his starting place is the single-celled bacterium—hence his emphasis on the language of "cognition." In the context of a single-celled microorganism it would be odd to talk about "perception" given the fact that there are no discrete perceptual faculties such as sight or hearing, only a cell membrane and the rudimentary sensorium that follows from this. Despite these variations in terminology, enactivism is united by its focus on total system states—that is, on how the organism *as a whole* interacts with its immediate environment.

Ultimately, I think these differences in terminology help signal the exigency of 4E theory and its relationship to contemporary neuroscience. Obviously, we cannot hope to advance our knowledge of the brain without analytical methods—i.e., breaking it down into its constituent parts and then devising clever experiments that either support or falsify a hypothesis. It is precisely this tension that emerges between the concepts of "perception" and "cognition" that allow us to interrogate many of the unexamined assumptions that are baked into our theorizing of mental life. Such tension is crucial for bridging the first-third personal descriptive divide.

So far in this project I have utilized the language of "cognition" because that is the language born from Matura and Varela's work on autopoiesis and the larger discourse of second-order cybernetics that emerged during the 1970s. This use of the term "cognition" is foundational for Thompson's enactivism because it claims to unify, implicitly, the autonomy of biological life with subjectivity, intentionality, and mind. Given this broad scope, in the context of systems theory "cognition" is often used as a catchall term for a wide array of mental phenomena (e.g., perceiving, thinking, intending etc.).[21]

Despite these terminological differences, in the opening line of Noë's *Action in Perception*, we can see echoes of the core cognitive dynamic outlined by Thompson and autopoiesis. He says, "The main idea of this book is that perceiving is a way of acting. Perception is not something that happens to us, or in us. It is something we do. Think of a blind person tap-tapping his or her way around a cluttered space, perceiving that space by touch, not all at once, but through time, by skillful probing and movement" (Noë 1). Here we can see the element of activity and agency "something we do," and the idea of ongoing-ness, "not all at once, but through time." Given this shared foundation, it might be tempting to read second-order cybernetics directly into Noë's enactivism, but such an interpolation would distract us from the most salient aspects of his inquiry—the uniquely human modes of perception. To this end, it's worth noting that the term "autopoiesis" does not appear at all in *Action in Perception*, and this omission speaks to my previous

---

[21] 4E theorists from across the discourse provide myriad frameworks for scaling up, extending, embedding and embodying the basic system-environment coupling which second-order cybernetics names *cognitive*. These theorists are also useful for descending from the (at times) diffuse and all-encompassing theoretical perspectives of Thompson, Varela, and company, who often extend their systems theoretical analysis to far-flung corners of the physical environment (e.g., bacterial chemotaxis), and to far-flung corners of intellectual history. Alva Noë and Andy Clark provide two such frameworks. They limit their discussion of cognition and perception to the human animal and its quotidian perceptual contexts, and ground their claims in a wealth of empirical, cognitive scientific studies. They also represent distinct facets of the 4E debate. Noë at least pays lip service to the notion that mental representations can play a part in cognition, but he leaves this firmly to one side in order to pursue the embodied and phenomenological consequences of enactivism. While Clark advances what has been called "'extended functionalism,' according to which the mind is the joint product of intracranial processing, bodily input, and environmental scaffolding…" a combination which is clearly amenable to a computational/representational framework (Newen et al. 9). These two theorists represent different poles within 4E cognitive theory, demonstrating that "the 4E approach as such does not presuppose a specific view on representation and computation…" (Newen et al. 9).

point: that Noë helps us descend from some of the more abstract theoretical heights and into the quotidian contexts of human sensorimotor perception.

Throughout the opening two chapters, Noë goes on to respond to many of the basic points of contention, particularly around visual perception, that we have already broached in this chapter between embodied and representational theories of thought. In chapter two, entitled "Pictures in the Mind," he traces a clear link between what he calls the "snapshot conception" and the recourse in cognitive science towards mental representations. Vision is the dominant sense perception, so it is easy to see why "when we try to understand the nature of sensory perception, we tend to think in terms of vision, and when we think of vision we tend to suppose that the eye is like a camera…to see, we suppose, is to undergo snapshot-like experiences of the scene before us" (Noë 35).[22] There is an immediacy and cogency to vision that can make it difficult to peel away its more minute and strange phenomenological aspects. Noë argues that "something like the snapshot conception provides the starting point for much empirical work on vision" (Noë 35) and that the basic problematic vision science faces is how to reconcile the cogency of visual perception—its "richly detailed high-resolution visual experience"—when the sensations recorded by our retinas are in reality so limited (Noë 36)?

The answer for cognitive science has traditionally been that the brain "fills in the gap in the internal representation of the scene" (Noë 38). A clear example of this kind of

[22] It's worth pointing out that this quotation echoes some of what Cain says regarding perceptual experience: "I open my eyes, orient them to the world and have a perceptual experience without having any awareness of executing an extended process the earlier elements of which justify the later elements" (Cain 10).

"filling in" is our failure to perceive the optic blind spot—the place on the retina where the optic nerve connects to the eye, and on which no photoreceptors exist. The result is that, physically speaking, there *should* be a blind spot, but because we do not perceive it (it only reveals itself through clever visual experiments) we erroneously conclude that our brain *fills it in*. Noë says, "orthodox visual theory in this way frames its central problem as that of constructing an internal representation sufficient to support our detailed, high-resolution, gap-free, snapshot-like…visual experiences of the world despite the imperfections and limitations of the retinal image itself" (Noë 39). Despite the fact that, "no contemporary theorist believes that we see by seeing internal pictures" (46), as this would be a gross oversimplification, "to assume filling in occurs in the absence of this evidence is to commit the homunculus fallacy" (47). The conclusion Noë draws is that visual theory continues to smuggle in homuncular explanations, which depend either directly or indirectly on "pictures in the head, what theorists today might describe as internal neural structures that are spatially or topographically isomorphic to that which they represent" (Noë 48)[23].

---

[23] Dennis Bray provides us with a helpful gloss of this snapshot/pictorial model criticized by Noë, one which has been updated into computationalist/connectionist terms as internal neural structures that are spatially and topographically isomorphic. Bray says: "Vision, for example, is not due to an actual image projected onto the brain as if on a screen. The visual scene is actually fragmented into different features and then put back together again. Light enters the retina, where it activates up to 120 million photosensitive cells (rods and cones). These are reduced to about 1 million axons in the optic nerve before being carried to the primary visual cortex. There the signals are broken down into separate features—orientation, color, movement—before eventually being recombined. The internal 'view' thus created is based on a hypothesis of what the outside world is like and can be easily duped—for example, by optical illusions. Note that light itself never enters the brain. No matter how dramatic the visual experience, it is completely dark in there—just pulses of electricity moving along axons" (Bray 239). Obviously, no contemporary theorist would suggest that we see via *actual* pictures in the head. The point is that these pulses of electricity moving along axons come to represent the outside world by retaining its features via formal (isomorphic) equivalence. In this view, the world is objectively *out there*, waiting to be transduced and represented by the brain.

In the snapshot conception, "the relation between moving and perceiving is only instrumental. It is like the relation between the lugging around of a camera and the resulting picture. The lugging is preliminary to and disconnected from the photograph itself" (Noë 2). In this view, the image recorded on our retina (the photograph) is totally uncoupled from the rest of our body (the lugging of the camera). Noë's goal in *Action in Perception* is to develop enactivist theory enough to account for this implicit disconnect between moving and perceiving, and to demonstrate how our perception of the scene depends on the myriad ways that the retinal image changes as a result of movement. This is what Noë calls *sensorimotor dependencies*, and at other times, *sensorimotor contingencies*. Noë's purpose is to disabuse us of the snapshot conception and its attendant homuncular baggage by explaining our visual experience via these sensorimotor dependencies and bodily affordances.

By foregrounding the role of sensorimotor dependencies in perception, Noë's sensorimotor approach emphasizes perception as a space of open-ended space of bodily possibility rather than a static space reproduced in representations or internal computations, which as we've seen are highly disembodied processes. Furthermore, it gives pride of place to the concrete environment that is ready-to-hand because it is from this immediate coupling to the environment that one first becomes grounded in the sensorimotor loops that ground cognition and agency.

For his part, Thompson distinguishes Noë and O'regan's sensorimotor approach in the following ways: in contrast to behaviorism, "perceptual experience, being in part constituted by endogenous knowledge (skillful mastery), mediates between sensory stimulation and motor behavior" (256). Recall that behaviorism eschewed interiority (i.e., perceptual *experience*), and examined only what could be concretely measured and manipulated from without, such as the conditioned stimuli of the experiment or a subject's responses. It is a straightforward sensory input—behavioral output model, and no subjective experience is required to explain the behavior. This was its biggest strength, and before long, its most glaring weakness.

In contrast to cognitivism, "experience does not intervene in a linear causal fashion between sensory 'input' and motor 'output.' Sensory stimulation does not *cause* experience in us, which in turn *causes* our behavior, because 'skillful activity (consisting of behavior and sensory stimulation) *is* the experience' (O'regan and Noë 2001b, p. 1015)" (qtd. in Thompson 256). This last observation may seem a bit counterintuitive, but it gets to the heart of system-environment couplings we have been espousing all chapter. Recall the perception circuit that Albright expounded for us. Sensation activates a perceptual process by which we refer the sensation to the world outside ourselves and to the knowledge (i.e., the various contexts) within. Notice how this view still seems to break up perception into an input/output model, only now the representational and computational "experiences" of a subject intervene between sensation (input) and behavior (output). In the enactivist, sensorimotor approach "perceptual experience emerges from continuous and reciprocal (nonlinear) interactions of sensory, motor, and cognitive processes, and is thereby constituted by motor behavior, sensory stimulation, and practical knowledge" (Thompson 256). Here *nonlinearity* refers to a description of the overall state of the system, a technical term signifying that a system's outputs are not directly proportional to their inputs, i.e., they cannot be reduced to any one set of inputs. All are determined with respect to each other (reciprocally) and are constantly undergoing changes. The important takeaway here is that no one part of the process can be uncoupled from any other. For enactivism, the totality of this circuit *simply is* the experience we call perception.

Thompson concludes his examination of Noë and company by arguing that the dynamic sensorimotor approach should be understood "not as an attempt to close the comparative explanatory gaps in a reductionist sense, but instead as an attempt to bridge these gaps by deploying new theoretical resources for understanding perceptual experience and neural processes in a coherent and overarching sensorimotor framework" (Thompson 257). These passages help us to further categorize Noë's sensorimotor approach in relation to Thompson's autopoietic enactivism. "Cognition" represents a totality of involvements, all the behavioral and perceptual activities performed by an organism per se; while "perception" represents a subset of that idea and refers specifically to the embodied cognitive activity of organisms with nervous systems.

The Systemic Brain:

As stated in the introduction, this dissertation adopts a systems theoretical approach. Systems theory is about placing objects and actors in larger contextual wholes. As Capra and Luisi tell us, "…all natural phenomena are ultimately interconnected…their essential properties, in fact, derive from their relationships to other things" (Capra and Luisi 2). This means that at all levels of analysis and description—and *regardless* of the particular object in view (e.g., a brain or a part of a brain, a body or a part of a body, a bacterium, an animal, a speaking subject etc.)—I will attend to the ways in which these objects/actors are defined by their relationship to other things, and to larger networks of interaction in which these assemblages participate. Therefore, what I hope to accomplish in this section is to articulate how the brain fits into this systems approach.

Stated simply, the brain is a dynamic system: its parts reciprocally determine one another, forming, dissolving, and reforming myriad neural assemblies that can respond to a changing environment by successfully coordinating various patterns of perception, action, and thought.[24] Neural assemblies are "large scale coalitions of neurons that operate in collective activity over a [sic] spatio-temporal scales of millimeters and milliseconds, i.e., at "meso-scale" level of brain organization thereby linking cellular (micro-scale) and entire neuronal systems (macro-scale) events (Eytan and Marom, 2006)" (Badin et al. 1). Or stated more simply, "a distributed subset of neurons with strong reciprocal connections" (Thompson 331). Neural assemblies are dynamic patterns of nerve cells that fire together, and as Badin et al. indicate, these assemblies occupy an intermediate scale of brain activity, one that effectively mediates between the firings of individual neurons and the integration of macro-scale neuronal systems. Neural assemblies determine the kinds of operations and information processing that a given region of the brain can perform. Neuroscientist Richard Passingham says, "This is the principle of 'localization of function', and it can hold because each area receives a unique pattern of connections and sends out a unique pattern of connections to other areas"

[24] All systems are dynamic systems. All connections and interconnections between parts and their larger wholes are ongoing and processual. Modifiers such as "dynamic" or "static" simply denote whether or not time is considered in the description of the system.

(Passingham 16). The incoming patterns of activation determine what kinds of information a given brain area can process, and its outgoing patterns determine what other neural processes this area can influence.

It is now a well-established fact in cognitive neuroscience that different regions of the brain are responsible for organizing and analyzing different types of sensory information, and furthermore, that the global properties of these various systems—for example, the unified character of our visual field—are themselves composed of myriad subsystems or pathways. Thompson says, "There is now little doubt in cognitive neuroscience that specific cognitive acts, such as the visual recognition of a face, require the rapid and transient coordination of many functionally distinct and widely distributed brain regions" (Thompson 330). *Transient* in two senses (a) because it takes time for the action potentials of individual neurons to fire and for larger neuronal assemblies to form their distinct patterns and (b) because these assemblies are perpetually superseded by new ones. In sum, different regions of the brain are responsible for specific types of information processing (e.g., somatosensory, visual, auditory etc.) and each type of information processing can be composed of many different neural pathways.

Passingham elaborates on this dynamical organization via the notion of *parallel processing*. He says, "…the brain is not a single pathway from input to output…Information from vision, hearing, touch, and smell is relayed from the sense organs to separate regions of the brain. These are referred to as the primary sensory areas. The information is then relayed through a series of secondary areas" (Passingham 13). This hierarchical, yet interdependent organization shows that "…there are separate or 'parallel' pathways," meaning that "different operations are carried out simultaneously: we see at the same time as we hear. So the brain is unlike a desktop computer in which the instructions are followed one after another in a series. Instead the brain has a parallel architecture…" (Passingham 15).

A good example of the compartmentalization and integration of different neural processes in this parallel manner can be seen in the "dual-stream hypothesis" of visual processing propounded by David Milner and Melvyn Goodale. Their research has distinguished "two largely autonomous visual systems" known as the ventral and dorsal streams (Noë 11). The dorsal stream is a "semiautonomous processing stream…that guides fluent motor action in the here and now" (Clark 182). This is because the dorsal stream is the only visual pathway to have "direct connections with the areas in the frontal lobe that control movement" (Passingham 14). In other words, because of its direct connections with the part of the brain responsible for motor control, this subsystem forms a dynamic and reciprocal interconnection that enables humans to integrate visual stimuli and bodily movement into a unified process of visuomotor control. In contrast, the ventral stream involves "conscious visual awareness…geared toward enduring object properties, explicit recognition, and semantic recall" (Clark 181-182). Whereas the dorsal stream can enable seamless and subconscious realization of visuomotor tasks, the ventral stream has been shown to operate in the total absence of visual stimuli via visual imagination and recall (Clark 182).

These insights demonstrate the dynamic, multilayered, and functional organization of the brain. Passingham says, "The concept of a system is fundamental in cognitive neuroscience. Though areas differ in the operations that they perform, they do not operate in isolation. To put it bluntly, no behavioural task depends on a single area of the brain…*anatomical systems support functional systems*" (Passingham 29-30, emphasis added). This is a crucial point for understanding the brain as a dynamic system. An anatomical system refers to "areas that are closely interconnected" (29), meaning that the spontaneous activity of different neural assemblies in these areas is closely coupled: they exert a well-defined physiological (Hebbian) effect on one another. In contrast, a functional system is a product of several integrated anatomical systems and is defined by the roles (or functions) being performed.

For example, in the case of visuomotor skills discussed above, it is the anatomical interconnection between the dorsal stream of visual processing and the area of the frontal lobe associated with motor skills that enables the overall *functional* ability of visuomotor skill. Thus, functional systems (e.g., visuomotor skill) depend on the brain's anatomical systems, but are not reducible to them. This dynamic and functional organization makes the brain an extremely flexible and adaptive organ. Not only can it respond effectively to a huge array of novel stimuli, but it can even overcome the effects of physical trauma by "rewiring" itself and redistributing its functions across different parts of the brain, a physiological mutability known as *neuroplasticity* (Costandi 2).

Dynamic systems theory and the role of "emergence" will be a central concern in chapter three, but to conclude this chapter, I want to explicate one of the most decisive ways that enactivism, in particular, has contributed to contemporary neuroscientific practice and helped realize its goal of integrating first-personal experience with third-personal descriptions: neurophenomenology.

Neurophenomenology: Discovering the Mechanism of Time Consciousness

As usual, Capra and Luisi provide us with an extremely lucid description. They say:

> The basic premise of neurophenomenology is that brain physiology and conscious experience should be treated as two interdependent domains of research with equal status. The disciplined examination of experience and the analysis of the corresponding neural patterns and processes will generate reciprocal constraints, so that research activities in the two domains can guide one another in a systematic exploration of consciousness. (Capra and Luisi 263)

Neurophenomenology is perhaps the clearest and most concerted contemporary scientific methodology aimed at uniting subjective experience and objective description, and as we will see, it is particularly well-suited to studying the "temporal dynamics of conscious experience" (Thompson 312). Thompson even goes so far as to call neurophenomenology an "offshoot of the enactive approach" (Thompson 312). This is because it begins by acknowledging the epistemological circularity that confronts any inquiry into the workings of mind. As the quotation by Capra and Luisi attests to, neurophenomenology gives equal pride of place to first-personal experience and uses it as a means of constraining and guiding neuroscience and its search for neural correlates of consciousness.

Thompson says, "A crucial feature of this [neurophenomenological] approach is that dynamic systems theory is supposed to mediate between phenomenology and neuroscience" (Thompson 329). Once again we are told that it is only by understanding and analyzing the brain as a dynamic system that researchers can ever hope to fulfill the ambitions of integrating first-personal subjective experience with the brain-imaging and mapping data that constitutes most of neuroscientific research and practice. Dynamic systems theory functions as a crucial mediator between phenomenology's first-personal descriptions of subjective experience, and cognitive neuroscience's third-personal descriptions of the brain (Thompson 331). Thompson says:

> Experimental neurophenomenology is guided by the theoretical proposal…that the most promising current candidate for the neurophysiological basis of consciousness is a flexible repertoire of dynamic large-scale neural assemblies that transiently link multiple brain regions and areas…In this approach, the framework of dynamic systems theory is essential for characterizing the neurophysiological processes relevant to consciousness. (Thompson 340)

Without question, the most important work accomplished by neurophenomenology has involved the study of time consciousness. Because the brain is a dynamic system connecting many disparate brain regions and functions via parallel processing, conscious awareness/experience and its underlying processes of perception and action are not always in sync. Armstrong says, "The neurobiological processes underlying our experience of time are somewhat elusive, but the scientific consensus is that there is no central clock or timing mechanism in the brain, just as there is no single language module and no 'homunculus' functioning as a central processor" (Armstrong 56). The dynamic fluctuations of the brain's neural firings (i.e., the myriad ways and means by which different senses are connected and integrated) often create discontinuities in our subjective temporal experiences.

These discontinuities begin at the neuronal and synaptic levels. Citing a 2014 book chapter by Dean V. Buonomano entitled "The Neural Mechanisms of Timing on Short Timescales" Armstrong says, "the responsiveness of neurons is 'strongly dependent on their recent history of activity,'…thereby providing 'an ephemeral memory of what has happened in the past few hundred milliseconds'" (qtd. in. Armstrong 56). Armstrong continues: "oscillatory coupling and decoupling of neuronal assemblies across the brain and between the brain and the body provide further mechanisms for processing durations at different scales" (Armstrong 56). Here Armstrong elaborates on the basic dynamical organization outlined above. Given the brain's dynamic organization, the "oscillatory" (meaning periodically synchronous/asynchronous) coupling and decoupling of neural assemblies can occur at many different timescales. This means that the senses, and our conscious awareness of their output (percepts), operate at different timescales. For example, we can successfully swerve our cars to avoid hitting a child who has run out into the middle of the road several milliseconds before we consciously recognize the fact that it was a child we avoided (Armstrong 60).

Investigation into the temporal dynamics of neuronal assemblies is work that Francisco Varela helped pioneer. Armstrong says, "Francisco Varela, the foremost neurophenomenologist of time, proposes three 'scales of duration' to distinguish different 'windows' of integration…" (Armstrong 56). Here *integration* refers to the process of coupling different neural assemblies. These are (1) the elementary scale (also known as the 1/10 scale), corresponding to between 10-100 milliseconds; (2) The integration scale (also known as the 1 scale), which varies from .5 to 3 seconds; (3) the narrative scale (also known as the 10 scale) which can last anywhere from a few seconds of one's working memory to longer term syntheses (Armstrong 56-57; Gallagher and Zahavi 84).

These measurements indicate the distinct periods of time during which different neural assemblies synchronize. As Gallager and Zahavi note, "Although Varela characterizes these different scales in objective measures of milliseconds and seconds, there is no set or rigid completion time or fixed integration period. The integration window is necessarily flexible depending on a number of factors…" (Gallagher and Zahavi 85). For example, the 1/10 scale represents, "the minimum amount of time needed for two stimuli to be consistently perceived as non-simultaneous, a threshold which varies with each sensory modality" (Gallagher and Zahavi 84). It is crucial to establish a minimum threshold for non-simultaneous perception because experimenters need to be able to distinguish between distinct perceptual events that can be registered by their subjects, and that can serve as the basis for integration into the 1 scale.

Accordingly, the 1 scale corresponds to our perception of the "now" (Thompson 333). As Thompson points out, the idea that "the integration-relaxation processes at the 1 scale are strict correlates of present-time consciousness" (333) are not merely correlative, they are meant to be *causal* (334). Thompson says, "The aim is to explain how the temporal structure of experience is caused by and realized in the dynamic structure of biological experience" (334). *This shift from correlation to causation is crucial because this is how neurophenomenology bridges the gap between subjective experience and objective empirical description*. The working hypothesis is that these ever-shifting patterns of neural assemblies are what actually cause our subjective experience of the present moment—the now.

Finally, the 10 scale represents the long-term synthesis of neural assemblies such as memory formation and codification. Gallager and Zahavi explicitly describe the 10 scale as "the narrative scale" which attests to the inherently narrative nature of cognition, and which will serve as an important premise for the final chapter of this dissertation.

Varela's work on neurophenomenology demonstrates this synthesis of dynamical systems theory and neuroscience because "cognitive time is not based on any external or internal uniformly ticking clock, but rather arises from an endogenous and self-organizing neurodynamics...." In one of the most bold and intriguing moments of synthesis across disciplines, Thompson tells us that "...According to Varela, this dynamics can be described as having a retentional-protentional structure" (Thompson 335). Varela related this dynamic process of neuronal assembly-disassembly-reassembly to Husserl's original description of time consciousness, in which "ordinary acts of perception carry along with them moments of the past (retentions) as well as anticipations of the future (protentions)" (Moran 138). Thus, not only does this contemporary neurophenomenological work meaningfully bridge first and third personal modes of inquiry, but it also hearkens back to, and legitimates, one of the conceptual foundations of phenomenology.

Conclusion:

In this chapter we have examined the mind-as-computer metaphor that both kick-started the "cognitive revolution" and furthered the explanatory gap. We have unpacked the computational and representational underpinnings of cognitive science and seen how these conceptual and logical assumptions helped to either bracket from consideration, render suspect, or repudiate outright the first personal *experiences* that make our cognitive lives meaningful. This does not mean that we scrap quantitative, reductionist, or analytical practices root and branch. Instead, we need to qualify and ameliorate this reductionism by weaving enactivism its recourse to phenomenology into our understanding of mind. Enactivism provides us with the conceptual and methodological tools to recover the first personal experiences and intentionality that define mental life: we are neither computers nor programs but bodily wholes whose cognitive activities emerge through dynamic interaction with the world. In the following chapter we will examine how the explanatory gap insinuated itself in the discourse of biology and its understanding of organisms.

Chapter Two:

Systems Theory and The Beginning of Ends

> Without question, phenomena such as life and mind owe some of their mysterious character to limitations in our present state of science. I'm confident that these limitations of our current theoretical tools can be overcome and that these phenomena can also become integrated into the larger fabric of natural science. The question is whether in accomplishing this, their distinctive ententional characteristics (function, representation, end-directedness, self, and so on) will be explained rather than merely explained away.
>
> —Terrence Deacon, *Incomplete Nature*

> There seems to be an intimate fear that the awe with respect to life and the living would disappear if a living system could be not only reproduced, but designed by man. This is nonsense. The beauty of life is not a gift of its inaccessibility to our understanding.
>
> —Maturana and Varela, "Autopoiesis: The Organization of the Living"

In keeping with Levinas, I believe that ethics is defined by an experiential encounter with a *face*—that is, the outward expression of an interiority and autonomy that calls us beyond the bounds of solipsism and into an ever-broadening circle of care and existential vulnerability. The central claim of autopoiesis is that *all life is cognitive*; all living processes are cognitive processes. As Humberto Maturana unequivocally tells us, "A cognitive system is a system whose organization defines a domain of interactions in which it can act with relevance to the maintenance of itself, and the process of cognition is the actual (inductive) acting or behaving in this domain. *Living systems are cognitive systems, and living as a process is a process of cognition*" (Maturana 13). Cognition, so conceived, is a systems-level phenomenon that is coextensive with biological life. Cognition is defined precisely by the interiority and autonomy, mutatis mutandis, as the ethical encounter. It stands to reason then, that all biological life enmeshes us in an ethical relation, and there is no living thing that we do not owe an ethical obligation to. This is the radical claim subtending my exploration of autopoietic enactivism, and one which I believe fundamentally advances the aims and ambitions of the environmental humanities.

Autopoietic enactivism fundamentally encompasses two disciplines: cognitive neuroscience and biology. Chapter one addressed its relationship to the former. This chapter will address its relationship to the latter. We cannot understand the full scope and import of autopoietic enactivism without addressing both sides of this disciplinary equation. Therefore, this chapter begins by constructing a kind of historical systems theory "biopic," and its emergence from the biological sciences. This will provide essential coordinates for understanding how the second-order cybernetic systems theory that is autopoiesis, and two of its most interesting theoretical offspring—Thompson enactivism and Deacon's emergent dynamics—will attempt to reconcile the aforementioned explanatory gap between subjective and objective modes of description.

In their book *The Systems View of Life*, physicist and systems theorist Fritjof Capra and biochemist Pier Luisi furnish us with an overarching metaphor—a pendulum—which signifies the back-and-forth movement evinced in our collective understanding of organisms and living systems—between the poles of mechanism on the one hand, and holism on the other. In keeping with this metaphor, this chapter will oscillate between watershed moments in this trajectory. On the mechanistic side of the aisle, I will examine the work of Claude Shannon and his "information theory," as well as the so-called "neo-Darwinist" school of evolutionary biology which Shannon's theory helped facilitate and which has tended to bracket from consideration the organic wholeness of the organism via its recourse to Darwinian natural selection and modern genetics. I will alight on one important theorist from within this tradition—Daniel Dennett—and examine a direct response he wrote to Thompson's book *Mind in Life*. Next, on the holistic side of the aisle, I will examine developmental systems theory (DST) which Thompson convincingly offers as enactivism's holistic answer to the reductionist tendencies at work in neo-Darwinism and beyond. Together these oscillations between information/neo-Darwinism and DST/enactivism will show how materialist conceptions of "information" insinuate themselves into the discourse of biology, effectively recreating the cognitivist position we encountered in chapter one, but this time transposed into our collective understanding of organisms and evolutionary processes.

In the last half of the chapter, we will turn directly to the autopoietic-enactivist urtext: Maturana and Varela's essay "Autopoiesis: The Organization of the Living" from their 1980 book entitled *Autopoiesis and Cognition*. This section is titled "Autopoiesis and its Discontents" in order to signal the theory's vexed relationship to teleology and mainstream life science research. I end the chapter by examining Thompson's "body-body" problem, which is his autopoietic-enactivist response to the explanatory gap, and one which tries to synthesize both the cognitive scientific and biological exigencies we have been tracing in these first two chapters.

Everything is a System:

In laying out the coordinates for this systems theory "biopic," I will be drawing from three of my primary intertexts in this dissertation: Capra and Luisi, Deacon, and Thompson. The goal of this section is to define a system and its associated concepts and terminology, and to map key moments in the development of modern biology and the various ways it has tried to come to terms with the problems posed by teleological/purposive systems.

In their introduction to *The Systems View of Life* Capra and Luisi frame the reemergence of systems thinking with the following foundational premise: "Twentieth-century science has shown repeatedly that all natural phenomena are ultimately interconnected, and that their essential properties, in fact, derive from their relationships to other things" (Capra and Luisi 2). They trace the origins and development of scientific practice in the West through the metaphor of a pendulum, arguing that at various moments in its history science has oscillated between the study of quantities and constituents, to the study of patterns and relationships. Each swing of the pendulum can be characterized by a guiding question. On the mechanistic side one can ask, "what is it made of?" and on the systemic side, "what is its pattern?" They say, "The basic tension is one between the parts and the whole. The emphasis on the parts has been called mechanistic, reductionist, or atomistic; the emphasis on the whole, holistic, organismic, or ecological" (Capra and Luisi 4). They argue that presently we are in the midst of a reemergence into systems thinking; that we can see this renewed interest in patterns and relationships at work in the most fundamental subatomic levels described by quantum physics, in the largest levels of Earth Systems Science, and most importantly for our purposes, at the level of the brain and nervous system.

Simply put, a *system* is an integrated whole whose properties depend on the relationship between its parts (Capra and Luisi 64). Systems are defined by their boundaries, and a boundary is defined by a delimited organizational pattern. These two notions—boundary and pattern—are interrelated. In the case of a simple thermodynamic system, for example a cup of coffee, we are confronted with an obvious physical boundary—the cup—that contains the hot liquid within. But notice that in defining this boundary we still must make reference to the relationship between its parts (the liquid and the container) and specify how they interact. An important aspect of systems thinking more generally is the notion of *irreducibility*; while we can and do talk productively about discrete parts of systems, irreducibility signifies the primacy of the whole in relation to its parts. Stated simply, systems theory is always in the business of taking reductionist insights and reintegrating them back into the context of the larger whole.

In the introduction to "The Biology of Cognition," Humberto Maturana describes the role of the observer as follows: "The basic cognitive operation that we perform as observers is the operation of distinction. By means of this operation we specify a unity as an entity distinct from a background, characterize both unity and background with the properties with which this operation endows them, and specify their separability"

(Maturana xix). Whether it is a simple cup of coffee, the motions of planets around the sun, or the interpretation of a work of literature, every system is defined by its boundary, and these boundaries are made explicit by specifying and delimiting its pattern of organization.

Systems theory distinguishes two generic ways in which a system's boundary can be specified. *Heteronomous systems* are systems whose boundaries are determined from without. Our cup of coffee is a case in point. There is a rather obvious material boundary that impresses itself on the observer, but there is nothing in principle preventing one from expanding this system to include the human hands holding the cup, or the room surrounding the hands that hold the cup. The point here is that the boundaries of a heteronomous system are mutable precisely because they depend on the observer. This mutability does not prevent the practice of sound science. For example, the two main criteria used to define the solar system—the farthest reach of the sun's gravity and the farthest reach of the sun's solar winds—demarcate ~~very~~ different boundaries, but this variation in the formal description of the solar system has not prevented the forward march of astronomical study. This is not to say that the boundaries of natural systems are *arbitrary*, but they can be differentially defined. If anything, such mutability has enriched our overall conception of the solar system.

In contrast *autonomous systems* are systems whose boundaries are "defined by its endogenous, self-organizing and self-controlling dynamics…" (Thompson 43). In an autonomous system, boundary and pattern (i.e., *systemic identity*) are generated from within. The paradigm case of an autonomous system is biological life. It is a self-generated locus of causal processes. In their original work on autopoiesis, Maturana and Varela use the terms *autopoietic* and *allopoietic* to describe autonomous systems and heteronomous systems respectively. In their 1992 book *The Tree of Knowledge* they assert, "We are not proposing that living beings are the only autonomous entities…We are proposing that the mechanism that makes living beings autonomous systems is autopoiesis. This characterizes them as autonomous systems" (48). By describing autopoiesis as *the mechanism* of autonomy their purpose was quite literally to ground this description in a mechanistic explanation. Autopoiesis is the mechanism (procedure) by which autonomy and cognition are realized.

It is the autonomous and self-organizing character of cognitive systems that gives living things their quintessential purposiveness and end-directedness, what has traditionally been called its *teleological* organization, "literally, the logic of end-directedness" (Deacon 24). One of the most important formulations of this problematic was Kant's critique of teleology in the *Critique of Judgement* in which he distinguishes between an artifact and a natural purpose. In the case of artifacts, we are confronted with part/whole relations but "This idea or concept [of the whole] resides outside the entity in the mind of an intelligent designer" (Thompson 134). A watch, for example, is a system whose purpose (keeping time) is determined by the watchmaker. This heteronomy means that every single feature of the watch's organization can be exhaustively defined and mechanistically explained: there is no inscrutable organizing principle that needs to be accounted for.

But an organism is a natural product, and "…the cause of its production resides within it, as a function of its autonomous organization. An organism is self-producing and self-organizing, for each of its parts reciprocally produces the others" (Thompson 134). This means that in order to understand the workings of an organism, we "must take the idea of the organism *as the ground of our cognition of it*. That is, we must view it as the basis for our ability to cognize it as an organized being and to grasp its structure and workings in a unified way" (Thompson 135). This is the moment in which purposes, ends, or constitutive norms are smuggled into the description of life's organization, and consequently, demonstrates the deep—and as we will see, problematic—interconnection between autonomy and teleology. Maturana and Varela describe this tension as follows: "Machines are generally viewed as human made artifacts with completely known deterministic properties which make them, at least conceptually, perfectly predictable. Contrariwise, living systems are *a priori* frequently viewed as autonomous, ultimately unpredictable systems, with purposeful behavior similar to ours" ("Autopoiesis…" 83). The qualifier "a priori" recapitulates this idea that we assume or infer something essential about living organization which is incapable of empirical description. The norms that govern an organism's processes of self-maintenance and self-determination are endogenous, and "the parts are related to each other reciprocally as end and means…" thus "…We cannot explain this reciprocal relation as a necessary consequence of matter in motion" (Thompson 135).

The autonomous and purposive character of living systems has remained a recalcitrant thorn in the side of modern science, particularly biology, because it is at odds with the materialist principles of scientific inquiry. Purposes and ends are materially absent causes: they cannot be measured, quantified, or reduced to component parts. How is it that science can account for the animate unity and dynamism of living processes when we know, at bottom, that everything must be made of inanimate *stuff*? How is it that organisms grow and develop towards well-defined terminal states of being such as the acorn to the oak? How to explain complex behavior in animals when this behavior so profoundly exceeds what is observable and measurable by physics and chemistry alone? How to account for the abstract imaginings of human consciousness or the complex linguistic signifying structures that characterize our intersubjective interactions? These represent big problems for the reductionism and materialism inherent to the scientific method. In all these phenomena we are confronted with a natural purposiveness and end-directedness that is assumed in the observer's conceptualization of a living system, and because it is assumed, this purposiveness escapes rigorous empirical description.

Despite the fact that living things confront us with something inherently teleological and holistic, in 1859, the publication of Charles Darwin's *On the Origin of Species* baptized modern biology on mechanistic lines. Darwin's theory of evolution by natural selection provides a *non*-teleological explanation of life and is without doubt the most robust and sacrosanct scientific theory at work in our modern world. There are three features to Darwin's theory (1) reproduction and inheritance of traits (2) spontaneous variation of those traits, and (3) reproduction in excess of what a local environment can sustain, which creates competitive pressures among organisms. These three features are all that are required to explain the highly complex and well-adapted biological forms that pervade the natural world. As Deacon tells us, "The core distinguishing feature of Darwin's explanation—and what made it so revolutionary—was…the *after-the-fact* logic of this mechanism" (Deacon 111). Evolution by natural selection did not appeal to a logic of ends. According to Darwin, the acorn becomes the oak because of countless generations of inheritance, competition, and selection that shape an organism's individual development (ontogeny) and species-specific morphological features (phylogeny) after-the-fact—a kind of winnowing process that simply preserves what succeeded in a particular environmental niche without any precise "end" in view. Darwin's theory of natural selection explains how inheritance and variation of traits and competitive pressures can lead to the spontaneous (i.e., natural) selection of the most advantageous of those traits, and thus the adaptations we see preserved in phylogenesis.

As Deacon argues, natural selection "is not exactly a mechanistic theory…" insofar as it does not trace the precise actions of "matter in motion," i.e., the precise series of efficient causes in which x gives rise to y gives rise to z. Deacon continues, "…it can best be described as a form of statistical inference that is largely agnostic about the mechanisms it depends on" (Deacon 151). Given the theory's reliance on the kind of holistic interactions between multiple members of an ecological community, we can also understand it as a kind of proto-systems approach: it is concerned with the interactions *between* parts (i.e., members of a lineage, an ecological community etc.) and with the morphological effects that such long-term (dynamical) interactions produce. Despite this focus on process and interaction over time, Darwin's theory posits no overarching organizational purpose or end. This was the decisive move for modern biology, and its ripple effects were felt across disciplines, especially those which sought naturalized grounds for the development of complex forms and patterns in nature.[25]

By the turn of the twentieth century "cell theory, embryology, and microbiology – had established the mechanistic conception of life as a firm dogma among biologists… [but] it remained largely ignorant of the coordinating activities that integrate those operations into the functioning of the cell as a whole (Capra and Luisi 63). Capra and Luisi cite this failure to integrate the cell's increasingly well-defined constituent parts into a functional whole as the impetus that gave rise to *organismic biology*.

Early organismic biologists understood that living things were not decomposable phenomena, but in order for their approach to remain scientifically valid they needed to avoid appealing to some intangible "vital force" that would animate living things and so relieve them of further explanatory burdens, a long-standing scientific and philosophical approach known as vitalism. To walk this tightrope, they appealed to "organizing relations [which are] patterns of relationships immanent in the physical structure of the organism. Later on, the concept of organization was refined to that of 'self-organization,' which is still used in contemporary theories of living systems" (Capra and Luisi 64). As we have already encountered in our exploration of cognitive science, the dynamic processes at work in the

---

[25] For example, Karl Marx and Fredrich Engels were both deeply impressed by Darwin's anti-teleological rewriting of natural history and cited him as an ally in *their* respective rewriting of the laws of historical development, and its concomitant political and economic forms (e.g., feudalism, capitalism, communism). In a letter from Engels to Marx dated shortly after the publication of Darwin's book (December 12, 1859), Engels writes: "All in all, Darwin, whom I am reading right now, is superb. Teleology had hitherto not yet been destroyed in one of its aspects, and now this has been done. Moreover, hitherto there has never been so sweeping an attempt to prove historical development in nature, especially with such success" (qtd. in Zavadskii et al. 63).

human brain evince self-organization. The brain and the nervous system constitute a self-organizing system that gives rise to the emergent phenomenon of mind.

The systems theories advanced by Maturana and Varela, Capra and Luisi, Thompson, and Deacon all acknowledge that the reductionism of modern science is "a powerful tool for breaking up the work involved in the exploration of the complex system that is an organism" but as Deacon says, it also "precisely brackets from analysis what is most relevant: the 'organic wholeness.' The life of an organism is not resident in its parts. It is embodied in the global organization of the living processes" (Deacon 135). Life is an emergent phenomenon: it possesses an irreducible wholeness that strict reductionist approaches cannot capture.

Yet as Capra and Luisi's metaphor of the pendulum attests to, every inroad into a systems-based explanation gets counterbalanced by ever-more creative ways of mechanistically reducing the explanation. Thompson says, "In the twentieth century, classical Darwinism was transformed into what is sometimes called neo-Darwinism, first during the 1930s as a result of the so-called 'modern synthesis' between evolutionary theory and genetics and then again with the rise of molecular biology in the 1950s" (Thompson 171). In 1953 when Francis Crick and James Watson successfully modeled the structure of DNA and helped complete our understanding of the mechanisms by which traits were passed down, mainstream science once again began to assume that it could provide a fully materialist account of life's end-directed organization. By precisely identifying the "cellular and molecular basis for the units of inheritance in the DNA on the chromosomes in the cell nucleus, the term *evolution* came to be used in a narrower sense than before to mean changes in gene frequencies in a population" (Thompson 172). Thanks to the Modern Synthesis, scientists no longer needed to be agnostic about natural selection's underlying mechanisms because now they could explain them in precise molecular detail. Inheritance and variation could be tied to DNA and how it codes for proteins—the building blocks of life.

Around this same time a new term was adopted in the life sciences that seemed to capture an organism's end-directed developmental processes and purposive behaviors: *teleonomy*. Whereas the *logos* in teleology is freighted with the rational and purposive properties of the observer's own mind, the *nomos* in teleonomy denotes lawlike or habitual convergence. Deacon says, "By coining a term which implied target-directedness but was agnostic about how this behavior was produced or came to exist, biologists and engineers could continue to use teleologically loaded descriptions, but without the metaphysical baggage that tended to come with them" (Deacon 116). It was in this spirit of teleonomy that neo-Darwinism appealed to genes and the proteins they coded for. But as Thompson points out, the problem with this view is that "genetic processes are described in the language of 'information,' 'instructions,' and 'coding,'…" (Thompson 173), a move which hides the end-directed quality at a deeper level of description, and which does not actually explain how such a quality could have arisen in the first place.

Information Theory: Semantic (Content) Information vs. Shannon Information

In chapter one, we followed the rise of cognitivism and its recourse to computation. Claude Shannon's information theory was central to this program because it allowed theorists to described cognitive processes in terms of "information processing."

Henceforward will see how such “information talk” insinuated itself into the discourse of biology as well, in effect recreating the cognitivist position in our collective understanding of organisms and evolutionary processes. This also means that we have reached the stage in our historical overview in which the problematic of teleology in biology dovetails with that of intentionality/mind that we encountered in our overview of cognitive science. Both disciplinary explanations converge on the mechanistic notion of “information” cited by Thompson at the end of the previous section.

In 1948, five years before Watson and Crick’s success, Claude Shannon published “A Mathematical Theory of Communication,” which laid the foundation for modern information theory. Shannon information theory is an essential point on our pendulum’s mechanism/holism swing because it continues to be a hegemonic heuristic in contemporary scientific practice for understanding orderliness and pattern in nature, and as Deacon tells us, information has become a “central unifying concept in the sciences” (Deacon 372). Given the pervasive influence this term wields throughout scientific practice in general, and the ways in which it structures our understanding of order/pattern in cognitive science and biology in particular, I want to spend a few pages unpacking its foundations.

As part of his work for Bell Telephone Laboratories, Shannon needed “to define and measure amounts of information transmitted through telegraph and telephone lines in order to estimate efficiencies and establish a basis for charging for messages” (Capra and Luisi 93). In order to quantify this process, that is, in order to develop a mathematical theory that could accurately describe the “information,” that was being transmitted along these telephone and telegraph lines, Shannon realized that "communication signals must be treated independently of the meaning of the message" (93). He accomplished this separation of content from signal by applying statistical thermodynamics—specifically the concept of entropy—to the carrying capacity of communication mediums.[26] In this context, "information" has nothing to do with the meaning of the message, instead it describes a calculable measurement of order—i.e., repetition or nonrandomness—in the signal (93). This orderliness in the signal medium is simply the physical instantiation of the underlying message. As Capra and Luisi say, "...the main concern of information theory is the problem of how to get a message, coded as a signal, through a noisy channel" (93). In information theory, the signal and physical characteristics are what matter, not the content of the message.

Enactivist theorist Daniel Hutto helps us place Shannon's theory in a more matter-of-fact register by distinguishing between a weak and strong sense of information. He says, "the weak notion is the familiar one that derives from the work of Shannon...It assumes

---

[26] Deacon says, "Because this measure of signal options is mathematically analogous to the measure of physical options in thermodynamic entropy, Shannon also called this measure the 'entropy' of the signal source" (Deacon 378). Entropy refers to a reliable tendency towards disorder in any system over time, one whose values can be calculated using the Boltzmann distribution. Deacon says, "This reliably asymmetric habit of nature [entropy] provides the ultimate background with respect to which an attribute of one thing can exemplify an attribute of something else" (378). In other words, because "non-correlation and disorder are so highly likely..." not to mention, statistically quantifiable, "...any degree of orderliness of things typically means that some external intervention has perturbed them away from this most probable state" (378). According to Shannon's analysis, *information* refers to "the improbability of receiving a given transmitted signal, determined with respect to the probabilities of all possible signals that could have been sent" (378). To sum up these insights, we can say that given (1) expected and calculable disorder of the signal medium (its entropy); (2) the improbability of redundancy in the received signal, it becomes possible to precisely quantify the likelihood of those redundant (i.e., informational) features. To make an analogy to visual perception, our ability to calculate the physical entropy of the communication medium serves as the "ground" upon which the informational "figures" (redundant patterns) can stand out.

that informational relations are nothing more than covariance relations; they exist wherever correlations between facts, events or properties obtain" (Hutto 55). In their book *Radicalizing Enactivism*, Hutto and Myin provide us with a lucid biological example: how tree rings can be said to carry information about the age of a tree, as each ring is the product of a new season's growth. Notice that in this extension of "information" into this biological context we are no longer dealing with communication signals, but with the more general insight (derived from Shannon's work) that "information" describes a measurable degree of orderliness and repetition in some physical substrate.

Information theory is functionalist. It is a method for abstracting and computing the orderliness of potentially *any* physical configuration because it is not wedded to a particular substrate and thus can be mobilized in myriad disciplinary contexts. We can see how Shannon's thermodynamic analysis of communication mediums, together with Turing's work on computation, helped establish a powerful materialist backbone for the burgeoning cognitive revolution. So long as one thing can vary lawfully or reliably in relation to another, we have grounds for treating it in terms of physically instantiated "information."

On the one hand, this functionalism makes such information-talk a powerful heuristic for connecting lawful or habitual covarying system processes across quite diverse disciplines and methodologies. On the other hand, it also means that "information" no longer refers to any of the things we normally would associate with a mental or biological process—i.e., purpose, intention, reference, or aboutness. This is what Hutto refers to as the "strong" sense of information, "the kind of contentful information—the message—that some communications convey" (55). The objectivity of the scientific method means that the weak sense of information pervades, but this weak sense simply amounts to a computation of order from disorder.

Stated simply, information theory presents big rhetorical—and philosophical—problems. The very term seems to imply a mentalistic property because "information" in our everyday, received sense of the word (i.e., without the weak/strong distinction) denotes a semantic content. In point of fact, all that information theory can do is describe/predict/explain the appearance of redundant patterns in a physical substrate, and by extension, how such redundancy can contribute to what our minds perceive as orderly arrangements. *In order to bear semantic (or contentful) meaning, such patterns still need to be interpreted, i.e., placed in some encompassing context, and ascribed an intention.* Given the rhetorical and philosophical ambiguity which this term inaugurates, Deacon argues that this redefinition has "…cemented the Cartesian cut into the formal foundations of physics…if it is assumed that we can simply replace the term *information* with *order*, thought becomes synonymous with computation, and any physical difference can be treated as though it has mentalistic properties" (374). By retaining the term "information" despite the lack of semantic content, Shannon's theory—and its extension across the science—functions as yet another rhetorical and conceptual sleight-of-hand. It enabled scientists and theorists from nearly every contemporary discipline (but particularly those working in biology and neuroscience) to explain life and mind in quasi-teleological terminology while at the same time divesting their objects of their basic teleological character.

"Information" Meets Evolution:

The methodological sleight-of-hand identified by Deacon in the above quotation is precisely where Thompson and Deacon's work converge. Both theorists attempt to course correct for the reductionist and materialist inheritances that resulted from Shannon's work; both theorists follow this thread down into the most fundamental neuroscientific and biological contexts. [27]

Thompson argues that the same basic computationalist logic is at work in neo-Darwinist approaches, in which "the genome is a set of coded instructions" and thus "essentially abstract and causally privileged in its role as a program in the cell. The cell therefore amounts to little more than a 'vehicle' driven by its genes" (Thompson 174). This is why Thompson likens neo-Darwinism's recourse to genetic "information" to the

---

[27] To be sure, we can also see this kind of "information" talk in the way that autopoiesis and cognition gets taken up by other systems theorists. For example, cyberneticist Bruce Clarke describes autopoiesis in explicitly informatic terms: "to maintain their autopoiesis, (self-referential) systems must remain operationally (or organizationally) closed to information from the environment. On that basis, they can construct their interactions with their environment *as* information" (Clarke, "Introduction" 9). In this passage "information" results from a cognitive system's operational closure, meaning that the system maintains an ongoing and invariant pattern of organization. In the case of a single cell, this pattern is its core suite of metabolic processes that sustain its physical boundary and that distinguishes the cell as a unity distinct from its environment. We know that in order for a cell to hold together as a unity, these processes must remain stable (or functionally invariant), and this is precisely what Clarke means by "closed to information from the environment." According to Clarke, this operational closure is the condition by which a cell can engage with the environment "*as* information" by which he means those sense-making processes that select salient features (information) from the environment because they are essential to the cell's autopoiesis. Clarke's passage is cumbersome for a few reasons. First, the redundant use of "information" is both ambiguous and nontechnical. Second, Clarke speaks about this process as though there were a clear sequence of cause and effect—i.e., *first* operational closure and *then* sense-making selectivity—but we know from our previous chapter that in self-organizing systems cause and effect have become totally interdependent. This is the essence of circular causality. Clarke demonstrates how the concept of "information" can insinuate itself into discussions around autopoiesis and cognition. Even though Maturana and Varela do not use the concept of "information" in their explication of autopoiesis, opting instead for descriptions such as internal/external "perturbations" (91) and related concepts such as "transformations," and "homeostasis," (78-80)—on its surface, *information* functions in many equivalent ways because it too relies on predictable covariation of states inside and outside of the autopoietic system.

cognitivist position in neuroscience, in which the "mind is essentially a matter of a computer brain inside the head" (173) because both positions "perpetuate the dualisms of hardware versus software, matter versus information, body versus mind" (174). As with cognitivism, recourse to genetic "information" tends to ignore the reciprocal interactions between an organism and its environment and upon which all evolutionary and genetic processes ultimately depend.

In other words, this focus on ever-more atomized system states (such as genes on the DNA strand) as opposed to totalized dynamical interactions gives rise to logical inconsistencies with respect to natural selection. In order for natural selection to occur, there needs to be a self-generated locus of causal processes—autonomous systems—that will compete with one another in order for the dynamics of evolution to play out. It makes no sense to ascribe this competitive behavior to genes or proteins. Genes and proteins do not act for their own self-interest and survival, *organisms* do. As Deacon tells us, "Natural selection could not have produced the conditions that made natural selection possible" (136). This is precisely where enactivism, and its recourse to the behavioral dimensions implicit in "autonomy," intervenes. In the context of evolutionary biology, autonomy is meant to describe and explain the self-seeking of organisms upon which the processes of natural selection depend.

According to the enactivist framework, competition results from the needs of whole organisms embedded in their environments. This is the position of developmental systems theory (DST), which emerged in response to the reductionist tendencies inherent in the modern synthesis, and what I have thus far been referring to as neo-Darwinism.[28] In effect, DST is autopoietic enactivism's answer to the reductionist tendencies of the Modern Synthesis and its influence on evolutionary theory. DST is a theory of evolution made possible by second-order cybernetics and autopoiesis, and it speaks directly to the exigencies of 4E theory and its investment in the whole organism.

According to the broad program outlined by DST, "it is unacceptable to say that DNA contains the information for phenotypic design, because this statement attributes an intrinsic semantic-informational status to one particular type of component and thereby divests this component of its necessary embedding in the dynamics of the autopoietic network" (Thompson 57). Here the notion of "semantic information" refers to context-independent representational potential, the idea that when DNA "codes" for proteins that it maintains some "special type of symbolic causal relation or a special type of intrinsically informational molecule that rises above the dynamic fray" (56). According to DST, genes very clearly cannot *signify* in this univocal sense, uncoupled from what Thompson calls the dynamic fray. As Nicholas Shea tell us, "Relying on a notion of genetic representation will be anathema to many DST theorists (Moss 2001), especially Susan Oyama, who identifies information talk as the source of gene centrism and genetic determinism (Oyama 1985)" (Shea 61).

[28] As Roberts et al. tell us, "For many, the locus classicus of developmental systems theory is Susan Oyama's 1985 book, *The Ontogeny of Information*, wherein Oyama rejected dichotomous views of development relying on the division of ontogenetic causes into genetic causes and generic (everything else, but usually mainly environmental) causes. For Oyama, as for other adherents to DST, developmental information resides neither in the genes nor in the environment, but rather emerges from the interactions of disparate, dispersed developmental resources—hence, the ontogeny of information" (954).

Thompson places his autopoietic enactivism firmly in this DST lineage alongside Moss and Oyama. Thompson says, "If genocentrism is homologous to the computationalist view that the mind is a computer in the head, then developmental systems theory is homologous to the enactive view that the mind is embodied in the active organism and embedded in the world" (187). He argues that the only way this genetic coding of information makes sense is if the content-bearing parts—the arrangement of nucleotide molecules into base pairs and genes—is integrated into the ongoing dynamic processes of organisms *and* their environments; in other words, in an autopoietic system. Thompson says, "Information is not intrinsic to the static linear array of the DNA sequence, but is rather dynamically constituted in and by the cell as an autopoietically organized, three-dimensional entity—by the cell as a body. In summary, the linguistic mode is emergent from the dynamical mode, and information exists only as dynamically embodied" (Thompson 57).[29] This is the same logic that we will trace all the way up the cognitive ladder. In chapter five we will examine linguistic systems, and I will argue for an embodied and context-dependent version of semiotics that avoids abstract systems in favor of immediate and pragmatic uses.

In this dynamical recontextualizing advanced by DST and enactivism, genes, proteins, and the well-defined biological functions they give rise to are subordinated to an organism's immediate environmental contexts and constraints, and to the sense-making processes that ensure its survival. To be clear, this subordination does not deny

[29] Here "linguistic mode" simply refers to the suite of univocal, gene-centric descriptions of information, and which helps Thompson distinguish the "dynamical mode" he is advancing in its place.

the role of evolutionary history nor the effects that functional organization exerts on an organism's ontogeny, but it does make good on the earlier assertion by Deacon that "natural selection could not have produced the conditions that make natural selection possible" (Deacon 136). Reproduction, and all the processes of genetic inheritance tied to reproduction, cannot occur without autonomous self-seeking, which is precisely what the theory of autopoiesis purports to explain. Thompson says, "Although autopoiesis and reproduction go hand in hand in living cells, there is a logical asymmetry between the two. Reproduction presupposes autopoiesis, but autopoiesis does not necessarily entail reproduction, for a system can be self-producing according to the autopoietic criteria without being capable of reproduction" (167). This is the decisive qualification that Thompson and Maturana and Varela place on evolutionary theory, and while no self-respecting evolutionary biologist would deny the causal role that individual organisms play in the emergent processes of evolution, the point I want to emphasize here is that it becomes easy to lose sight of this fundamental causal agency. In much the same way that cognitivism has tended to assume or bury in a deeper layer of description the precise causal role played by an interpretive intentionality in its descriptions of mental and perceptual "information processing," so too can evolutionary theory lose sight of the specific causal role played by the autonomy of individual organisms. In the following section, we will examine one case study to this effect. My larger contention is that autopoietic enactivism offers biology a cogent theory for placing this individual autonomy front and center in the complex web of life's many dynamic and emergent processes.

A Neo-Darwinist Rejoinder:

My goal in this section is to continue tracing the complex interrelationship between materialist and holistic approaches to life and mind. This section focuses on an acerbic response to Thompson's *Mind in Life* written by philosopher Daniel Dennett entitled "Shall we Tango? No, but Thanks for Asking." Dennett's academic career places him at the nexus of the themes and disciplines with which this dissertation is concerned, and he has directly written about the two primary sources underlying this research—favorably in the case of Terrence Deacon; critically in the case of Evan Thompson. Dennett articulates a materialist counterpoint from which to consider Thompson's enactivism, and he offers a clear defense of neo-Darwinism against some of Thompson's more ungenerous characterizations.

Elsewhere Dennett describes the ongoing "tug of war" in modern thought inaugurated by Descartes' paradigmatic statements opposing mechanistic bodies and immaterial minds (Dennett, "Aching Voids…" 321). Dennett says, "There are no entirely apt labels for the opposing sides of this gulf…" yet he does go on to offer up an illustrative few: reductionism/holism; and more suggestively Enlightenment/Romanticism, the former representing the reductionist and materialist camp, and the latter representing the holistic camp (Dennett 321). Dennett then proceeds to sort various figures onto either side of this divide. On the Enlightenment side he places people such as Darwin, Turing, Crick, and himself, while on the Romantic side, we find Thompson, Maturana, Varela, and Deacon (321-322). In sum, Dennett is a self-described Enlightenment-materialist philosopher of mind and biology. With this context in mind, we can turn to his response to Thompson.

The pith of Dennett's rejoinder to Thompson revolves around a rhetorical move that he calls "ratherings" where a theorist will advance some version of the following syntax: "'it is not the case that blahblahblah, as orthodoxy would have you believe; it is rather that suchandsuchandsuch—which is radically different'" (Dennett, "Shall We Tango..." 25). For Dennett, such a construction "implies—without argument—that there is an important incompatibility between the claims flanking it" (25). In effect, ratherings create a straw man of an established position in order to underscore the novelty or radicality of one's own claim. Dennett asserts that this technique pervades much of Thompson's book, but particularly in those sections that deal with information and genocentrism, (under which Thompson files Dennett's own work), and the counterpoint that Thompson endorses in its place—developmental systems theory.

Dennett points out all the ways in which Thompson misreads him. For example, when Thompson speaks about the "dualism" of hardware/software, Dennett argues that "It is about as secure and useful as any 'dualism' in science (e.g. matter and anti-matter, or — closer to home — living and non-living)" (27). Here Dennett draws a distinction between conceptual binaries that function heuristically, and Cartesian substance dualism, arguing that just because one draws such binary distinctions between the material and the organizational aspects of a system does not necessarily mean that one subscribes to dualist metaphysics. He continues, "Some sciencephobes have used the epithet 'dualism' somewhat ironically (one supposes) to attack any science that uses the concept of information, as if the criticisms of Descartes' brand of substance dualism could be somehow harnessed to their campaign against the use of computer science concepts in the humanities!" (27-28). Even though he seems to be constructing a bit of a strawman himself here—point Dennett. We ought to be wary of conflating the metaphysics of Cartesian substance dualism with any invocation of Shannon information.

With respect to Thompson's attack on neo-Darwinism, Dennett takes exception to his characterization of the gene as a unit of "pure" (i.e., context-independent) information. He says, "I wonder if anybody has ever subscribed to that myth" (28)? To illustrate this point, he draws a parallel to natural language:

> I take it everybody recognizes that the sentence 'Snow is white' only carries information about the colour of frozen precipitation on the assumption of the whole world of English speakers and readers. It carries no information at all intrinsically. Similarly, genes carry information only within the larger system of gene-readers, gene-recipe executors, etc., etc. I doubt that any 'genocentrist' has ever thought otherwise. (28)

While we have seen how this kind of oversight insinuates itself often enough in the context of cognitive science, it is not at all clear that this is an express feature of neo-Darwinist discourse. Dennett says, "In his campaign against orthodoxy, Thompson cites a veritable Hall of Fame of the would-be revolutionaries of biology, from such warhorses as Levins and Lewontin, Margulis and Sagan, and Lovelock, to Kauffman, Keller, Oyama and Moss" (29). He argues that individually, none of these figures has "made much of a dent on orthodoxy so far, and pooling them into a chorus…" is ultimately unpersuasive (29).

Again point Dennett. At least in the context of biology, Thompson does, at times, seem to overplay his hand by overestimating the novelty and radicality of his claims. I agree with Dennett that if we look past the "ratherings" what we are really left with is a sustained injunction to not lose sight of the dynamical interactions between the overall systemic unity and its environment. To this end, he says, "So far as I can see, autopoiesis is an excellent summary of what it takes for a collection of molecules to be alive, but it doesn't predict anything in biology that hadn't already been well understood by earlier theorists, or dissolve any puzzles that had been bedevilling those theorists" (25). From the "Enlightenment" perspective of biological science this statement is true. Autopoiesis is largely *not* a predictive theory designed to function in any kind of empirically rigorous manner (i.e., structure a specific hypothesis or identify any outstanding empirical "gaps" that can then be answered with experimentation). Autopoiesis was advanced foremost as a way to reimagine the foundation of living organization and the observer's relationship to it, and as such largely functions as an epistemological theory—i.e., what we can know and how we can know it—over and above any specific empirical or scientific claims. Therefore, I disagree with Dennett's characterization of autopoiesis in this passage.

The most interesting moments in this article come when Dennett examines Thompson's dichotomy of autonomy/heteronomy. We know that for Thompson, "semantic" (content-bearing) information such as that maintained in genes, only makes sense when understood in the context of autonomous systems, while heteronomy can only account for information as computational covariance. But Dennett calls this firm line in the sand between heteronomy and autonomy a "red herring." He says, "If an AI is designed…to muck about in the world and devise their own categories, the charge is vacated. As I and others have argued, all meaning in organisms is constructed by self-regarding processes that gerrymander the 'given' categories of physics to suit their purposes" (30). This example of an AI that can "muck about in the world"—by which I assume Dennett means, an AI that can train itself/optimize its own learning procedures. On the one hand these AI are heteronomous because the input/output parameters are programed from without by the computer scientist. The system is provided with a fixed initial data set and a well-defined target end-state. On the other hand, they are autonomous because in between these two explicit layers lies a hidden network of weighted connections that determines their own computational architecture, one that progressively changes itself as it is trained to compute a more ideal output. It is precisely this autonomous behavior that so easily leads non-specialists to impute mentalistic qualities to large language models such as Chat GPT. But for all this nuanced autonomous behavior, the AI is still subject to a heteronomous regime. It has got an input and an output layer.

This criticism also makes perfect sense. Autonomous cognitive systems are subject to the "heteronomous" regime of physics and chemistry, whose fundamental laws serve as the explicit input/output layers that constrain all autonomous, self-organizing processes. In other words, the laws of physics and chemistry function as the "computer scientist" in this analogy, and from these inputs and their resulting constraints, evolutionary dynamics have created biological intelligence that, over millions of years of "mucking," has self-organized such that it can define and delimit its own phenomenology. Once again, point

Dennett. When utilized as an explanatory heuristic, as means of distinguishing one general type of process from another, the dichotomy of autonomy/heteronomy is subject to context and definition and thus cannot be a firm line in the sand. But according to autopoietic enactivism, autonomy is not merely a heuristic, it is an irreducible, emergent *property* of the autopoietic system, and in this context, it does not have merely explanatory, but causal significance. This is an area where Thompson's analysis wins out. I do not believe that Dennett's rejoinder does much to undermine Thompson's autopoietic enactivist position or the pride of place it affords to autonomy.

Dennett explicitly endorses Maturana's assertion that "*Living systems are cognitive systems*…[and that] This statement is valid for all organisms with and without a nervous system" (Maturana 13), but he also says that "the 'autonomy perspective' is not required for making sense of sense-making" (Dennett, "Shall We Tango…" 30). To substantiate this claim he invokes the "agency" of macromolecules. He says:

> I say that agency is born 'in the first macromolecules that have enough complexity to "do things"'. Thompson points out, truly enough, that such macro-molecules cannot exist without being parts of autopoietic systems, but still they are agents, even if not fully autonomous. Think of motor proteins—little porters trudging along on their actin or tubulin highways carrying freight to where it is needed. Think of proof-reading enzymes. (30)

The idea advanced here is that just as a cell and its sense-making can be observed to "do things," so too can macromolecules with their well-defined functions be reasonably said "to do things." But the agency Dennett wants to ascribe to biological macromolecules (carbohydrates, lipids, proteins, and nucleic acids); the self-organizing cellular structures made from them (microtubules, lipid bilayers, protein and carbohydrate matrices etc.); and all the functional processes that these macromolecules contribute to a cell's overall sense-making (for example, the sensing of helpful/harmful chemicals by enzymes in the cell membrane), all depend on their incorporation into the organization of the cell as a whole. Dennett acknowledges this. They cannot exist outside this larger set of circularly causal processes in which part and whole are totally interdependent.

This is the one moment in Dennett's rejoinder where I think he falls victim to the kind of reductionism for which Thompson initially calls him to account, embedding the property one wants to explain—e.g., the sense-making that defines "cognition"—in a still deeper layer of explanation. In this case, the deeper layer takes the form of "agency," a move that conflates biological *function* with the full phenomenological implications of "autonomy." As long as the system holds together as an operationally closed unity, *autonomy* entails the subordination of all other processes to this larger systemic unity ("Autopoiesis…" 80). This subordination of underlying *functions* to the unity-cum-autonomy of autopoiesis is arguably one of the clearest expressions of "self" in the formal self-reference of cognitive systems. The sense-making that defines cognition must be situated at the level of the autopoietic whole, and this means that if a given biological function is part of a larger system, the autonomy of that system necessarily supersedes any lower-order expressions of "agency. In this passage, Dennett places agency—the simple capacity *to do things*—on a par with autonomy; in essence, conflating function with cognition.

Overall, I think it is telling that the pith of Dennett's rejoinder focuses largely on a question of rhetoric. Dennett acknowledges the validity of, and many of the upshots to, the enactivist position, but he also demonstrates for us that the "autopoietic" side of Thompson's enactivism shares far more of its position with neo-Darwinist accounts than otherwise, and that any attempt to differentiate it on radical or revolutionary grounds is misplaced. As I suggested in chapter one, we are better served by thinking of enactivism as a kind of sustained qualification rather than an outright repudiation of a larger orthodoxy. Enactivism has demonstrated its ability to disrupt orthodoxy in the context of cognitive neuroscience (e.g., cognitivism), but this success should not be used as a knee-jerk justification for radicality in other disciplinary contexts. Perhaps Thompson's enactivism is better off courting neo-Darwinist approaches like Dennett's while simultaneously doubling down on the epistemological and phenomenological implications that follow from the emergence into autonomy.

Finally, I think this case study of a self-described materialist/neo-Darwinist such as Dennett supports the organizing principle of this dissertation: that autopoietic enactivism has real import for the interdisciplinary and translational aspirations of the environmental humanities because of its emphasis on the commonalities that exist between vastly different scales of observation and between organisms which, on their surface, evince wildly different cognitive processes. I argue that Dennett and Thompson articulate positions that are not so radically opposed as their academic posturing would, at first glance, suggest. In some ways I think their debate serves as a cautionary tale: while it is useful to engage in the kind of broad intellectual and disciplinary historicizing that structures this chapter, we also need to be wary of (re)producing strawmen that can undermine or distract us from the interdisciplinary and translational goals at the heart of the environmental humanities.

Regardless of whether Dennett is guilty of “genocentrism,” the takeaway for our purposes in this chapter is that any kind of implicit or unexamined recourse to Shannon “information” tends to reify, or simply define out of existence, the end-directed and purposive qualities of both life and mind, and by extension the embodied and context-dependent ways in which this purposiveness is enacted. Ends and purposes cannot be reduced to material (efficient) causes and thus stand at odds with the materialist and reductionist tendencies of the scientific method. *Purposiveness* is life and mind’s quintessential feature.

Autopoiesis and its Discontents:

Maturana and Varela’s theory of autopoiesis took the hugely important first step equating living organization with mental processes—this is its revolutionary synthesis and the reason it continues to persist and reverberate across cognitive science, systems theory, and various ecological and 4E approaches. Despite this profound advance, it also explains life and mind’s quintessential purposiveness by effectively *reducing* it to the autonomous character of the autopoietic system—this represents the pith of its “discontents,” and the focus of this section. Autopoiesis purports to explain the appearance of purposiveness in cognitive and biological systems by appealing to their autonomous organization, and thus, “in its original formulation was explicitly mechanistic and antiteleological” (Thompson 141).

In “Autopoiesis: The Organization of the Living” Maturana and Varela are expressly committed to a naturalized and scientifically valid explanation of “the nature of living organization” (75). To accomplish this, they “make a starting point of the unitary character of a living system, and maintain that the evolutionary thought through its emphasis on diversity, reproduction and the species…has obscured the necessity of looking at the autonomous nature of living unities” (75). And their approach will be “mechanistic” because it will not rely on any sort of vitalistic forces or principles “which are not found in the physical universe” (75).

On the surface then, there seem to be two competing exigencies—maintaining holism and maintaining mechanism. They try to balance these exigencies by clearly distinguishing between their heteronomy as observers, and the autopoietic unity’s autonomy. From the outset, the notions of unity (wholeness) and autonomy are conjoined. Autonomy is a *property* that gets instantiated at the level of the unified whole. It is a property that depends on the autopoietic unity’s organizing relations, and which gets expressed in its phenomenology. In Maturana and Varela’s rendering, even though autonomy is the most important and elusive property of living organization (73), it does not function as an unexamined purpose or constitutive norm. It is a property capable of mechanistic description.

To underscore these mechanistic pretensions their first chapter develops the core self-referential dynamics of all autopoietic and allopoietic systems in terms of “machines.” This machine metaphor helps eliminate the whiff of vitalism, but to be clear, for

Maturana and Varela a machine is tantamount to a system (77).[30] They describe the basic pattern of organization that constitutes an autopoietic machine as “an homeostatic (or rather relations-static) system which has its own organization…as the fundamental variable which it maintains constant” (79). In this context, “autonomy” signifies that autopoietic machines “subordinate all changes to the maintenance of their own organization” (80).

Their second chapter begins by rejecting all teleological and teleonomic notions—including the notion of function—from the concept of “organization” because they view these as nothing more than descriptive glosses that originate in the observer and which “reflect our considering the machine or system in some encompassing context” (85). As we have seen, in biology this encompassing context takes the form of evolution by natural selection, which outlines a larger set of organizing relations (inheritance, variation, competition and scarcity) that can legitimate descriptions of biological functions by subordinating them to this larger dynamical process.

Functional, teleonomic, and teleological descriptions all define living organization within some kind of encompassing context in which a specific feature of the system can be exhaustively (i.e., mechanistically) defined by its relation to some well-defined function or output. In contrast autopoietic machines, because of their self-organizing dynamics, “do not have inputs or outputs” (81). Talk of biological functions can, for example, help explain how morphological or behavioral features contribute to an

[30] This conceptual and terminological conflation of machine-life-system can also be explained, in part, by the theory’s cybernetic antecedents.

organism's ontogeny, but the contention here is that such descriptions, because they delimit some specific subset of input/output relations (for example, how the protein hemoglobin has the biological function of transporting oxygen molecules) that they do not necessarily belong to the autopoiesis of the organism, nor its phenomenology.[31] The mere fact that "we can divide physical autopoietic machines into parts does not reveal the nature of the domain of interactions that they define as concrete entities operating in the physical universe" (82). This lack of input/output parameters is the essence of what became known as *operational closure*, but which Maturana and Varela first define in terms of "the autopoietic space." They say:

> We as observers can project all cellular processes upon a system of three orthogonal coordinates, and legitimately say, as valid in the projection, that specification is mostly produced by nucleic acids, constitution by proteins, and order (regulation) by metabolites. The autopoietic space, however, is curved and closed in the sense that it is entirely specified by itself, and such a projection represents our cognitive relation with it, but does not reproduce it…In such a system any deformation at any place is not compensated by bringing the system back to an identical state of its components as it would be described by projecting it upon a three-dimensional Cartesian space; rather it is compensated by keeping its organization constant…(92)

[31] They say: "We can analyze a physical autopoietic machine in its physical parts and treat all its partial homeostatic and regulatory mechanisms as allopoietic machines (sub-machines) by defining their input and output surfaces. Accordingly, these sub-machines are not necessarily components of an autopoietic machine because the relations that define such a machine need not be those that they generate through the input-output relations that define them" (82).

In this passage we can see how autonomy inaugurates a situated perspective that cannot be reduced to third-personal description. The unity is operationally closed because it only refers to itself, i.e., its own homeostatic relations. It does not refer to outside states or conditions *at all*, it merely *responds* to outside conditions (perturbations) by maintaining its autopoiesis. Thus, all that a third-personal description can accomplish is to specify the organizing relations between components or very general rules for how these components can or will interact. It cannot reproduce this situated perspective. This passage also starts to tease out the so-called "virtual" dimensions at work in self-reference: the autopoietic unity is not material, it is relational and processual. Yet these immaterial features which constitute the virtual whole determine and are determined by the physical interactions of all its underlying parts. Explaining this reciprocal causality in precise thermodynamic detail will motivate much of Terrence Deacon's work, but for now, the takeaway with respect to teleology is that none of these functional or purposive descriptions can be brought to bear on the concept of "organization" per se because:

> [organization] only states relations between components and rules for their interactions and transformations, in a manner that specifies the conditions of emergence of the different states of the machine which, then, arise as a necessary outcome whenever such conditions occur. Thus, the notions of purpose and function have no explanatory value in the phenomenological domain which they pretend to illuminate, because they do not refer to processes indeed operating in the generation of any of its phenomena. (86)

Here "phenomenology" denotes a domain of interaction that is "defined by the properties of the unity or unities that constitute it…" (116). Because the autopoietic machine preserves a set of dynamic processes (constitutive relations) which are both topologically and operationally closed, its "phenomenology" signifies a situated perspective which will resist a totalized description from without on the basis of a purpose or function. References to purposes/functions are reflections of the observer's heteronomy; not the autopoietic machine's autonomy. Accordingly, "Living systems, as physical autopoietic machines, are purposeless systems" (86). This rejection of teleology is precisely where Terrence Deacon's theory of emergent dynamics and Thompson's autopoietic enactivism, will intervene, and as I will argue in the following chapter, *surpass* the theory of autopoiesis. Deacon's primary contribution to the discourse of systems theory is a mind-independent explanation of organizational pattern. In this way he tries to circumvent the problems posed by the observer and resuscitate teleological descriptions on the basis of sound materialist science. Thompson, for his part, will address this teleological discontent through his use of phenomenology.

In this explicitly anti-teleological context Thompson asks, "What bearing does this argument [the notion that a purpose/function possesses only communicative, not explanatory value] have on the Kantian thought that to exist by being self-organizing is to be intrinsically purposive?" (Thompson 145). This question is important because there is a tendency to conflate autonomy with the intrinsic or "natural" purposiveness that has traditionally differentiated biological life from machines. Thompson offers us three responses that together represent his attempt to reconcile Maturana and Varela's treatment of autonomy with Kant's notion of natural purposes.

First, he states that Maturana and Varela's rejection of "extrinsic functional descriptions of a system's components is logically compatible with the whole system being intrinsically purposive" (Thompson 145). This is because extrinsic functions presuppose a larger context while intrinsic purposiveness only "pertains to the overall organization of the whole system itself" (145). Second, "in laying down the conditions of the autopoietic organization, no reference is made to any ends, purposes, goals or functions of the component processes or the whole system. Thus the theory of autopoiesis does not presuppose or appeal to intrinsic purposiveness in an unanalyzed way but rather explicates this notion naturalistically" (146). Third, "intrinsic" according to this naturalized account "must be taken to mean constitutive and not nonrelational (and hence unanalyzable)" (145). By which he means that the autopoietic unity evinces intrinsic purposiveness that is explicable solely by the relations existing between parts. Thompson then goes on to say:

> Given this line of reasoning, it might be better to call this sort of constitutive purposiveness immanent purposiveness. The thought here is that purposiveness is neither a nonrelational property of something internal to the system (as 'intrinsic' can misleadingly suggest) nor a property determined by something outside the system (by something that transcends the system). Rather, purposiveness is a constitutive property the whole system possesses because of the way the system is organized. (Thompson 146).

Philosophically, I find this explanation wholly convincing. When we begin to examine life and its self-organizing processes we are confronted with one very undeniable and immanent purpose—its own perpetuation. However, from a materialist, (or as Dennett would say, "Enlightenment" point of view), I find this explanation *unsatisfying* because it somewhat too glibly glosses over the defining characteristic of a purpose or end—existing in relation to an *as yet unactualized state*. To this end, Deacon says:

> in their [Maturana and Varela's] effort to make the autonomous observer-self a fundamental element of the natural sciences, the origin of this self-creative dynamic is merely taken for granted, taken as a fundamental axiom. The theory thereby avoids the challenges posed by phenomena whose existence is determined with respect to something displaced, absent, or not yet actualized, because these are defined in internalized self-referential form. (Deacon 6)

In other words, by making a starting point of "the unitary character of a living system" ("Autopoiesis…" 75), Deacon argues that they have begged the question, and I do not see how Thompson's description of "immanent purposiveness" does much to move beyond the critique laid out by Deacon regarding the *as yet unrealized state* of existence which defines purposiveness.

Even though Deacon's critique is leveled at Maturana and Varela, it poses a problem for Thompson because like Maturana and Varela, he also maintains that the highly developed operations of nervous systems are simply more complex instantiations of the basic cognitive dynamic outlined by autopoiesis. For example: "…this circular organization and mode of coupling with the environment *are recapitulated in a more* *complex form by the nervous system*" (Thompson 260, emphasis added). Compare this to Maturana's statement in "The Biology of Cognition" where he says, "*The nervous system expands the cognitive domain of the living system by making possible interactions with 'pure relations'; it does not create cognition*" (13). By "pure relation" Maturana seems to mean something that exceeds the immediate system-environment coupling that constitutes cognition—in other words, the ways in which the nervous system can be said to represent or mediate between features of its environment that are absent or not yet actualized. How is it then that nervous systems can evince the teleology of "pure relations" if they are merely expanding the phenomenological bounds of basic autopoiesis?

This drastic scaling up of "cognition" from basic autopoiesis to nervous systems is a potential explanatory gap in enactivist theory—at least in the context of mainstream neuroscience which is wary of such systems theoretical abstraction given that higher-order features of vertebrate mental life are largely understood in terms of mental representations or information-processing of one kind or another. Therefore, if enactivism wants to make good on its ambition to unify, implicitly, the autonomous and cognitive core of biological life with those more complex features of subjectivity, intentionality, and mind, then it will have to explain how such "pure" representational, seemingly contentful informational features of mental life relate to its most basic descriptions of autopoietic cognition. This is the theoretical bridge that Deacon will construct for us.

Not only am I sympathetic to Deacon's critique, but as we will explore in the following section, I believe Thompson's subdivision of the body-body problem into separate "scientific" and "enactivist" domains reproduces some of these same indeterminacies. Autopoiesis, and by extension enactivism, have a vexed relationship with mainstream life-science research, which I believe stem from two primary causes.

First, it does not precisely define the emergent transitions that give rise to such autonomous organization in the first place.[32] Maturana and Varela state that the organization of a machine "only states the relations between components and rules for their interactions and transformations, i*n a manner that specifies the conditions of emergence* of the different states of the machine which, then, arise as a necessary outcome whenever such conditions occur" ("Autopoiesis…" 86, emphasis added). The implication is that if all the conditions of emergence are met—already a rather liberal hedge—then the emergence into autonomous organization will follow as a matter of course. This is really no explanation at all. It amounts to a description of what is already physically present in the system and the environment. It amounts to a generalized account of self-reference, but not *how* that homeostatic pattern came to be. Second, it does not reach down into the thermodynamic fundament. They say:

[32] Daniel Hutto draws our attention to this vexed relationship to mainstream cognitive and life-science research. He says, "For even among those who are prepared to accept that the enactive approach holds promise for understanding how organisms develop and interact over time, a standard verdict is that it lacks the independent explanatory resources to provide a genuinely alternative understanding of the basis of mentality (Ramsey, 2007; Clark, 2008)" (Hutto 47). Here we see the recognition of a *lack* in enactivist theory, even among its most ardent supporters. My contention is that Deacon's emergent dynamics will offer *precisely* the kind of "independent explanatory resource" that Thompson and company lack. Deacon's emergent dynamics recapitulates many of the core ideas in autopoiesis and autopoietic enactivism, but it does so in a unique, thermodynamically grounded analysis of emergence and dynamical systems.

> Although indeed energetic and thermodynamic considerations would necessarily enter in the analysis of how the components are physically constituted…these considerations do not enter in the characterization of the autopoietic organization. If the components can be materialized, the organization can be realized; the satisfaction of all thermodynamic and energetic relations is implicit. (89)

By excising the thermodynamic and energetic influence on a system's formal properties, Maturana and Varela come (rhetorically, if not conceptually) close to implying that the material realization of organizing relations (i.e., structure) can be meaningfully disentangled from pattern—in other words that form can be wrested from matter. This is another reason why their mechanistic pretensions are not as mechanistic as they could and should be.

Despite these shortcomings, it is clear that Thompson's autopoietic enactivism more directly confronts these outstanding problems, indeterminacies, and explanatory gaps, specifically via his adoption of phenomenological approaches that read purposiveness into the very foundations of biological life. These phenomenological approaches will occupy the bulk of chapter four. For now, I would like to focus on a key moment in Thompson's book in which he brings us to a philosophical and scientific crossroads regarding the phenomenon of emergence, and enactivism's place within this problematic.

The Body-Body Problem:

Throughout these first two chapters we have encountered, in one form or another, the problems posed by life and mind's non-material intentions, purposes, and autonomy in their relationship to the materialist and mind-independent descriptions of science. This ongoing dialectic between the im/material is known as the "explanatory gap." In the context of cognitive science this gap has largely been explained by means of computations, mental representations, or connectionist neural networks; while in the context of biology this problem is explained by recourse to genetics and Darwinian natural selection. As we saw at the beginning of this chapter, both positions are legitimated by the pervasive influence of Shannon information theory. In this final section, I want to focus on an important moment in Thompson's book in which he addresses this explanatory gap head on and offers perhaps his most unique and phenomenologically inspired answer to the problem. We will see that he is not so much interested in *closing* the explanatory gap, as in redefining its boundaries in an inherently more productive philosophical manner.

Daniel Hutto helps set the stage for us: "Thompson sees adoption of the enactive approach as a way to put aside the mind-body problem, once and for all, and to refocus our investigations on the more fertile—phenomenologically-inspired—body-body problem" (Hutto 45). The pith of this phenomenological reframing is that mental experience is not ontologically distinct from bodily experience—that subjectivity exists on a continuum with bodily agency (i.e., our intero, extero, and proprioceptive abilities) and represents a difference of degree, not kind. In the parlance of phenomenology, that there is *something* *it is like* to have a body, and that this phenomenal quality exists prior to the emergence of higher-order mental states.

I want to begin by reiterating the rather large "gap" that emerges in autopoietic enactivism's own analysis. We know that for enactivism the paradigm case of cognition can be found in the sensorimotor loop of non-nucleated bacterial cell.[33] But whereas the organization of a bacterium is strictly metabolic in nature and limited to this (relatively speaking) narrow suite of processes; cells in larger bodily systems contribute to incredibly complex wholes that themselves assume a unique systemic identity. A case in point, the nervous system. Best estimates for the number and makeup of cells in the nervous system are roughly 90 billion neurons and approximately 360 billion glia cells, which are distinguishable from the former by their lack of action potentials—the ability to fire electro-chemical impulses. How does enactivism actually account for and address this drastic scaling up of complexity from single cells to networks of billions of neurons?

Thompson's answer is that "…this circular organization and mode of coupling with the environment *are recapitulated in a more complex form by the nervous system*. The nervous system establishes and maintains a sensorimotor cycle, whereby what one senses depends directly on how one moves, and how one moves depends directly on what one senses" (Thompson 260, emphasis added). Despite this drastic difference in complexity and scale, the nervous system—like the single-cell—is still a system, and therefore shares a

---

[33] All those molecular processes which constitute and reproduce in an ongoing fashion the organizational "form or pattern that remains invariant through any kind of structural change…" and which "holds [the cell] together as a distinct entity" (Thompson 98).

basic set of processes such as *operational circularity* (outputs fed back into the system as inputs), *recursive self-constitution* (each part of the system assumes its identity/function due to its participation in the larger whole), and *boundary formation*. Obviously, this move requires a very high level of abstraction, which may be anathema to some neuroscientists, although there is nothing in it that contravenes the basic tenets of systems theory.

Thus, Thompson's approach is to work from the "bottom up" so to speak; to argue that what obtains in the smallest case necessarily obtains in the most complex. In this way, the body-body problem takes the seemingly insuperable issues posed by qualia and tries to insert these features back into the underlying dynamics of simple organisms.

Thompson's first point of intervention has to do with the conceptual dichotomy of mental/physical. These are the preferred terms of Thomas Nagel, and at several points in this section Thompson refers to it as "the Cartesian formulation of the hard problem" (237). Even though in *Meditations on First Philosophy* and *Discourse on Method*, "body" was already Descartes' preferred term, it is a body unlike anything we encounter in the discourse around complex biological systems. For Descartes, the "body" is a wholly inanimate object subject to physical laws and possessing no motive force of its own.[34] Nagel, like Descartes, reduces bodily phenomena to essentially inanimate objects of third-

[34] For example, In *Meditations* Descartes says, "…by 'body,' I understand all that is capable of being bounded by some shape, of being enclosed in space, and of filling up a space in such a way as to exclude any other body from it; of being perceived by touch, sight, hearing, taste, or smell; of being moved in several ways, not of course, by itself, but by whatever else impinges upon it. For it was my view that the power of self-motion, and likewise of sensing or of thinking, in no way belonged to the nature of the body" (64). Here "body" is resituated in abstract spatial and geometrical terms. We know bodies according to their shape, according to their mutual exclusion and ability to be circumscribed. We know them by our senses, and we perceive them as being subject to, and animated by, outside forces (efficient causation).

personal description, and Thompson's first move is to substitute this inanimate, physicalist Cartesian "body" with the notion of a lived/living body of phenomenology. According to Thompson, when we engage in this lexical revision, we are no longer confronted with the "contrived" issue of animate subjects versus inanimate objects, a move which already forecloses any possible synthesis, but with "…*the emergence of living subjectivity from living being*, where living being is understood as already possessed of an interiority that escapes the objectivist picture of nature. *It is the issue of emergence that we need to address, not the Cartesian version of the hard problem*" (236, emphasis added).

To this end, Thompson highlights two ways in phenomenological analysis in which "body can be disclosed…as a material thing (*Körper*) and as a living subject of experience or living body (*Leib*)" (235). This distinction is not meant to recapitulate the Cartesian split but highlight different ways in which a body can be disclosed (contextualized). In true phenomenological fashion it acknowledges the body's objective and subjective contexts without essentializing either one. This initial distinction then serves as the foundation for further one: that between the "structural morphology of the physical body and its living and lived dynamics" (236). Morphology refers to the body's structural features, while the "dynamics comprises the lived flow of life, that is the flow of intentional movement and lived sensations (interoceptive, exteroceptive, and proprioceptive)" (236). The observation here is that the structural features of an organism's body are not decomposable, they are always embedded in a dynamic process of living. In essence, Nagel's Cartesian formulation of the hard problem ignores the full implications of a dynamic systems approach, in which part and whole are totally interdependent.

Despite this conceptual jostling, we are still confronted with a gap: between the structural features of a living body (morphology) and how these features constitute a lived body—a locus of feeling, intention, autonomy. But the difference now is that it is a gap with a common denominator. On both sides of the gap we have a "body" understood as an organismic whole, with its "living" (dynamic processes) and "lived" (agential) aspects. By Thompson's reckoning, this gap "is no longer between two radically different ontologies (mental and physical), but between two types within one typology of embodiment" and furthermore "the gap is no longer absolute because in order to formulate it we need to make common reference to life or living being" (Thompson 237). In other words, *body*, properly understood, is already beyond the gap because its essential features are always disclosed within a larger dynamic and organismic whole. He says, "The lived body is the living body; it is a performance of our living body, something our body enacts in living (Thompson 237).

Thompson's purpose in this section is to reframe the contours of a much larger and longstanding debate, and I believe his attention to these minute lexical, conceptual, and rhetorical shifts succeeds in reframing the philosophical non-starter that is Nagel's hard problem. He also manages to touch on the crux of the debate earlier in the argument: *emergence*. That is, how precisely does the lived body—as the locus of subjective, causal agency—emerge from the living dynamics and structural morphology of the body? Thompson concludes this section on the body-body problem as follows:

> The scientific task is to understand how the organizational dynamic and dynamic processes of a living body can become constitutive of a subjective point of view, so

> that there is something it is like to be that body. For the enactive approach, this task takes the form of trying to understand a lived body as a special kind of autonomous system, one whose sense-making brings forth, enacts, or constitutes a phenomenal world. (Thompson 237)

Though he does not say the word emergence here, it is writ large in the gap he describes. The task for both science (i.e., strict materialist and empirical approaches) and enactivism (i.e., empiricism situated and qualified within a larger phenomenology of life and mind) is to come to terms with the emergent properties of life (intentionality, agency, autonomy, and qualia etc.) via its underlying dynamic systems processes. Thompson is very clear that autonomy is the province of enactivism, but he is rather cagey with respect to the scientific task. In a word he leaves this so-called "scientific task" strategically vague and doubles down on his program of autopoietic enactivism, asserting that (a) living beings are autonomous systems and (b) organizing his analysis so that all roads begin from this autonomy. This is another philosophical and rhetorically savvy way of bracketing out some of the scientific problems posed by emergence and is one reason why I cannot find Thompson's description of the enactivist task wholly satisfactory. This shortcoming is directly traceable to the theory of autopoiesis on which his enactivist account rests.

This chapter is subtitled "the beginning of ends"—my attempt to signal the ongoing problems posed by teleology in the history of mental and biological inquiry. Despite some of its shortcomings, we ought to understand the body-body problem as one small but powerful argument in a larger, systems-theoretical paradigm shift. We ended chapter one with an enactivist-inspired contribution to cognitive neuroscience—

neurophenomenology—that tries to bridge disciplinary and methodological divisions that arise from the explanatory gap. Thompson's "body-body problem" is an extension of the same unifying impulse that motivates neurophenomenology. For the same reason, I would also argue that it is an exemplar of environmental humanist discourse—a concerted attempt to move beyond unproductive philosophical language and the entrenched disciplinary divisions between science and the humanities that can follow from it.

Conclusion:

The purpose of this chapter has been to examine the recent history of mechanistic and holistic explanations of life. We have seen how evolutionary biology—like cognitive science—makes use of Shannon information theory in order to explain—or mechanistically explain away—the problems posed by life's teleological organization. We compared the neo-Darwinist model of life and evolutionary development to the holistic explanations advanced by enactivism and developmental systems theory. Most importantly, we came to terms with the autopoietic enactivist urtext: how Maturana and Varela's theory of autopoiesis attempted to resolve this holistic/mechanistic dialectic via recourse to the autopoietic unity's inherent autonomy. This is a move that rejected teleological organization and purposive/functional descriptions, and which I argued has led to ongoing discontent with autopoietic enactivism.

In the following chapter, we will take up the "scientific task" outlined by Thompson in his body-body problem, that is, understanding how the organizational dynamics of a living body can become constitutive of a *subjective* point of view. This task is nothing short of bridging the hard problem of consciousness: purposes and intentions are the defining features of subjectivity, and as Terrence Deacon will argue, they pose problems for third personal description because they are expressed as material absence. Deacon's theory of emergent dynamics will attempt to explain the immaterial, purposive, and autonomous character of life and mind in a scientifically valid, mind-independent way.

Chapter Three:

"Emergent Dynamics": A Case Study in Emergence and Cognition

> A simple trick from the backyard astronomer: if you are having trouble seeing something, look slightly away from it. The most light-sensitive parts of our eyes (those we need to see dim objects) are on the edges of the region we normally use for focusing.
>
> —Jonathan Safran Foer, *Eating Animals*

Since its inception and codification in the early twentieth century, systems theory has lived by the axiom "The whole is greater than the sum of its parts," a phrase often attributed to the first gestalt psychologist Christian von Ehrenfels, but which can be traced all the way back to Aristotle's *Metaphysics*. The phrase sums up our problematic in a nutshell: life and mind are wholes (unities) whose essential properties are not reducible to their underlying parts. Enumerating all the neurons in our brains will not explain subjective experience. Describing all the physical components of an organism will not explain its purposive behaviors. Nor can such mechanistic methods furnish us with an explanation of how we *ought* to comport ourselves as social creatures. Empirical facts cannot explain the values that define our existence.

Thus, the central complaint of every contemporary theorist and theory we have encountered so far has been that over the course of the twentieth century, as mechanistic science has excelled to ever-greater descriptive heights, it continues to encounter a causality that it cannot adequately explain. Life and mind are characterized by their intentional, purposive, and teleological organization—existing for-the-sake-of something that is non-intrinsic. Purposes, functions, and ends are causes that cannot be neatly reduced to underlying physics and chemistry. To be a teleologically organized whole means to exist in relation to immaterial norms that (seemingly) transcend the materiality of the world.

In the previous chapter, we saw how the theory of autopoiesis attempted to come to terms with this problem. Maturana and Varela argue that any description of living organization that makes use of "purposes" or "functions" presupposes a conception of order or pattern already at work in the observer's own mind and as such *cannot* count as a valid materialist (i.e., mind-independent) description. Their solution was to abstract only the most basic pattern of organization characteristic of all living things: (1) operational or organizational closure made possible by (2) a semi-permeable physical boundary that differentiates system and environment and that (3) selectively allows energetic inputs from without. The result of this system/environment coupling and its ongoing autopoietic (self-making) dynamic is a domain of interaction they deem *cognitive*. In this way, the theory of autopoiesis unifies life and mind from the very start and from the ground up. Accordingly, *all life is cognitive.* Perhaps most importantly to the logic of their approach, this living organization evinces *autonomy*. Autonomy is a property entailed by this autopoietic organization and is tantamount to the end-directed, purposive behaviors with which life and mind confront us. Maturana and Varela argue that autopoiesis provides us with a mechanistic explanation of autonomy, and as such, obviates teleology.

Despite the cogency of their position the profound unification of life and mind their theory accomplishes, Maturana and Varela make a starting point of the unitary character of the living ("Autopoiesis…" 75). In other words, they take the single cell as their starting point, but it is a starting point that brackets from consideration its own thermodynamic and energetic preconditions. Therefore, the goal of this chapter is to dig one final layer deeper.

This chapter will use Terrence Deacon's novel approach to systems theory which he calls *emergent dynamics*, and the minimal proof of principle he constructs using this method—*autogenesis*—as a means of obviating observer-dependent descriptions of organizational pattern, a move which he argues can (re)integrate teleological descriptions in a scientifically valid (materialist) procedure. Deacon's theory is not committed to the paradigm of living organization per se, but rather to the deeper causal problematic surrounding teleology and its underlying thermodynamics. Thus, as the epigraph to this chapter suggests, emergent dynamics is a means of looking slightly away from our primary object so as to bring some of its most essential features more firmly into view. In effect, Deacon's theory will explain how it is that this autopoietic/autonomous unity came to be in the first place, and hence the subtitle of his book: "How Mind *Emerged* from Matter."

Deacon, in a move redolent of Derridean deconstruction, begins by inverting the sedimented logic that has been operative in systems theory since its inception. He says, "In contrast, [to this aforementioned part/whole axiom] and not simply to be enigmatic, I will try to explain instead how the whole is *less* than the sum of its parts" (Deacon 43). To accomplish this inversion, his logic will revolve around the role of *constraint*—that is, "the state of being restricted or confined within prescribed bounds. Constraints are exhibited by what is not there but could have been" (Deacon 548). As the passage suggests, Deacon will attempt to enter into the problematic through the back door by trying to rigorously define the contours of systemic organization negatively, by what is *not* there but could have been.

To arrive at his theory of autogenesis Deacon will explicate three general system types: homeodynamic, morphodynamic, and teleodynamic. These are systems which more traditional systems theories would recognize as thermodynamic, self-organizing, and cognitive/autopoietic systems, respectively. In other words, Deacon is operating in precisely the same playing field as cybernetics, second-order cybernetics, and autopoiesis. But by redefining these systems via a more generalized and dialectical notion of causal change—what he will term orthograde/contragrade; by attending at every step of the analysis to the constitutive role of constraints in dynamical behaviors; and by using this dialectical pattern of change to reframe our understanding of thermodynamic work—these systems will no longer be identical to their thermodynamic, self-organizing, and autopoietic analogues.

Deacon's emergent dynamics gives us the tools to interconnect *all* forms of existence and *all* types of system, whether it is the simple thermodynamics that define our coveted morning cup of coffee, the complex geometrical patterns of snowflakes, the homeostatic and life-protecting organization of the biosphere (Gaia), or the most complex expressions of human consciousness. Deacon articulates an emergentist logic that bridges the hard and human sciences, and as such offers the environmental humanities what I would argue is its most robust and thought-provoking theory to date.

In short, Deacon provides us with unambiguous language and novel concepts for understanding the nature of an emergent transition, and his stepwise analysis of these three system types, each nested and ramified within its successor as it crosses the threshold of an emergent transition, provides a framework for understanding how the materially absent features of cognition and subjectivity, such as purposes, intentions, and functions, can *emerge* from underlying material and energetic processes. Deacon's goal is to preserve these quintessential features of mind and life; not define them into existence as Thompson and his forebears seem to do.

Purpose and Absence:

Incomplete Nature begins by invoking the partial, absent and incomplete aspects of systems which Deacon calls *absential*. This refers to any general attribute or property that is not materially present. These could be objects of reference, specific system states, abstract qualities or general types of properties (Deacon 3). He quickly refines this neologism by providing us with another: *ententional*. Whereas "absential" refers to a generalized relationship to constitutive lack, "ententional" aims to actively unite an entire class of phenomena in which absence can play quite varied and distinct roles. He says, "I propose that we use the term ententional as a generic adjective to describe all phenomena that are intrinsically incomplete in the sense of being in relationship to, constituted by, or organized to achieve something non-intrinsic" (27). Ententional language can be found in all manner of discourse. In philosophy its watchwords include *intention*, *meaning*, *value, reference, representation*; in the life sciences it can take the form of *function*, *purpose*, or *information*. For Deacon, the dividing line between animate life and inanimate matter is the constitutive role that such materially absent organizational features play in the maintenance of life's observable, measurable, materially present processes. He says, "This paradoxical intrinsic quality of existing with respect to something missing, separate and possibly nonexistent is irrelevant when it comes to inanimate things, but it is a defining property of life and mind" (3).

In defining these initial neologisms Deacon invites us into the basic problematic of his project. He quips, "What is absent *matters*, and yet our current understanding of the physical universe suggests that it should not. A causal role for absence seems to be absent from the natural sciences" (Deacon 3, emphasis added). The challenge Deacon sets is to demonstrate in a scientifically valid procedure how the materially absent can be rendered causally efficacious. *Incomplete Nature* attempts a naturalistic account of such constitutive absences and outlines how the scientific method, which is beholden to measurement and materialism, might begin reintegrating what it has spent its entire history systematically exorcising: the teleological dimensions of mind and life. We need to understand "ententionality" as a rhetorical and conceptual inroad into a larger methodological realignment.

Central to the book—indeed contained in the very title—is the notion of *emergence*. This concept has been a key feature of systems theory since its inception and arose from the recognition that when considering the boundaries of certain systems, one could observe "different levels of complexity with different kinds of laws operating at each level" (Capra and Luisi 65). A case in point is temperature, which is a bulk thermodynamic property clearly observable at a certain level of complexity—our coffee cup—but which becomes "meaningless at the level of individual atoms where the laws of quantum theory operate" (Capra and Luisi 65). In other words, as we ascend from the scale of subatomic particles and individual atoms to the bulk properties of myriad atoms and their interactions we are confronted with a novel property which is not exhibited at the lower levels.[35] Capra and Luisi say, "In the early 1920s, the philosopher C.D. Broad (1887–1971) coined the term 'emergent properties' for those properties that emerge at a certain level of complexity but do not exist at lower levels (65). Thus, the concept of emergence is characterized by a *transition* into a new level of complexity, one which "assumes that new general types of properties and physical dispositions—new causal habits of nature—can arise de novo" (Deacon 185). This "de novo" (i.e., "from the beginning" or "anew") is an important qualification because it underscores just how radical these emergent transitions can appear to a casual observer, as though the causal patterns that predominated prior to this transition have been supplanted by something inexplicably different.

[35] In the case of temperature, we can observe its effects by coupling (in this case, placing in direct contact) different systems. When we do this, we can observe a transfer of heat between systems. This is the essence of the zeroth law of thermodynamics, which states: "if A is in thermal equilibrium with B, and B is in thermal equilibrium with C, then C will be in thermal equilibrium with A" (Atkins 5). The zeroth law asserts the existence of a universal property (temperature) that allows us "to anticipate when two systems will be in in thermal equilibrium regardless of their composition and size" (Atkins 5).

As I have stated, the crux of the life-mind problematic can be located in this phenomenon of emergence because this is precisely where the prevailing laws of physics and chemistry appear to do an about face and start to operate in counterintuitive—and thus hard to explain—ways. In the opening pages to his chapter on emergence, Deacon makes this radicality clear: "The appearance of the first particles, the first atoms, the first stars, the first planets, and so on, marked fundamental new epochs in the 13-billion-year history of the universe, yet none of these cosmic transitions contorted the causal fabric of things as radically as did the appearance of life and mind" (144). For billions of years the basic laws of the universe described by physics and chemistry were sufficient: stars coalesced, heavy elements formed, then atoms and molecules, followed by planets, atmospheres, organic compounds, living things, and at the very end of this dizzying causal chain, human consciousness, symbolic thought, and society. These last three are not somehow separate from the latter.

As humanists, it can be odd (and perhaps even a bit alienating) to think of ourselves within this causal lineage as complex systems that organize matter and energy, but that is precisely what we do and precisely how emergent dynamics challenges us to reconceive of our mental and biological underpinnings. Deacon asks us to put aside cultural and ideological signifiers that compel us to see ourselves as agents of history, and instead understand ourselves as sites of thermodynamic interaction in which causality is channeled according to a new warp and weft.

Emergence is often articulated through the interlocking notions of autonomy and dependence. Emergent phenomena are dependent on their underlying physics and chemistry, but they exercise autonomy with respect to these parts because the whole is observed to have qualities that the constituent parts do not have on their own. This helps explain why autopoietic enactivism places such emphasis on the property of autonomy. By appealing to autonomous organization, Thompson and company believe they can offer a naturalized account of the origins of cognition, and by extension, subjectivity. It should come as no surprise that a paradigm case for emergence is biological life. Living systems are self-determining in ways that cannot be reduced to their underlying physics and chemistry. This contradictory relationship between autonomy and dependence is what defines the emergentist problematic. Deacon says:

> The fundamental challenge of classical emergentism is to make good on the claim that higher-order (supervenient) properties can in some critical sense not be reduced to the properties of their component lower level (subvenient) base, while at the same time being entirely dependent on them. So, if one agrees that there can be no difference in the whole without a difference in the parts, how can it be possible that there is something about the whole that is not reducible to combinations of properties of the parts?" (166)

On their surface emergent phenomena appear to instantiate new causal or physical laws. So the challenge becomes how to trace in stepwise fashion—and thus, in keeping with sound science—the precise series of energetic and material interactions that give rise to an emergent transition (à la materialism), without sacrificing what we observe to be unique in the emergent whole? This is the crux of Deacon's project: to show in scientifically valid detail how the ententional properties of mind *emerge* from matter, and without sacrificing the purposive and teleological character that makes it unique.

Accordingly, there are two methodological tendencies which Deacon repeatedly critiques and avoids. These are the "pitfalls of both crude reductionism and naïve emergentism" (231). The former assumes that material absence precludes causal efficacity, while the latter too easily avers the existence of the whole without due regard for its appearance from underlying parts. A good example of crude reductionism is cognitivism and its recourse to computationalism. The problem with this approach is that:

> There is no point where ententional dynamics just fades smoothly into thermodynamics. Minds are not just made of minds of simpler form made of minds of yet simpler form that eventually become so "stupid" as to be modeled by simple mechanisms. Nowhere down this ever smaller rabbit hole can we expect that the normal laws of causality will imperceptibly transform into their mirror image. (Deacon 139)

Cognitivism eliminates ententionality from its explanations by assuming that subjective mental states are epiphenomenal: effects of neural processes that have no causal efficacy of their own. This kind of reductionism avoids the problems posed by emergence by hiding the novel causal properties/laws in a deeper layer of description. A good example of naïve emergentism—at least according to Deacon's high standards—is the theory of autopoiesis.

As Deacon tells us, “The concept of emergence began to take on a new meaning in the last years of the twentieth century” (169), particularly the emergent process of self-organization which was due largely to “the insights gained from computation” (169). This marriage of computation and systems modelling is known as dynamical systems theory. In the field of dynamical systems theory mathematical models are used to “describe and predict the way an actual system changes over time” (Thompson 39). Simply put, a *dynamical system* is one whose state changes over time (Haddad 25), and in truth, *all* systems are inherently dynamical systems—subject to entropic change—because every observable unity or pattern of organization is undergoing some form of energetic and/or material flux.[36] As such, time is absent from no system. The distinction simply marks the manner by which a system is studied and mathematically modeled: either as a static arrangement in which time is eliminated from the description of parts and wholes, or dynamically, as a system in process.

All dynamic systems consist of “three elements—namely, a setting called the state space…and in which the dynamical behavior takes place…a mathematical rule or dynamic, which specifies the evolution of the system over time; and an initial condition or state from which the system starts at some initial time” (Haddad 25). These are terms which ought to sound familiar from our description of computation from chapter one—a

[36] “Open systems” exchange matter and energy. “Closed systems” exchange energy but not matter. “Isolated systems” exchange neither matter nor energy. Classical thermodynamics will speak in terms of “isolated systems,” but this is understood as an *ideal* system. In nature perfectly isolated systems do not exist. There is no such thing as a perfect insulator.

Turing machine is by this definition a dynamical system.[37] These "computational model systems allowed complete control and manipulation of the initial variables of an interactive process involving large numbers of component interactions and which could be set running for hundreds of thousands of iterations to see where these interactions led" (Deacon 169). The results were "both difficult to predict and surprising" (169) and help explain why emergence began to take on new meaning. For example, in chapter one we encountered an example that straddles the line between crude reductionism and naïve emergentism: artificial neural networks and their so-called "black box models" (Bray 111). We also need to include in this category cellular automata, which Deacon says are "Probably the most widely studied and influential computational exemplars of this new way of thinking about emergence" (170).

Ultimately, what distinguishes the mathematical modeling of dynamical systems theory is its qualitative nature. As Thompson says, "This approach is said to be qualitative because it uses topological and geometrical techniques to study the general or global character of the system's long-term behavior (its behavior in phase space), instead of seeking to predict the system's exact future state (a specific arrangement at a one precise moment)" (Thompson 40). The mathematical models used in DST create a qualitative picture of a system's behavior.[38] For example, *attractors*, which are stable regions in this

---

[37] This language also seems to closely follow Maturana and Varela's definition of "organization"—i.e., components and the rules for their interaction. Couple this with their assertion that "The cell defines its physical boundaries through its dimension of production of constitutive relations that specify its topology. There is no specification in the cell of what it is not" ("Autopoiesis…" 91), and I think the case can be made for why autopoiesis also demonstrates characteristics of both crude reductionism and naïve emergentism.

[38] In the previous chapter DST was used to denote "developmental systems theory." For ease of use, this chapter will use DST to denote mathematical dynamical systems theory.

topological (i.e., multidimensional geometrical space) towards which a system will tend to converge. These models arrive at the qualitative picture by mapping, in an iterative and mechanistic fashion, the innumerable interactions that follow from its precisely defined initial states. In other words, these models extrapolate the generalized behavior of natural systems rather than determine some discrete end state. DST and its many offspring, e.g., "…complexity theory, chaos theory, synergetics, or just complex systems theory (among other names)" (Deacon 173) helped to initiate a "shift in perspective from part/whole to process conceptions of emergence" (174), and to "legitimate its usage, free of philosophical baggage" (174).[39]

Despite these advances, Deacon makes it clear that his method *cannot* proceed according to this paradigm, as dynamical systems theories:

> …are ultimately forced to explain away the end-directed and normative characteristics of organisms, because they implicitly assume that all causally relevant phenomena must be instantiated by some material substrate or energetic difference. Consequently, they are as limited in their power to deal with the representational and experiential features of mind as are simple mechanistic

[39] To be clear, Thompson and Deacon are both inheritors and beneficiaries of the methodology and insights of DST because it helped to define and legitimate emergentist discourse in the hard sciences. Up to this point, I have cited Thompson and Deacon in turn because both texts produce clear expositions of its key features. It is telling that both Thompson and Deacon tread the same ground: both feel the need to acknowledge their systems theoretical forebears before distinguishing their own respective enterprises in relation to it. For example, Thompson—ever the phenomenologist—is wholly uninterested in mathematical modeling, and his exposition of the subject is used to ground his philosophical and phenomenological claims in a broader empirical context. For his part, Deacon also avoids mathematical minutiae, although he does provide the reader with many helpful illustrations and visual figures of attractors derived from DST methods, and incorporates many of its insights into his emergent dynamics framework.

> accounts. From either perspective, absential features must, by definition, be treated as epiphenomenal glosses that need to be reduced to specific physical substrates or else excluded from the analysis. (Deacon 5)

While it may seem strange to say given the complex and emergent processes that DST is able to successfully map, for Deacon, DST still falls within a reductionist paradigm because it has no procedure for explicitly accounting for the causal efficacy of absence. This is the key difference between DST and Deacon's emergent dyanmics. Iteration and the complex dynamical regularities that can follow from it are ultimately reducible to the system's initial state and rule, regardless of how strange and inexplicable of a qualitative picture it produces. While DST is far from "crude," its mathematical substratum renders it incapable of accounting for the ententionality that defines Deacon's investigation.

Yet it is the reductionism at the heart of DST that accounts for the renaissance of emergentism across the hard and social sciences (Deacon 174). At the same time, the profound order and regularity that can emerge from these initial chaotic states reinforces the fact that there exists no precise definition of a natural system because such systems are said to stand out from a background "as a single whole, as some observer sees and conceptualizes things" (Thompson 39). Even though we can map any number of self-organizing behaviors via the methods of DST, the observer's role in defining the system occurs in the initial state and the effectuation of the rule. In all cases, dynamical systems are salient *as systems* because some observer can distinguish and differentiate the organizational pattern, boundary, or states that define it.

Thus, a key step in Deacon's emergent dynamics will be to uncouple the notion of order—that is, organizational form or pattern—from subjective and situated preference. Deacon's method involves imagining the dynamics of systems in terms of material processes and their resultant configurations, but in a way that avoids imposing on these processes their formal/organizational characteristics from without, i.e., according to the preference or rules of some observer. This is the pith of Deacon's intervention into the discourse of dynamical systems theory. He will attempt to describe the dynamics of systems in a mind-independent way, one which does not depend for its description on the preference or perspective of some situated observer.

A whirlpool is a case in point. It is materially indistinguishable from the stream in which it emerges. Both are a moving collection of water molecules, but the vortex that is the essence of this dynamical system depends on the perceptual capacities of an outside observer to define and differentiate it, to distinguish this organizational foreground from its otherwise identical material background. Deacon says, "to attribute physical regularity to some perceived or measured phenomenon presumes a prior mental regularity or habit with respect to which the physical regularity is assessed" (189). In other words, the observer forms an essential part of the description by imparting his or her mind-dependent notions of order onto the system in question. In the case of the whirlpool, the notion of a vortex is an abstraction that arises from induction.

This is precisely why Maturana and Varela reject all teleological, teleonomical, and functional descriptions from autopoiesis. In autonomous systems the boundaries are by definition *self*-determined, and so any notions of order and formal organization that an observer implicitly wants to ascribe to the system become part and parcel of the actual, immanent organization of those boundaries. Stated differently, Maturana and Varela could not imagine a mind-independent description of order. Furthermore, the imperfect equivalence between the observer's notion of order and a system's immanent generation of order can obscure the precise nature of an emergent transition. One need look no further than the vexed descriptions of mentality and cognition we have seen in our examination of the cognitive turn: for much of the twentieth century, mind has consistently resisted coherent materialist explanations precisely because our own cognitive predispositions and sense of "order" obscure its mechanisms.

All of this is not to say that the observer perspective articulated by Maturana and Varela somehow ceases to exist or be epistemologically relevant: *All cognition is a situated and operationally closed phenomenon*. But Deacon's lodestone is not "cognition." He wants to understand the emergence of a new type of *causality*, and the implication of his theory is that understanding this emergent transition can, in fact, be done without recourse to the perspective of some outside observer. Therefore, we cannot conclude prima facie that one (cognition) is simply tantamount to the other (ententionality). As stated, central to Deacon's mind-independent description of "order" will be the notion of *constraint*. But before we arrive there, we first need to come to terms with a basic view of thermodynamics and its laws. This is the ground zero that distinguishes emergent dynamics from autopoiesis and enactivism.

Back to the Roots: The Basic Laws of Thermodynamics

The dissipation of a gas, the birth and decline of a hurricane, the steady cooling of my coffee as I write this passage—despite all appearances to the contrary, all of these systems follow the same logic as my stream of consciousness, that steady parade of thoughts popping into and out of my awareness throughout the day. All are systems whose underlying processes are happening *spontaneously*, without the intervention of an external influence. Deacon's goal is to reenvision all of these systems and their underlying thermodynamic processes via their most general habits of change, and in this way unite disparate and seemingly incompatible orders of nature under the banner of the famous second law of thermodynamics, which describes a general tendency towards *disorder* that is occurring at all scales of existence. The centrality of the second law to all of physics and chemistry means that its codification as a linguistic expression, and the unstated assumptions that are baked into its conception of "order," has influenced how scientists employ and understand this law in the exploration of novel problems such as the description of life and mind.

The study of thermodynamics can be divided into two essential types, each based on a different scale of observation. *Classical thermodynamics* refers to the study of "bulk properties"—things like the pressure, volume, and temperature of systems when viewed from the outside. Classical thermodynamics is largely a product of the nineteenth century and predates the widespread acceptance of atoms (Atkins 8). In contrast, *statistical thermodynamics* attempts to "account for the bulk properties of matter in terms of its constituent atoms. The 'statistical' part of the name comes from the fact that in the discussion of bulk properties we don't need to think about the behavior of individual atoms, but we do need to think about the average behavior of myriad atoms" (Atkins 9). Thus, thermodynamics encompasses both statistical inference and bulk mechanical description. Both modes describe the behavior of natural systems in ideal states. Throughout *Incomplete Nature* Deacon refers exclusively to the first and second laws and alternates between their classical and statistical expressions.[40]

The first law of thermodynamics is known as the law of conservation of energy and states that in a closed system energy is neither created nor destroyed, only converted into other forms. In other words, in an isolated system, the total amount of energy will remain the same.

The second law of thermodynamics is known as the "law of entropy." As with the first law, the second law can be stated either in terms of bulk properties or the statistical

[40] Ludwig Boltzmann was responsible for the turn towards statistical thermodynamics because he formalized the description of temperature in microscopic (atomic) terms. This formal expression became known as the *Boltzmann distribution.* Understanding it requires the following premise: that an atom can exist with only certain energies (Atkins 9). Atkins says, "At a given temperature—in the bulk sense—a collection of atoms consists of some in their lowest energy state (their 'ground state'), some in the next higher energy state, and so on, with populations that diminish in progressively higher energy states" (9). Here a *population* refers to the number of molecules at a given energy state. When a system has reached thermodynamic equilibrium (a situation in which atoms can still jump from level to level but where the overall population at that level remains the same) then the distribution of these populations can be accurately described by the Boltzmann distribution. In simple terms, the Boltzmann distribution refers to the "distribution of molecules over the available energy levels" (Atkins 13). The Boltzmann distribution says, "that for states of progressively higher energy, the populations decrease exponentially: there are fewer balls [atoms] on the high shelves than on the lower shelves" (Atkins 10). Or stated more formally, "The Boltzmann distribution is an exponentially decaying function of the energy. As the temperature is increased, the populations migrate from lower energy levels to higher energy levels. At absolute zero, only the lowest state is occupied; at infinite temperature, all states are equally populated" (Atkins 12). We can posit absolute zero and infinite temperature as inherent to the logic of the distribution, but these states are not empirically possible. In the bulk thermodynamics sense, the *state* of the system is its condition as viewed from outside, again, a function of observable properties like temperature and pressure. Thus, we can see how the Boltzmann distribution describes the condition of its parts as viewed from within.

inference of atomic properties.[41] Deacon says, "Although Rudolf Clausius coined the term entropy in 1865, it was Ludwig Boltzmann who in 1866 recognized that this could be described in terms of increasing disorder" (Deacon 378). Atkins provides us with a helpful example: imagine the molecules of a gas distributed over their various energy levels in a box. Now imagine that the overall energy in the box (i.e., the temperature of the gas) is kept constant, but the walls of the container are expanded. The Boltzmann distribution tells us that the same energy spread over a larger volume would cause the number of these allowed energy states to increase. In other words, there are more possible positions that the same number of molecules can occupy. Therefore, "the chance of choosing a molecule from one level in a blind selection decreases. That is, the entropy increases as the gas occupies a greater volume" (Atkins 53). As the box expands, there are more possible states for the molecules of gas to occupy, thus decreasing the observer's predictive powers, an inference that Boltzmann rendered in terms of *disorder*.

The second law of thermodynamics states that in an isolated system entropy will always increase towards its maximum: "in simple terms, things just tend to get as mixed up as they possibly can" (Deacon 108). This maximum state of entropy is also considered the equilibrium point of a system because it signifies that "any future change is as likely

---

[41] The second law of thermodynamics is the product of the accumulated efforts of French engineer Sadi Carnot and his work on heat/steam engines, Lord Kelvin, and Rudolf Claussius. In classical thermodynamics, the second law can be stated in a number of different ways. In Lord Kelvin's formulation: "no cyclic process is possible in which heat is taken from a hot source and converted completely into work" (Atkins 41), essentially that there is an "upper limit" in a system's capacity to transform heat energy into mechanical work; and in Clausius' formulation: "heat does not pass from a body at low temperatures to one at high temperature without an accompanying change elsewhere" (Atkins 42), this is where we get the notion that heat always flows "downhill," unless work is performed elsewhere to change this spontaneous tendency. Clausius sought a single equation that would synthesize these various formulations of the second law, and the result was the introduction of a new thermodynamic function: the entropy (Atkins 47).

to occur in one direction as the other, and over time directional changes will tend to cancel out, resulting in no net change overall" (Deacon 108).

Emergent dynamics asks us to put aside the commonplace (mechanistic) manner of describing energy in terms of its underlying substance (e.g., differences in mechanical force, differences in chemical bonds, differences in temperature, differences in electrical charge etc.), and realize that it is *difference as such* that makes energetic processes what they are. He says, "What 'flows downhill,' so to speak, in these transformations, is basically just a 'difference'" (Deacon 218). It is the tendency to equate energy (and by extension work) with their underlying substances that obscures the different types of change that can occur within thermodynamic systems, and most importantly, obscure the role of constraint on which much of his analysis will depend. He says, "Energy is more accurately, then, a relationship of difference or asymmetry, embodied in some substrate, and which is spontaneously unstable and self-eliminating—a tendency described by the second law" (Deacon 219).

The fact that all systems are subject to the second law means that their development over time is also one directional, always tending towards disorder, and in this way the overall movement of a system can also be described as asymmetrical (see above). Given the "upper limit" of a system to turn energy into work, all things in the universe will tend towards disorder. This is the irrevocable forward march of asymmetric change. Deacon says, "And once mixed up, things tend to stay that way. This spontaneous asymmetry in the pattern of dynamical change is the essence of the second law of thermodynamics" (220). The universe may have begun with a bang, but the second law of thermodynamics tells us it will most likely end with a cold, dark whimper. Thermodynamic equilibrium will beset everything. This is the so-called "heat death of the universe." It denotes a state of maximum entropy, meaning that heat gradients (i.e., energetic differences) will no longer be capable of interacting to perform thermodynamic work.

Life is Negentropy: Towards a New Causality

What makes living processes so special is that they defy this universal tendency described by the second law. Deacon says, "living and mental phenomena violate this presumably universal law. Expending effort to adjust conditions and to mobilize resources in order to alter future conditions toward an otherwise unlikely state is one of the most common attributes of living systems" (108). This strikes me as one of the best descriptions of "life" in the dispassionate language of systems theory one is likely to encounter. *Stated more simply*, life seems to fly in the face of this supposedly immutable movement towards disorder. It exists and persists in a far-from-equilibrium state. As soon as life arrived on the scene, not only has it been incredibly hard to eradicate—scientists have identified no less than five mass-extinction events over the course of Earth's history—but it has proliferated into ever-complexifying expressions of order (i.e., speciation). In other words, life seems to go against this universal second law. Life exists in a far-from-equilibrium state and gives rise to dizzyingly complex expressions of order. Life is "negentropy" (i.e., it negates entropy by increasing its orderliness).

Deacon starts to resolve this seeming paradox of living systems by going back to basics and focusing only on a thermodynamic system's most general characteristics of change. Therefore, now that we have some essential history and concepts in hand from classical thermodynamics we can better appreciate two more neologisms that will operate in all three classes of dynamic systems Deacon will go on to describe: homeodynamic, morphodynamic, and teleodynamic, and which speak directly to this issue of change. These concepts are *orthograde* and *contragrade* change.

Orthograde changes are those “changes in the state of a system that are consistent with the spontaneous, ‘natural’ tendency to change, irrespective of external interference,” while contragrade changes are those “changes in the in the state of a system that must be extrinsically forced, because they run counter to orthograde tendencies” (Deacon 223). In a simple thermodynamic system orthograde change is the spontaneous tendency towards disorder/entropy, while contragrade change arises from “the interaction between contrasting orthograde processes” (223). This means that, at bottom, all change is the result of orthograde processes, and there are many possible expressions of orthograde change “…involving different properties of things, such as temperature, mass, movement, electric charge, structural form…” (224). In classical thermodynamics, an “isolated system” is one that does not exchange matter or energy with its surroundings, but this is really just an abstraction for understanding how thermodynamic systems function. While we can certainly understand and measure distinct processes and properties—things like mass, electrical charge, and temperature—in practice these processes are constantly impinging on one another. Stated more simply and holistically, Deacon says, “it is simply because the world is highly heterogeneous that there can be contragrade processes” (224).

To understand why contragrade change must derive from the collision of two opposing orthograde tendencies, consider the coupling of two orthograde tendencies in a simple thermodynamic system: sugar dissolving in water. As Deacon tells us, orthograde change is the precondition for contragrade change. In this example, the orthograde tendencies inhere in the differing properties of sugar and water as each substance has a unique equilibrium point. Once these molecules begin to interact "The molecular collisions and electrochemical interactions between water and sugar molecules are contragrade, because the interaction among these molecules changes their spontaneous motions and ranges of movement" (Deacon 225). As these molecules continually bump into each other they change each other's "spatial positions and dynamical values" which contributes to a global change towards a new equilibrium point (225). I like to think of it like a layer cake: orthograde dynamics at one scale of observation (the atomic scale of things like mass, electrical charge, etc.), provide the basis for contragrade change at another scale (the molecular), which in turn provides the basis for orthograde change at the level of the coupled system (the glass of water). Ultimately, by redirecting our attention towards the spontaneous or forced manner in which change is organized, the orthograde/contragrade distinction "provides a language for describing the dynamical relationships that link different levels of a process" (226). As we move up the scale of complexity and through different types of emergent transitions, we can see how the orthograde dynamics at one level (i.e., the new equilibrium point of the sugar-water system) depend on contragrade processes at another level, back and forth up the systemic ladder like a layer cake.

Deacon acknowledges that at this point in the analysis orthograde and contragrade seem to merely restate the basic thermodynamic concepts of entropy and mechanical work, respectively. But as we ascend the scale of complexity and new system properties emerge, these neologisms will become necessary because they provide a number of affordances over the conventional language of thermodynamics and mechanics. Furthermore, by focusing on orthograde/contragrade change Deacon believes he can obviate the rather one-dimensional way that causation is conceptualized in physics and chemistry, that is, in terms of antecedence. For Deacon, this narrow view of causation often obscures the different types of change operating in a system and thus one's ability to describe the nature of the emergent transition under observation.[42]

We must always keep in mind that the second law of thermodynamics is effectively the *only* game in town. Therefore, the only way to make anything happen—i.e., create any kind of order—is to turn this process against itself by coupling systems with different equilibrium points. With these foundations of classical thermodynamics in place, as well as Deacon's concepts of orthograde/contragrade change, we are now in a better position to discuss the core of Deacon's emergent dynamics: *constraint*.

[42] To illustrate this point let us consider two facets of our sugar/water system (1) sugar spontaneously dissolving in water; and (2) extracting that sugar by chemical means. Deacon says, "both changes are caused in a generic sense, [but] what we mean by 'cause' is quite different in these two situations. Both are consistent with the laws of physics and chemistry; they nevertheless differ radically in how these causal influences are organized" (Deacon 225). In the former example, change is organized along orthograde lines; in the latter it is contragrade. It is Deacon's contention that focusing on general characteristics of change—whether spontaneous or forced—can unite a host of quite disparate physical phenomena such as temperature, mass, movement, and even formal (geometric) organization. This in turn enables more seamless movement between different types of dynamic processes as we begin to climb the scale of emergent transitions.

Constraint: A Mind-Independent Approach to Order

This takes us back to where we left off at the end of the first section ("Purpose and Absence"), and to Deacon's inverse, constraint-based approach to the concept of order/organization. The orthograde/contragrade distinction will allow him to develop a new dialectical understanding of causality that will unlock a mind-independent notion of order and pattern. We will see how order is the result of imposing constraints on a system.

Constraint is "the absence of certain potential states" in a system (198). They are "exhibited by what is not there but could have been" (Deacon 548).[43] A simple example of constraint is an internal combustion engine composed of fixed cylinders and movable pistons. By igniting gasoline and confining the explosion within the fixed walls of the cylinder this expanding thermal energy is constrained. It is this constraint on the expanding gas (an orthograde process, given that an ignited gas wants to *spontaneously* expand) that enables the internal combustion engine to transform thermal energy into mechanical work. In this example, the constraints are expressed in the immovable walls of the piston which are responsible for eliminating the myriad system states of the gas, winnowing it down such that it enables the bidirectional motion of the piston. Thus, in

---

[43] As we can see from Deacon's definitions, this does not just signify that the gas is inhibited or restricted in some generic sense, but that its myriad potential states/configurations have been eliminated. This is a subtle but important distinction because if one can find a way to more precisely circumscribe and define these unrealized potential states, then their absence might actually be able to assume some explanatory value. We will return to this idea momentarily.

the context of a simple thermodynamic system we see that constraint is “a critical factor in the capacity to perform work” (Deacon 198).

While the above example is helpful for understanding constraints in the context of classical thermodynamics, remember that what is ultimately at stake for Deacon is a *mind-independent notion of order/formal organization*. To get to this point we need to reconsider what precisely is meant by “pattern of organization,” a phrase which has been *the* primary point of departure for complex systems theory, beginning with organismic biology (Capra and Luisi 64).

To this end, Deacon says, “…patterns are an expression of redundancy. Redundancy is a defining attribute of dynamic organization…a form of organization in which some process or object is characterized by the repetition of similar attributes” (187). If we think about what defines a whirlpool, it cannot be the material itself, as its water molecules are constantly moving in and out of the system. Systems theory unequivocally tells us that it must be the formal organization, i.e., its spiral nature. But when we appeal to formal organization in this way, we produce a description of the system that relies on the perception of similarities and redundancies, of which the observer forms an essential part. In this way ententionality has been smuggled into the description.

This is where the concept of constraint intervenes. It is “a complementary concept to order, habit and organization because something that is ordered or organized is restricted in its range and/or dimensions of variation, and consequently tends to exhibit redundant features or regularities” (Deacon 548). In other words, constraint is like the photographic negative of a subjective, observer-dependent notion of order, pattern, and organizing relations. Deacon says:

> The general logic is as follows: If not all possible states are realized, variety in the ways things can differ is reduced. Difference is the opposite of similarity. So, for a finite constellation of events or objects, any reduction of difference is an increase in similarity. Similarity understood in this negative sense—as simply fewer total differences—can be defined irrespective of any form or model and without even specifying which differences are reduced. A comparison of specifically delineated differences is not necessary, only the fact of some reduction. It is in this respect merely a quantitative rather than a qualitative determination of similarity, and consequently it lacks the formal and aesthetic aspects of our everyday conception of similarity. (190)

Something that is patterned (similar) is also less different. Deacon is asking us to conceive of order *not* on the basis of repetition of similarities but on the elimination of differences. Constraint provides us with the photographic negative of our mind-dependent notion of order. It is "quantitative" because one proceeds by eliminating all possible variations that might have occurred but did not. This of course requires that we have a "finite constellation of events or objects," meaning that we can conceive, in principle, the totality of possible states such that the elimination of differences will ultimately yield the kind of redundancy we associate with an ordered arrangement. When we define order in this way we no longer appeal to abstract ideals (e.g., spirality per se); or to features of the system as they appear to a situated observer (e.g., the repetition of curves around a vortex). Instead, we proceed by eliminating possible system states, a process that can be quantified and computed by statistical means rather than subjectively described.

This may all sound rather abstract, but it is important to recognize that this quantification of order via constraints has a hugely important exemplar in the hard sciences: Claude Shannon's information theory. Thus, there is an important precedent for the kind of nontypical understanding of order that Deacon is advancing here. As we have seen, Deacon is a rather fierce critic of "crude" reductionist appropriations of Shannon's theory (e.g., cognitivism). But here he shows us how this kind of computational and statistical logic can be incorporated into new, nuanced emergentist perspective.

Ultimately, "Employing the concept of constraint instead of the concept of organization…not only avoids observer-dependent criteria for distinguishing patterns, it also undermines value-laden notions of what is orderly and not" (Deacon 195). This reframing of order represents a clear intervention into the subjective tendencies inherent in the language of thermodynamics and dynamical systems theory. Order and disorder are misleading metrics. They imply a situated and normative perspective; whereas the concept of constraint (and by extension orthograde/contragrade conceptions of change on which it depends) allows one to describe the state of a given system merely in terms of its spontaneous or forced tendencies. Finally, because constraint provides a way of describing physical systems and their interactions in terms of states *not* realized, it is an absence made manifest, one that is directly responsible for the formal patterns of organization that define systemic interactions. By emphasizing constraint rather than patterns of similarity and their material/energetic substrates, Deacon believes he can avoid the extremes of both naïve emergentism and crude reductionism.

For our purposes, the final step in the reimagination of order/organization is to translate constraint directly into the language of thermodynamics, a methodological maneuver that plays *the central role* in Deacon's emergent dynamics approach. Deacon says, "Rather than order or disorder, then, I suggest that we begin to think of entropy as a measure of constraint. An increase in entropy is a decrease in constraint and vice versa" (228). Armed with a clear definition of constraint and its relation to spontaneous and forced change, Deacon believes that he can more precisely pinpoint what defines an emergent transition. In the coming analysis of his three system types, we will see how these conceptual shifts can improve our ability to describe an emergent transition.

Recasting Thermodynamics: Homeodynamics

By virtue of coming to terms with these neologisms (orthograde/contragrade) and how they restate the basic ideas of the first and second laws of thermodynamics we have effectively described the first class of system Deacon describes in his book: *Homeodynamic systems* are characterized by an orthograde tendency which "spontaneously reduces constraints to their minimum and thus more evenly distributes whatever property is being changed from moment to moment and locus to locus" (Deacon 233). The word "homeo" from the Greek *homoios*, meaning "like" or "similar" denotes this elimination of differences, and the arrival of equilibrium (Liddell and Scott 487). The dissipation of a gas in a room, or the cooling of a coffee cup in one's hands evince this basic thermodynamic tendency towards constraint elimination. In other words, homeodynamics is Deacon's term for a classical thermodynamic system, but one rendered in the more general terms of orthograde and contragrade change and in its relationship to constraint.

As we move on to the next class of system we need to keep in mind (a) this basic idea that a homeodynamic system is characterized by an orthograde tendency to reduce constraints to their minimum, and (b) that in each emergent transition the properties and processes of the prior level become nested within the next. In other words, the constraint elimination that characterizes homeodynamic systems becomes subsumed to morphodynamic processes, just as morphodynamic processes become the grounds for ongoing teleodynamic processes. Even though the prior level seems to disappear from view, it is constitutive of the new orthograde tendency that will define the next system type to emerge.

Recasting Self-Organization: Morphodynamics

Morphodynamics refers to "a diverse class of phenomena which share in common the tendency to become spontaneously more organized and orderly over time due to constant perturbation, but without the extrinsic imposition of influences that specifically impose that regularity" (Deacon 237). The root "morpho" comes from the Greek *morphē* meaning "form" or "shape," and signals that this class of system deals explicitly with formal or geometrical regularity (Liddell and Scott 452). These are systems that directly confront the observer with the phenomenon of self-organization and dynamical regularity, and as the quotation above tells us, this formal regularity is the direct result of constant perturbation, i.e., energetic input from without. Accordingly, Deacon provides us with a formula for describing morphodynamic processes: “High, uniform rates of perturbation lead to the production of second-order intrinsic constraints that eventually balance the rate of constraint dissipation to match the introduction of constraints from outside the system” (253-254). Morphodynamic systems are distinguished by these balanced rates of constraint production and elimination, a parity which has important implications on the order that can be observed in the system.

A classic example cited in both Thompson and Deacon is Rayleigh-Bénard convection. This is a phenomenon that occurs when a temperature gradient (i.e., a sustained energetic input) is imposed on a fluid medium. The heat energy is transferred to the fluid via convection. As the fluid is heated it becomes less dense and rises, as the fluid cools it becomes denser and sinks. In Rayleigh-Bénard convection, the epicycles of fluid movement are further constrained by a number of other physical factors like viscosity, specific gravity (the density of a substance relative to water), the depth of the fluid, and the intensity of the temperature gradient itself (Deacon 251). Given the proper combination and degree of these constraints, the epicycles of fluid movement will regularize into well-defined hexagonal shapes because this is “the most densely packed way to fill a surface with similar-sized subdivisions” (252). The resulting geometrically patterned convection cells—known as Bénard cells—are phenomena that manifest throughout the natural world: on the surface of the sun; in columns of cooled volcanic basalt the world over, such as the Devils Postpile in California; and—closer to hearth and home—in oil in the bottom of a hot cooking pan.

There exist many examples of morphodynamic processes in nature, such as the formation of snowflakes, the action of a whirlpool in a stream, the folding of proteins in a cell, or even the totality of the Earth's biosphere, first formulated in self-organizing terms as *Gaia* by James Lovelock and Lynn Margulis. In all these cases, we find that the system's underlying homeodynamic tendencies towards constraint elimination can be indefinitely forestalled provided that energy is continually added from without, a dynamic which keeps the system in a far-from-equilibrium state. We need to recognize here that morphodynamic systems, though common enough in nature, still require very specific conditions to form. This is because constraint dissipation and constraint production need to achieve an ideal parity. Deacon says, "Balanced rates are achieved when the rate of entropy production increases to the point that it keeps pace with the rate of disturbance by generating more dimensions of internal constraint that can now be simultaneously eliminated" (254). The implications of the second law mean that entropy will always win out in the end. In the terms outlined by emergent dynamics, this means that there will always be an underlying tendency towards constraint elimination. Morphodynamic systems yield to this inexorable demand by simultaneously generating, propagating, and eliminating constraints, processes which, on their surface, appear to contravene this law because in a morphodynamic system this takes the form of more orderly, redundant, or regularized configurations.

In the case of Bénard cells the macroscopic constraint of geometric, hexagonal packing is preserved and propagated throughout the system because it turns out to be the most efficient way of eliminating heat (i.e., realizing the underlying homeodynamic orthograde tendency towards constraint elimination). Deacon says:

> Because correlated movement of large collections of molecules involves fewer dimensions of difference than uncorrelated local molecular movement, the system is by definition in a simpler and thus more orderly state. What the external perturbation coupled with the effects of buoyancy and the geometric constraint of hexagonal close packing together accomplish is to increase and redistribute constraint in a more symmetric way. This redistribution of constraint into other dimensions of difference provides more ways for constraints to be eliminated as well. This is what creates the higher rate of constraint elimination (and thus entropy production). (253)

This quotation gets to the heart of why self-organizing systems are so counterintuitive and strange to the average observer. The second law of thermodynamics dictates that things can only get *more* disordered over time, but under the proper conditions—chief among them, constant energetic input—localized forms of "order" can arise, and this localized order actually allows for a greater overall rate of constraint elimination, thus yielding to the demands of the second law and its ineluctable march towards *disorder*.

Bénard cells, like all morphodynamic processes, "exhibit an effective reversal of the typical thermodynamic orthograde tendency: that is for macroscopic constraints to be progressively eliminated through microscopic dynamical interactions" (Deacon 252). In a basic thermodynamic system it is the lower-level molecular interactions—for example, the myriad contragrade collisions between sugar and water molecules in the glass—that will eventually eliminate the macroscopic differences in the system and lead to equilibrium (maximum entropy). But in morphodynamic systems, in which far-from-equilibrium states are achieved and maintained, this bottom-up tendency of constraint elimination takes counterintuitive forms.

For his part, Thompson puts the matter not in terms of macro/micro, but in terms of global/local. In Bénard cells, these "local" behaviors are the individual heated fluid molecules colliding, creating density differences and ultimately convection currents, which together with other constraints (specific gravity, depth etc.) create the "global" system property we see in the form of hexagonal pattern. These hexagonal shapes in turn further "constrain or govern the behavior of the individual components, entraining them so that they no longer have the same behavioral alternatives open to them as they would if they were not interdependently woven into the coherent and ordered global pattern…*This two-sided or double determination is known as circular causality*" (*Mind in Life* 62, emphasis added). Thus, we see that morphodynamic systems also exhibit the essence of cybernetic control—circular causality. In a self-organizing system, part and whole have become totally interdependent.

The Gaia hypothesis presents us with the same morphodynamic organization but on a much vaster scale and with far more degrees of difference at play (i.e., internally generated constraints). In this example, the sun provides the energetic input that places Earth in a far-from-equilibrium state. This constant application of solar energy is what enabled distinct subsystems and their constraints to arise and propagate, such as the water and carbon cycles. Over the course of Earth's 4.5-billion-year history as these cycles became established, still more complex processes emerged, the most important of which was the emergence of anaerobic bacteria and later aerobic cyanobacteria whose capacity to photosynthesize led to what is now referred to as the "great oxidation event" that created our oxygen-rich atmosphere and ultimately, the explosion of complex life on Earth. The Gaia hypothesis describes the Earth as a unitary, self-regulating cybernetic system in which the biosphere is tightly coupled to these myriad abiotic processes (i.e., water, carbon, nitrogen cycles etc.), and which maintains Earth's life-supporting outer shell in a state of beneficent homeostasis. The dynamic coupling of these myriad processes meant that, "rather than burning out, Earth's highly combustible mixture of reactive gasses has remained in a far-from-equilibrium state for hundreds or thousands of millions of years" (Clarke 25). In effect, the Gaia hypothesis is a description of Earth as a morphodynamic system.[44]

The crucial point to remember in all of these examples is that this constant energetic input does not determine or linearly direct the system's precise structural regularities (orderliness). Constant energetic perturbation maintains the system in a far-from-equilibrium state, which then allows its internally generated constraints to manifest

---

[44] If we extend the layer cake analogy into the different system elaborated in emergent dynamics, we will see that teleodynamics assumes the top "layer" insofar as it is the last, and, accordingly, most complex system to emerge. Gaia would seem to provide an interesting counterpoint here because it is a morphodynamic system that emerges out of the juxtaposition of both homeo and teleodynamic systems. While Deacon's analysis is right to proceed in this order given that each system nests within its successor, it also has the effect of elevating—conceptually and normatively—a certain class of system, and thus with improper handling could easily contribute to a false sense of anthropocentrism. If we think about the myriad processes of Gaia (of which biological life is a constituent) in *rigorously* morphodynamic terms—meaning as a process which ultimately must yield to the demands of the second law—then a natural conclusion to draw is that biological life is essentially the most effective way of eliminating constraints! Albeit through the highly regularized, self-organizing, and as we will see, teleodynamic (cognitive) patterns of organization that abound in nature.

and propagate. As these various constraints interact and propagate, their interactions begin to stabilize (or simplify) creating redundancies and regularities which can be observed as a higher form of "order."

As Deacon acknowledges, most researchers and systems theorists would simply call these *self-organizing processes* (Deacon 244), because in all these examples we are confronted with what Heinz von Foerster has famously called *order from noise*. This is the essence of a "self-organizing" system. Where once there was chaotic interaction and a tendency towards disorder, we now observe regularized and patterned organization. In this way, we can see a fundamental congruence between the second-order cybernetic descriptions of order and self-organization, and Deacon's description of a morphodynamic system.

But Deacon begins his chapter on morphodynamics by distancing himself from the language of self-organization on grounds that should, by now, sound familiar. For one, the language of "self" organization can tend towards naïve emergentism because it erroneously assumes that the pattern of organization is both cause and consequence of the system—an example of fallacious circular reasoning. This conflation arises from thinking about the system in linear causal terms (i.e., decomposable inputs and outputs), when really (as Thompson's earlier quotation illustrates) such systems are *circularly* causal: part and whole are wholly interdependent, meaning that we cannot isolate or conceive of one without the other.[45]

---

[45] In this context, the emergent transition into morphodynamics can benefit from distinguishing between *intrinsic* versus *extrinsic* constraints. In the example of the internal combustion engine, the constraints

Secondly, the language of "self-organization" implies a kind of agency in which the process is being actively directed towards this regularity, when in fact a more apt description is to say that these processes "fall towards" regularity (Deacon 245). The appearance of a geometric/formal order from out of a "noisy" thermodynamic background makes it all-too-easy to ascribe causal efficacy to that perceived order. This is one explanation for the top-down views of causality we see in naïve emergentism, where the macroscopic constraints like hexagonal packing or the vortex in a whirlpool are assumed to be causative of the lower-level effects. Upon closer examination, we find that the order exhibited at the macro-scale does not unidirectionally determine the behavior of its constituent parts but represents a compounding of constraints generated at lower levels. This is not top-down causality, but simply "an alteration in causal probabilities" (Deacon 161).

---

were extrinsic because they did not alter the physical properties of the gas (things like mass, charge etc.). The constraining force of the piston walls merely limited the physical expansion of the gas. In contrast an intrinsic constraint is one that alters molecular properties. To this end, Deacon provides us with two illustrative examples. The circular shape of a wheel represents an extrinsic constraint on the relations between its molecules; whereas "…in the organism, the properties of the molecules are a consequence of their incorporation into this system" (Deacon 163). He continues, "One of the crucial differences is that the emergent relationship in the wheel example is synchronic; but whereas there is a synchronic analysis possible for the organism, the property that is crucial to the constitution of its proper parts is the dynamics of reciprocal synthetic processes that is intrinsically diachronic…Thus, at least for higher-order forms of emergence, the part/whole distinction and the synchrony/diachrony distinction are intertwined" (Deacon 164). In simple systems synchronic description suffices because it is constituted by extrinsic constraints expressed all at once as a function of bulk properties. In such a system, all of its parts are decomposable. In complex systems such as life, diachronic description is essential because its constraints are implicit and arise through continuous, ongoing global processes. Thus, so-called "top-down" causality—the idea that a "higher-order emergent phenomena can alter phenomena that they supervene upon…" (Deacon 553)—is only problematic when a system's intrinsic constraints are viewed statically and without accounting for the ongoing dimension of time. Top-down causal perspectives are a common whipping boy for crude reductionist approaches trying to discredit emergentist logic, but these approaches are only problematic when process, change, and the intrinsic nature of a system's constraints are not fully accounted for.

So, while self-organizing systems are widely recognized as emergent phenomena because they create order from noise and appear to bend the second law of thermodynamics according to a novel causality, Deacon's emergent dynamics allows us to more rigorously describe the underlying dynamics and, as a result, more precisely identify it as a discrete class of system. In the terms of emergent dynamics, these self-organizing systems are emergent because they create new "orthograde dynamical regimes" (Deacon 261). This turns out to be *the essence of an emergent transition*. The far-from-equilibrium states achieved by self-organizing systems induce a novel orthograde tendency. In the case of morphodynamic systems this novel orthograde tendency is *constraint propagation*. In contrast to the hegemonic, sedimented way in which systems theory has hitherto addressed the phenomenon of emergence and self-organization, Deacon's emergent dynamics eliminates all recourse to the imprecise language of "order" in favor of constraint and the resulting dynamics of spontaneous/forced change.

Recasting Autopoiesis: Teleodynamics and the Autogen

The final step in Deacon's revisioning of thermodynamics and systems theory—a.k.a. *emergent dynamics*—and the final level of his nested hierarchy, is teleodynamics. According to his approach *an emergent transition occurs when a new orthograde tendency is established*. Whereas the orthograde tendency of basic thermodynamic systems is the steady elimination of constraints (entropy), and the orthograde tendency of morphodynamic systems is the localized amplification of constraints (self-organization), what defines a teleodynamic system is the preservation and reproduction of constraints—preservation *through* reproduction. As the prefix denotes, this is the emergent transition associated with purposes, goals, or aboutness. In other words, ententionality first emerges in the transition to teleodynamics.

Teleodynamic systems are not identical to biological life. To avoid confusing these categories we ought to keep in mind the tools of emergent dynamic description as Deacon has painstakingly constructed them. Although life is perhaps the most immediate example of a teleodynamic process emerging from an underlying morphodynamic process, "this does not mean that the origin of life is the only threshold leading to teleodynamics, or that only life can be teleodynamic" (Deacon 271). Deacon's emergent dynamics describes an abstract set of causal principles that distinguish between different types of emergent transitions, *not* a description of life per se. In contrast, the theory of autopoiesis describes the organization of the living as such. It was drawn directly from the paradigm of the single cell and designed in service of it. Deacon's method assumes no such indelible connection to biology. It is precisely for this reason that I have chosen *Incomplete Nature* as the vehicle for describing autopoiesis. Returning to this chapter's epigraph—by looking askance at these processes I believe we can allow the "formal self-reference" that defines *cognition* in second-order systems theory to come more fully into view.

In his chapter on morphodynamics Deacon examines several well-trodden examples of morphodynamic processes such as snowflakes, auditory resonance, and Raleigh-Bénard convection. But in order to track the emergent transition to a teleodynamic system, Deacon creates a single "minimal proof of principle" (290), a thought-experiment that he calls an *autogen*, and the processes that give rise to this system, *autogenesis*.

Emergent dynamics demonstrates that the coupling of different homeodynamic systems can, under the proper conditions, lead to the emergence of a morphodynamic system. In like manner, "teleodynamic processes must be constituted by reciprocally synergistic morphodynamic relationships" (Deacon 325). In order to avoid being unnecessarily complex, as there is no upper limit to the coupling of morphodynamic processes, Deacon's description of autogenesis imagines the coupling of two and only two morphodynamic processes which will form the foundation of the new teleodynamic orthograde regime. The first is *autocatalysis*; the second, *self-assembly*. By reciprocally linking these two morphodynamic processes, a third phenomenon, *containment* becomes possible. Containment is the result of self-assembly and autocatalysis, and the key to their persistence. He says, "Containment creates physical individuality. A boundary that distinguishes between inside and outside is almost synonymous with the self/other distinction, both functionally and metaphorically" (296). Containment is the creation of a physical boundary, and it is both an intuitive and functional marker of systemic unity. In the above quotation, the qualifier "almost" will turn out to be the decisive point of difference between autopoiesis and autogenesis.

Catalysts are molecules that increase the rate of a chemical reaction without themselves becoming subject to, or used up by, that reaction. *Autocatalysis* compounds this process. It is "a special case of catalytic chemical reactions in which a small set of catalysts each augment the production of another member of the set, so that ultimately all members of the set are produced" (Deacon 293). In autocatalysis, like produces like, creating a positive feedback loop of chemical reactions that will persist until all underlying substrates are exhausted. Deacon says, "although autocatalysis is a function of molecular diversity, it turns around and generates molecular homogeneity, since molecules that tend to produce other molecules from a linked autocatalytic set rapidly replace other forms that do not" (294). Because these reactions feed on their own products, they rapidly undermine themselves. They are "one of the most relevant classes of non-equilibrium chemical process…because there are many analogues to this in the living cell metabolism and because it can arise spontaneously as a transient, locally deviant, non-equilibrium process" (292). In other words, autocatalysis is an ideal candidate for imagining autogenesis because it is a well-recognized, far-from-equilibrium (morphodynamic) process. It propagates constraints and creates localized order and homogeneity. But precisely because it creates homogeneity in surrounding substrates its existence is extremely fleeting. Autocatalysis is widely recognized as an essential factor in cellular metabolism.[46] *In this way, Deacon recapitulates a key aspect of autopoietic theory: the internal reaction network of the cell.*

The second process in the formation of an autogen is *self-assembly*. Deacon likens the basic process of self-assembly of macromolecules to that of crystal lattice formation,

---

[46] Here it is worth mentioning that there exists a higher-order form of autocatalysis "in which two or more autocatalytic cycles are interdependently linked to create cycles of cycles" (Deacon 294). These were dubbed "hypercycles" by the Nobel prize winning chemist Manfred Eigen, whose research on the subject began in the 1970s, but which to date, has yielded no experimental proof of their viability. Deacon acknowledges that autogenesis is superficially similar to Eigen's hypercycle theory in that they both imagine self-promoting processes that "also promotes the other in some way" (Deacon 308).

in effect, "an expression of the intrinsic geometry of component molecules, the collective symmetries these offer in aggregate, and the lower energy state of the crystallized forms" (298). Crystal lattice structures are highly regularized, repeated patterns of atomic organization. The crystalline structure provides more stability for the atoms under certain conditions. For example, the extremely durable crystalline structure of diamonds is a product of carbon atoms when placed under extreme heat and pressure, whereas the extremely weak crystalline structure of ice forms under far less intense temperature variation. In both cases, given the right conditions, atoms will spontaneously assemble into these repeated crystal lattice structures.

In the case of the autogen, we are not dealing with atoms but *macromolecules*—polymers made up of many subunits, or monomers. Ultimately:

> The accretion of molecules into crystallinelike arrangements is thus a function of thermodynamics, but the influence of their structural and charge characteristics *may* contribute to amplification of constraints, thereby generating regularities in the ways they form into aggregates; a morphodynamic consequence. (Deacon 298, emphasis added)

Deacon seems a bit cagey here. On the one hand, these structures seem to self-assemble as a spontaneous orthograde process; but Deacon leaves the mechanism intentionally vague and seems to take for granted that given a sufficiently rich/diverse molecular environment this process can become self-perpetuating and therefore qualify as a morphodynamic process. It is also worth noting that the term *macromolecule* has different implications in the fields of biology and chemistry, and Deacon does not specify precisely what he means—whether he is referring to a general class of large molecules formed by weak bonds between atoms, as defined by the field of chemistry—or biological macromolecules such as carbohydrates, lipids, proteins, and nucleic acids. He goes on to adduce several types of biological self-assembling structures, such as microtubules, lipid bilayers, protein and carbohydrate matrices etc., so it is safe to assume that for the purposes of autogen formation, biological macromolecules offer the clearest path forward. *In this way, autogenesis and macromolecular self-assembly recapitulates yet another step in autopoietic theory: the formation of a semi-permeable membrane or boundary.*

The basic pattern of autogenesis is as follows: Autocatalysis requires a steady availability of the necessary substrates to maintain its catalytic reactions, while self-assembly (understood on the basis of crystal lattice formation) requires a steady availability of identical macromolecules so that stable containment structures can form. Because it homogenizes the surrounding molecular soup, autocatalysis can create the ideal conditions for self-assembly; while self-assembly of macromolecules can create a physical boundary that can keep necessary autocatalytic substrates from diffusing away. When these two morphodynamic processes become coupled, “Raw materials would be produced that are conducive to container formation. An enclosing structure would tend to form in the vicinity of the most active autocatalysis and thus would spontaneously tend to enclose autocatalytic set molecules within it. And closure of the container would prevent dissipation of the components of the autocatalytic set” (Deacon 304-305).

Other theories of molecular self-organization such as Manfred Eigen's hypercycles, in which autocatalytic cycles are "interdependently linked to create cycles of cycles" (Deacon 294) do not provide a mechanism for obviating the runaway self-undermining process of autocatalysis. Autocatalysis is "self-promoting, but not self-regulating or self-preserving" (295) and so its tendency to rapidly use up underlying substrates makes hypercycles—which depend on these already unstable processes—extremely susceptible to even the smallest disruption (295). What distinguishes autogenesis is the phenomenon of closure/containment. Physical closure of the autocatalytic set is the mechanism that checks this runaway process, and which has the potential to inaugurate "self-repair, self-reconstitution, and even self-replication in a minimal form" (306). Deacon says:

> though each component process [of the autogen] is self-undermining in isolation and co-dependent, together they are reciprocally self-limiting, so that their self-undermining features are reciprocally counteracted. Thus, whereas substrate exhaustion leads to both autocatalytic and hypercycle cessation and component dispersion, an autogenic system will establish its capacity to re-form before exhausting substrates, *so long as closure is completed before this point is reached*. (Deacon 308, emphasis added)

Deacon asserts that "In its inert closed state, an autogen can maintain this potential across vast epochs of time and through diversely non-supportive contexts" (310). This capacity for stasis exists because the self-assembling macromolecular crystalline structures of the container are more energetically stable than the surrounding molecular soup and can resist the forces of molecular diffusion across all manner of unsupportive environments. Of course, once closure happens, it is no longer possible to sustain autocatalysis because the supply of necessary substrates has been cut off. Yet this closure does not impede the forward march of constraint preservation, it enables it.

Closure is the key to autogen self-maintenance (identity) because once full closure of the autocatalytic set is achieved the scales of causal probability will have tipped into a new spontaneous tendency to preserve the synergy of these constraints despite the constant changes in matter and energy that continually attempt to disrupt it from without. Closure means that the system dynamics have become sequestered and the thermodynamic work they perform has become sequestered along with them—now directed inwards towards the autogen's own self-maintenance. In this way the autogen evinces yet another basic requirement of cognitive systems: the openness-from-closure principle. If we compare this to a bacterium, the bacterium has developed highly sensitive mechanisms for "gatekeeping," allowing only beneficent molecules to pass through its macromolecular structures. It has also developed the capacity for chemotaxis. In contrast an autogen can only engage in this type of selective behavior by rupturing and reforming. In essence, the containment of the autocatalytic set and its continual ruptures and reformations acts like a highly pared down version of a cell membrane.

Deacon refers to autogenesis as a negentropic "ratcheting" effect. Recall that in the language of emergent dynamics, a decrease in entropy (negentropy) is an increase in constraints. In autogenesis once the critical tipping point of closure is achieved the balanced rates of entropy production and elimination that characterize a morphodynamic process becomes fundamentally altered: autogenesis actually forestalls entropy production. Deacon says, “Regularity is built up, only to be frozen by closure. Full dissipation is prevented at a point where optimal conditions for rapidly reinitiating the process are achieved. An autogen is thus effectively a negentropy ratchet” (317). Here the picture we get of autogen behavior is an ongoing process of container formation and dissolution—little bubbles of teleodynamic organization coalescing into, and popping out of, static and dynamical states. The ratchet is an apt metaphor because it illustrates that once achieved this tendency towards constraint preservation will tend to move forward, so to speak, not back.

The ratcheting metaphor is also apt because it underscores the purpose of emergent dynamics: to demonstrate the emergence of new orthograde tendencies in systems. This is the burden of proof for an emergent transition. In the case of autogens it is clear that the morphodynamics of constraint propagation have been supplanted by a new teleodynamic tendency towards constraint preservation (318). Deacon says, “In the face of catastrophic perturbation, the physically dissociated components nevertheless retain their systemic identity sufficiently well to be able to reassociate into identically organized unit structures. This indicates that these components are to some extent present *because* of their contributions to a higher-order whole” (321). All the physically measurable molecular phenomena are present, i.e., localized and engaged in their synergistic morphodynamic processes because of a higher order of organizational unity which is constraining their interactions from the top down and contributing to their persistence. Yet in the case of teleodynamics this new tendency also inaugurates a fundamental change in how thermodynamic work is organized. In contrast to a morphodynamic processs, an autogen's localized reduction in entropy is no longer the result of a continuous throughput of energy characterized by balanced rates of entropy production and reduction, but a new, stable, orthograde tendency towards constraint preservation. The interdependent suite of morphodynamic process that constitute the teleodynamic system determines the boundaries of constraint propagation, and in so doing, turn their thermodynamic work in on itself. In this case, the boundary condition is an actual physical boundary—a container that checks the runaway constraint propagation, and that effectively reduces the complexity of the system.

Deacon says, "The autocatalysis, the container, and the relationship between them *are generated in each replication precisely because they are of benefit* to an individual autogen's integrity and its capacity to aid the continuation of this form of autonomous individual" (321). Thus, the totality of these processes results in what Deacon labels an autonomous individual, an individual founded on purposive behavior, namely, the preservation of its own pattern of organization. It is the emergence of this new tendency towards constraint preservation that justifies the "telos" in teleodynamic, implying that *this new systemic unity exists with its own end in view*. By this reckoning, the system's autonomous properties are coextensive with its teleodynamic organization. We will revisit the implications of this autonomy in greater detail in the following chapter because according to Thompson's autopoietic enactivist account, autonomy is *the* defining feature of a living/cognitive system. In this chapter we have seen how Deacon has been able to explain this autonomy in precise, stepwise detail—i.e., how an autonomous individuality emerges from underlying thermodynamic processes.

At the same time, this individuality is "strangely diaphanous" (Deacon 310) in the sense that it is only individuated when closure is achieved and the autogen is inert; when it is open, actively accumulating molecules, and reconstituting itself it is no longer a discrete individual.[47] The realization of a closed boundary is the condition of an autogen's teleodynamic organization. But this boundary, when fully closed, effectively halts the underlying morphodynamic work cycles that constitute its invariant organizational pattern. *In this way, autogenesis recapitulates yet one more feature of the theory of autopoiesis: operationally closed/structurally open.* For Deacon, this combination of openness and closure is evidence of the so-called "virtual" character of autogen identity: the autogen exists and persists *between* two states, thereby resisting reduction to either one. It is for these reasons: (a) persistence of morphodynamic processes for the sake of a higher-order unity, and (b) the virtual character of that identity, that "an autogenic system confronts us finally with an unambiguous absential quality. This is the essence of teleodynamical causality" (Deacon 311).

Finally, I think this immanent purposiveness is further demonstrated by the fact that autogenesis also entails "a persistent, if weak, form of natural selection" (Deacon 313). We know there are three preconditions to natural selection: reproduction, variation and inheritance, and competition. Autogenesis satisfies all three.

Upon dissolution of the container one autogen can just as easily become two or more because the autocatalytic dynamics that help drive container formation have

[47] Recall Deacon's earlier quotation where he states that a boundary is "almost" synonymous with the self/other distinction (296).

already become established and localized. This is reproduction. Next, as autogens break apart, reform, and reproduce, it becomes possible for these new iterations to incorporate different molecules present in the environment, some of which could affect autogenic function. He says, “incorporation of functionally interactive molecules will make a given lineage relatively more or less successful in propagation” (Deacon 312). This variation and inheritance of functionally advantageous features (traits) inevitably leads to competition, i.e., more successful phylogenetic types or lineages. Working backwards then, we can use this minimal proof of natural selection logic as a further sign of purposiveness. If different autogenic unities are *competing* with one another for resources (in this case, advantageous molecules that aid in the reproduction of the autogenic work cycle), then it stands to reason that they are animated by a purpose: an implicit striving towards self-preservation.

Do Autogens Have Cognition?

To be clear, nowhere in his chapter on autogenesis does Deacon argue that autogens are *cognitive*. Nowhere in his explication of autogenesis does the language of cognition explicitly occur, although interestingly Maturana and Varela (along with Stuart Kauffman) are primary intertexts. Despite the omission of “cognition,” I think we have made a number of inroads towards the scientific methodology outlined by Thompson in his formulation of the body-body problem at the end of the previous chapter. Deacon’s emergent dynamics and its ability to track the constitutive role of constraints through various types of emergent transitions demonstrates how we might arrive at the rudiments of telos and autonomy via a scientifically valid, mind-independent description of systems and their hierarchically nested emergent processes.

Therefore, I would like to close this case study with the following questions: (1) Where does Deacon's autogen leave autopoietic enactivism's working definition of "cognition"—as both formal self-reference and active (i.e., agential) selection of salient environmental stimuli? Do autogens enact their environments in the same way that a single cell does?

Autopoietic enactivism's recourse to chemotaxis as a minimal proof of cognition demonstrates that *autonomy* in this theory implies not only the preservation of an underlying pattern of organization (what Deacon would render as the preservation of constitutive constraints), but the system's capacity to reorient itself and/or move towards more favorable conditions—sensation guiding *action* that shapes sensation. Bacterial chemotaxis—movement towards or away from helpful or harmful chemical concentrations—makes the agential and selection mechanisms inherent in formal self-reference visibly manifest. Thus, in addition to defining the basic formal characteristics of cognition, autopoiesis implies a behavioral characteristic, one that is coextensive with the formal and which issues from within its very systemic identity. By creating itself life also determines itself *as a function of self-movement*. It is an act, a doing.[48]

---

[48] We see this stated even more clearly in the *Tree of Knowledge.* Maturana and Varela say, "we wish to examine the phenomenon of cognition by considering the universal nature of 'doing' in cognition—this bringing forth of a world—as our problem and starting point…" (28). In this passage, we can also clearly perceive echoes of the enactivist position (e.g., "bringing forth") that would emerge only a few years later with the publication of *The Embodied Mind*.

Emergent dynamics outlines several interesting parallels to the organization of living systems as described by autopoiesis, but according to Deacon autogens are decidedly *not* alive. In this respect Deacon's autogen might seem to muddy the cognitive waters, especially given Maturana's unequivocal alignment of cognition with life.[49] In a claim that seems to further muddy the waters between autopoiesis and autogenesis, Deacon says that autogens are not alive precisely because they are *passive*:

> Autogens may be subject to natural selection and evolution but they are not alive in most senses. They lack the majority of attributes associated with living organisms today. Most significantly, they are effectively passive—though not inert—structures, because they do not *actively accumulate* and mobilize energy within themselves that can be used for self-repair and reproduction. (Deacon 315, emphasis added)

This passivity is the first key difference between autogenesis and autopoiesis. Autogens are fundamentally passive systems because they emerge wherever conditions are favorable to their formation, and their capacity to break apart and reform is initiated from without. To be sure, they are capable of decreasing their entropy, instantiating order, and reproducing a work cycle, but they are also subject to the vicissitudes of their immediate environmental conditions—a passivity not present in the single cell and its innate capacity for chemotaxis. Interestingly, this passivity Deacon likens to the structure of a virus (315), a comparison which further disassociates the autogen from the realm of the

[49] "*Living systems are cognitive systems, and living as a process is a process of cognition*. This statement is valid for all organisms, with and without a nervous system" (Maturana 13).

living and from the perspective of enactivism and placing it in a further cognitive “grey area.” Given these differences, the most convincing objection to an autogen’s “cognitive” pretensions would be that it lacks an *explicit* mechanism for selecting from environmental stimuli and “mobilizing energy” in an endogenous way. Nevertheless, this objection can be met by tinkering only slightly with the basic autogenic model.

In his initial version of the thought experiment, if an autogen breaks apart in an inhospitable environment reformation of the system is unlikely. In an alternative version, Deacon imagines a “containing capsule [which] has molecular features to which the relevant substrate molecules tend to bind” (Deacon 442). In this scenario more molecular binding on the surface of the container will tend to compromise an autogen’s stability because there are more points for potential energetic destabilization of the macromolecular structures maintaining its integrity. This means that “the process will tend to be more stable in environments lacking essential substrates and less stable in environments where they are plentiful” (442). Ultimately, “higher substrate concentrations…”—meaning a more favorable environment—“…would make disruption more probable, and subsequent use of local substrates would deplete their concentration and make the replicated autogens more stable and more likely to diffuse to new environments” (442). In other words, the addition of such simple molecular features on an autogen’s container makes it sensitive to local environmental conditions. Deacon continues, “Autogen lineages with this sensitivity to relevant substrates will effectively be selective about which environments are best to dissociate and reproduce in” (442-443). Even though these autogens are still dependent on environmental conditions to form and reproduce, such binding sites more starkly demonstrate a system-environment coupling based on the autogen's own organizational unity and autonomy—in other words, on its sense-making procedures—and not as a simple passive response to environmental conditions.[50]

Despite this passivity, the autogen supports the 4E claim that *bodily* autonomy is not just some cognitive epiphenomenon, but integral to cognition and its processes. If we grant the validity of Deacon's thought experiment, then it seems that autonomy can be reconceived on even more generalized grounds as the teleodynamic tendency towards constraint preservation, and given the constitutive role that closure/containment plays in the perpetuation of an autogen's identity, this more minimal sense of autonomy is still tied to a physical body: after all, it is physical closure—*in essence, a demarcated body*—that enables the emergent transition to a new teleodynamic order.

However, according to Deacon, it is only in the absence of closure that an autogen can truly be said to be in a non-equilibrium state (450). The fact that an autogen is fully individuated only when closed and that such individuation is marked by the cessation of *ongoing* non-equilibrium dynamics strikes me as too stark a departure from autopoiesis' and second-order cybernetics' emphasis on sense-making, and I think we must conclude that *autogens are not cognitive* in the sense that they do not engage in an *ongoing* process of self-making. Furthermore, if we combine this with the other conclusions we have drawn

---

[50] For a more detailed discussion of the representational capacities of autogens, see Deacon: "How molecules became signs." *Biosemiotics*, vol. 14, no. 3, 25 Sept. 2021, pp. 537–559, https://doi.org/10.1007/s12304-021-09453-9.

about autogenesis—that this teleodynamic shift into a non-material purpose or end can be understood in terms of constraint preservation, and that this shift can occur in the absence of what autopoietic enactivism would recognize as cognition—then that would mean that *aboutness can be shown to precede cognition*; *that aboutness recruits the cognitive processes which are elaborated in autopoietic enactivism*.

Finally, what I think this analysis of autogenesis shows is that Maturana and Varela were too conservative in their initial formulation of autopoiesis. They wanted to abstract the organization of autopoietic machines: the "relations between components and rules for their interactions and transformations…" ("Autopoiesis…" 86) a description of a living system's essential patterns which could reveal the "conditions of emergence…[that] arise as a necessary outcome whenever such conditions occur" (86). With the help of Deacon's emergent dynamics, we can see how this rather roundabout excision of purposes and functions from the organization of a living system demonstrates a failure to fully theorize the phenomenon of emergence. In another sharp divergence from Deacon, Maturana and Varela say, "The cell defines its physical boundaries through its dimension of production of constitutive relations that specify its topology. *There is no specification in the cell of what it is not*" ("Autopoiesis…" 91, emphasis added). In contrast, Deacon builds his entire theory from precisely this axiom.

Conclusion:

Many years have passed since Maturana and Varela first published *Autopoiesis and Cognition*. In that time, 4E approaches such as enactivism have become important players in contemporary cognitive scientific debates. Theorists like Thompson who work from within this autopoietic genealogy no longer cling to the stilted rejection of purposes and functions. Nevertheless, there remain aspects of Thompson's work that take the purposive features of biological systems for granted by not rigorously engaging with the thermodynamic issues surrounding emergence. These are the oversights that, in my estimation, continue to leave many mainstream cognitive science researchers "utterly mystified about what enactive approaches have to offer" (Hutto 45). Perhaps this is overstated. The problem, it seems to me, is not mystification but outright skepticism of its legitimacy because these accounts have not attempted the kind of stepwise synthesis of scientific materialist methodology with philosophical holism. With the help of Deacon's emergent dynamics, we can appreciate how the shift to teleodynamics instantiates a new causal tendency towards preservation and reproduction of constraints, features which themselves are not materially present, but which can, in principle, be rendered legible in those organizational patterns that we can observe.

In this chapter, Deacon has provided us with a lever capable of cracking open many of the holistic and causal assumptions operative in autopoiesis and second-order cybernetics. The most important of which is the recovery of teleological description within a mechanistic, and thus scientifically-valid framework. We do this by attending to what is *not there*. In contrast to systems theory orthodoxy, Deacon teaches us to see a system as *less* than the sum of its parts—as an ongoing process of dynamical interaction in which constraint and absence play the defining role.

One reason I personally gravitated towards Deacon was his early invocation of the *Tao Te Ching*, a text which I have read hundreds of times since I first came across it (and purloined) in the back of my junior-year high school history class, and which has formed the basis of my spiritual worldview. In the epigraph to chapter one, Deacon quotes chapter 11 of the *Tao*:

> Thirty spokes converge at the wheel's hub, to
>
> a hole that allows it to turn.
>
> Clay is shaped into a vessel, to enclose an emptiness
>
> that can be filled.
>
> Doors and windows are cut into walls, to provide
>
> access to their protection.
>
> Though we can only work with what is there, use
>
> comes from what is not there.[51]

Anyone who has practiced meditation will know the fullness-of-being that emerges when one begins to quiet the mind. When one starts to observe the motions of the mind dispassionately, a distance opens up between thought and one's identification with it. Paradoxically, one comes to intuit that the self we consciously construct—in all its dizzying, troubling, and multifaceted complexity—emerges out of a silent, unformed, and pervasive present-being. This is the essence of the Tao, and I would argue, the basic

---

[51] My preferred translation by Stephen Mitchell reads: "We join spokes together in a wheel, / but it is the center hole / that makes the wagon move. / We shape clay into a pot, / but it is the emptiness inside / that holds whatever we want. / We hammer wood for a house, / but it is the inner space / that makes it livable. / We work with being, / but non-being is what we use.

insight of most spiritual ways of life. The final sentence of this passage encapsulates Deacon's entire argument in *Incomplete Nature*, but the "work" he traces is not the spiritual work of connecting with the Tao, but the thermodynamic work that underlies all systemic organization and its emergent effects. Deacon says, "The core insight that guides this book can be grasped by taking the Taoist metaphor quoted at the beginning of this chapter seriously. We simply need to pay attention to the *holes*" (42). In the context of emergent dynamics, this means understanding how constraints structure the interactions of systems, and how these interactions give rise to new (emergent) patterns of organization.

I have certainly felt these absential forces at work in my own spiritual life. Chapter 48 of the *Tao Te Ching* sums up, for me, many of the practical aspects to this absential and other-oriented worldview. It also nicely underscores how this worldview manifests itself in a body of knowledge. It reads:

> In the pursuit of knowledge,
>
> every day something is added.
>
> In the practice of the Tao,
>
> every day something is dropped.
>
> Less and less do you need to force things,
>
> until finally you arrive at non-action.
>
> When nothing is done, nothing is left undone.

True mastery can be gained

by letting things go their own way.

It can't be gained by interfering.

Any reading of the *Tao* will quickly reveal the deep political and social concerns that motivate the text, and so this chapter is certainly *not* a call for torpor—for political or social disengagement. Instead, it describes a spontaneous, unforced, and dare I say *orthograde* process of mindfulness that underlies our sense of self and its many machinations. Wisdom texts and spiritual practices the world over point us towards a constitutive absence from which our sense of self emerges. They try to teach us that this unformed, empty, and silent place of non-being and non-action is, rather paradoxically, a source of immense strength, well-being, and compassion.

Incomplete Nature is a theoretical brick, but it is founded on a very simple axiom: what we observe is constituted by what we cannot. So while Deacon's work does not point us directly to the face of nature, it does teach us how to look for it in counterintuitive ways. In the following chapter, we will transition into the phenomenological concerns that constitute the pith of Thompson's enactivism in *Mind in Life*, and we will see if our expanding ethical and systems theoretical optics can allow us to see the rudiments of experience and sentience at work in basic bacterial bodies.

Chapter Four:

Enacting Sentience: A Phenomenology of Basic Minds

> Whoever wants to hold on to the conviction that all living things are only machines should abandon all hope of glimpsing their environments.
>
> —Jakob Von Uexküll, *A Foray into the Worlds of Animals and Humans*

> The normative value that comes into existence with sentience is special. No other physical process has intrinsic ethical status, because no other type of physical process can suffer. Explaining the emergence of sentience thus must at the same time be an explanation of the emergence of ethical value. This is a tall order.
>
> —Terrence Deacon, *Incomplete Nature*

In his 2009 book *Eating Animals*, Jonathan Safran Foer investigates the social, political, public health, and ethical consequences that stem from the twentieth century invention of the factory farm. For Foer, all these issues begin with the stories that we tell ourselves about the food we eat, and ultimately, the systematic forgetting we engage in, in order to consume factory-farmed food. Foer frames meat consumption as it is practiced in industrial societies like our own, in large part, as the ethical dilemma of the face. He foregrounds an all-too-common feature of our ethical lifeworld: our terrible capacity for willful ignorance. Either we never bother to look for the faces of non-human animals in the first place, or perhaps more insidiously, we experience the full appeal of a non-human face at some point in our lives, but out of convenience and ease, or perhaps just moral indolence and indifference, we turn away. We tell ourselves convenient fictions in order to deface and forget them. Factory-farmed animals confront us with an ethical responsibility of monumental proportions, and yet we have so thoroughly and unthinkingly bargained away their well-being, our own health, and the health of the planet. The contradictions here are profound.[52] Foer says: "Our situation is an odd one. Virtually all of us agree that it matters how we treat animals and the environment, and yet few of us give much thought to our most important relationship to animals and the environment" (Foer 74).

Eating Animals points to a pervasive ethical pattern that bears directly on this dissertation and on the present chapter: If we can exhibit such profound cognitive dissonance towards creatures with nervous systems, complex social structures, and manifest "intelligence," then how much more challenging and paradigm-shifting would it be if we could learn to see a face in our cognitive limit case—a bacterium. Does such a creature present us with a face? Not a face of flesh and features, surely, but the same expression of an interiority that calls us to responsibility in human faces. Is there something approaching interiority in such a simplified living system? Can we find the face in bacterial life? This is the basic question that motivates this chapter.

In the previous chapter, I argued that Deacon's theory of autogenesis and the larger methodology on which it rests—emergent dynamics—together represent one possible solution to the problems that teleology has long posed to the mechanistic and

---

[52] Foer says, "When surveyed, 96 percent of Americans say that animals deserve legal protection, 76 percent say that animal welfare is more important to them than low meat prices, and nearly two-thirds advocate passing not only laws but 'strict laws' concerning the treatment of farmed animals. You'd be hard-pressed to find any other issue on which so many people see eye to eye" (Foer 73).

mind-independent descriptions at the heart of scientific materialism. As Thompson tells us in his body-body problem, "The scientific task is to understand how the organizational dynamic and dynamic processes of a living body can become constitutive of a subjective point of view, so that there is something it is like to be that body" (Thompson 237). Deacon's theory did not take us quite that far. However, it did demonstrate, in a scientifically valid procedure, a general causal shift into purposive, end-directed organization via a discrete class of autonomous system (teleodynamics), and in ways that complement and echo the picture of autonomy developed throughout *Mind in Life*.[53] All this is good news for Thompson because it goes a long way towards demystifying the concept of "autonomy" on which so much of autopoietic enactivism depends.

In that same passage from the body-body problem, Thompson identifies the unique role that enactivism plays in the interdisciplinary project of unifying mind and life that we have been tracing. He says, "For the enactive approach, this task takes the form of trying to understand a lived body as a special kind of autonomous system, one whose sense-making brings forth, enacts, or constitutes a phenomenal world" (Thompson 237). Therefore, in this chapter, I will continue my explication of autopoiesis and cognition by looking to the sense-making processes at work in our paradigm case—bacterial life; and

---

[53] In the chapter "Emergence and Autonomy" Thompson distinguishes between a top-down and a bottom-up analysis of autonomy (44). The top-down approach refers to "the relational organization proper to autonomy" meaning systems that are "sources of their own activity, specifying their own domains of interaction", this can be everything from single cells, microbial communities, nervous systems, multicellular organisms, ecosystems etc. (Thompson 46). In other words, to be autonomous in this top-down sense "a system does not have to be autopoietic in the strict sense (a self-producing bounded molecular system)" (44). In contrast bottom-up approaches "try to work out the energetic and thermodynamic requirements for the instantiation of 'basic autonomy' in the physical world" (44). As stated in chapter two, part of the problem for autopoiesis is that it does not reach down into the thermodynamic fundament and explain autonomy in this bottom-up fashion.

show how even in this simplest of cognitive beings, we encounter the rudiments of a phenomenal world, that is, a world of individual experience which I call *sentience*.

This chapter begins with an examination of the epistemological underpinnings of autopoiesis, which holds that all knowledge is the result of system-environment coupling. I show how the property of autonomy described by Maturana and Varela inaugurates not just ontological emergence (i.e., a new class of "cognitive"/teleodynamic system), but epistemological emergence, that is, the transcendental conditions for all knowledge and knowing. In keeping with this position, enactivism argues that there is no such thing as "objective" knowledge, disinterestedly awaiting apprehension. All knowledge is enacted via the self-making and world-creating processes of cognitive systems.

Next, I expound on the origins of phenomenology and use this as a jumping off point for Thompson's analysis of bacterial chemotaxis (the most basic "cognitive" or sense-making process evinced by single cells). This section will provide us with the concrete means of understanding bacteria as sentient beings. We will see how Thompson—much like Deacon—is dissatisfied with the basic cybernetic picture of cognition articulated throughout *Autopoiesis and Cognition* because it brackets from consideration the teleological dimensions of life and mind. But whereas Deacon's answer is to develop a new account of causality, Thompson's answer is to rely on the phenomenological method as a way of reading these features into the basic autonomous and cognitive dynamics of autopoiesis. I follow this section with a short case study of its foil—Dennis Bray's 2009 book *Wetware*—which begins with an explicit rejection of the sentient features of bacterial life. I offer up Bray as a kind of cautionary tale, a way to show what can be lost from view when we cling too closely to materialist methods and/or fail to understand life phenomenologically. I believe this foray into Bray underscores the strengths and affordances of Thompson's approach.

Finally, using insights from Deacon, I end the chapter with an exploration of *representation* and to what extent it can be said to exist in simple cognitive beings like bacteria. Deacon's emergent dynamics recovered for us a general causality of *aboutness*, that is, existing in relation to not-yet-realized states of being. The goal of this final section is to translate this causal shift more explicitly into the language of "representation" and see how, and to what extent, this property can be said to manifest in a bacterium. This line of inquiry is in no way at odds with the enactivist position I have advanced throughout this project: the sustained qualification that *all* representational activity emerges from the phenomenon of structural coupling and its embodied dynamics.

I began this dissertation by arguing that much of our anthropocentrism is rooted in a conception of subjectivity based on signification, meaning-making, and language. Therefore, if we can show the rudiments—not just of sentience and sense-making—but also representation at work in bacterial life, then I believe we will have come a long way towards disabusing ourselves of one of our chief anthropocentric and ethical roadblocks. Ultimately, this exploration of representation and phenomenology will set the stage for the final two chapters of the dissertation in which I will turn once again to the level of human cognition, and to the embodied dynamics that underlie language and narrative.

As with all other aspects of autopoietic enactivism, this chapter tries to work from the bottom up. If we can start to see the rudiments of meaning, sense-making, and sentience in the world of a bacterium, then how much easier will it be to see these features in all other non-human organisms with which we come into ethical contact?

Ontological Emergence Entails Epistemological Emergence:

Throughout this project we have encountered, time and again, the methodological and philosophical problems that arise when thinking about life and mind. This is what Capra and Luisi described as the perennial swing of the pendulum between mechanism and holism. In the preceding chapters we have examined a number of systems theoretical approaches that try to bridge this seemingly insuperable explanatory gap. In chapter one, we saw how neurophenomenology can be used to check the reductionism of neuroscience by utilizing subjective insights from test subjects trained in phenomenological methods. In chapter two, we saw how Maturana and Varela's theory of autopoiesis attempted to provide a mechanistic explanation of living organization by describing autopoietic "machines" and by banishing all teleological and functional descriptions. Despite the problems inherent to this move, they nevertheless asserted a profound equivalence between life and mind: that *all* life is inherently cognitive. In chapter three, we saw Deacon put a new absential face on the second law of thermodynamics, and in the process, strikes a remarkably cogent and subtle balance between these holistic and mechanistic impulses. In short, all these approaches have attempted to strike some kind of methodological balance between the pendulum's two extremes.

Now that we arrive fully at Thompson's own unique contribution to bridging the explanatory gap—*phenomenology*—we encounter what, at first glance, may seem like the most hardline position yet. To those unfamiliar with autopoiesis and its claims the idea that a bacterium possesses cognitive properties often elicits a mixture of confusion and curiosity. But what makes autopoiesis such a powerful theory is precisely this unflinching willingness to extend "cognition" down into the very biological subsoil—to single cells and their sense-making. Thompson's autopoietic enactivism extends this position *even further*, arguing that a single cell does not just evince the basic self-referential and formal characteristics of cognition outlined by autopoiesis, (i.e., operational closure, energetic openness, a physical boundary, and autonomy), but that this autonomy inaugurates the rudiments of *sentience*.

Thompson clearly distinguishes between the explicit self-apprehension that largely defines the term "consciousness" and the more basic properties contained in his use of "sentience." He says:

> …a familiar idea from mind science is that not all forms of awareness imply consciousness in the sense of subjective experience. There are a number of different concepts of consciousness, but the one most relevant here is probably sentience, the feeling of being alive and exercising effort in movement. (Thompson 161)

Here we see that sentience and consciousness exist on a continuum, and in which sentience is advanced as a kind of proto-consciousness—a larval stage for the subjective self-awareness and self-apprehension that characterize our own cognition. Terrence Deacon echoes this view. He defines sentience in terms of "a perspective, an internal and private locus, a self-reference frame with respect to which non-self is discriminated…I will refer to this core feature of conscious experience as sentience. The term sentience derives from the Latin, and literally means 'feeling'" (Deacon 486). Thus, "sentience" is meant to invoke a paradigm of irreducible affect, in which the sentient being is in possession of itself—its own unity, purposiveness, and frame of reference. Thompson does not grant single cells the faculty of consciousness, but he does grant them sentience. By grounding our analysis of bacterial chemotaxis in this paradigm, we are acknowledging a common denominator between all living organization, ourselves included.

This intuition and incorporation of our own sentience in our understanding of other cognitive beings is the crucial step in Thompson's phenomenological approach. The researcher begins by acknowledging his or her own role as observers. Thompson says, "In the present context, the theory of autopoiesis provides a naturalistic interpretation of the teleological conception of life originating in experience, but our experience of our own bodily being is a condition of possibility for our comprehension of autopoietic selfhood." (Thompson 164). In other words, it is only because of the disclosure of our own purposive, embodied modes of being and cognition that we are capable of recognizing these processes in other organism—e.g., bacteria. This is a transcendental claim because it makes a categorical observation regarding "the conditions for the possibility of knowing life" (164), and it is a phenomenological claim because it begins from intuitions regarding one's own first-personal experience. According to Thompson, concepts such as *organism*, *self-organization*, and *autopoiesis* "are not derivable from some observer-independent, non-indexical, objective, physico-chemical description" (164). The idea is that such concepts depend on an intuition that we as observers already possess as agentive, autonomous, and purposive beings. He says, "Objectivism in biology…takes the organism for granted as a ready-made object…In contrast, phenomenology traces this category back to its cognitive source, which is the lived experience of our bodily being. Objectivism refuses to take this sort of reflexive step" (Thompson 164).

A word of clarification is in order here: This project adopts the position of philosophical naturalism the idea of "objective nature as the fundamental reality that makes possible the existence of conscious minds…" (Pace Giannotta 210). This strikes me as a basic requirement for any inquiry framing itself as a work of environmental humanities. In order to forestall fundamental rhetorical and logical incompatibilities between humanistic and scientific modes of inquiry, we need to acknowledge that something like "nature" actually exists, *out there*, and that all properties of life and mind are subject to its laws.

On its surface, this would seem to pose problems for autopoietic enactivism because it is expressly opposed an objectivist view of nature "conceived of as a mind-independent ontological domain that is endowed with certain 'objective' properties that exist independently" of its relation with a cognizer (Pace Giannotta 210). As Giannotta reminds us, "The central concept of AE [autopoietic enactivism], by which it brings into question objectivist naturalism, is that of *enaction*. According to this notion, the living organism is an autonomous system that 'brings forth' its 'environment' or 'world.'" (Pace Giannotta 211). Nevertheless, I think we can preserve this constitutive claim for autopoietic enactivism and incorporate it into a philosophical naturalism of the type Dennett, Deacon, and others adopt. Giannotta puts the matter as follows:

> In fact, the claim that subject and object of cognition are essentially correlated and cannot be conceived independently one from the other is still compatible with a

> view according to which the mind is grounded in a mind-independent world that, however, is not accessible to us in its being "in itself." In this view, "object" would be the object of cognition and not a mind-independent thing in the world, and this would confine the correlationist claim to the plane of knowledge. I will refer to it as *epistemological correlationism*. However, one can also conceive of the correlation in metaphysical terms, as a claim about mind and world conceived of as ontological domains that are fundamentally correlated. I will refer to it as *metaphysical correlationism*. (Pace Giannotta 211)

Stated simply, epistemological correlationism does not preclude philosophical naturalism. Nor would Thompson and co.'s stronger metaphysical (i.e., transcendental) claim regarding the autonomous enaction of self and world preclude what Dennett called "the 'given' categories of physics" (Dennett 30). We can grant the metaphysical claim that "world"/"environment" does not exist unless something *thinks* about it, and still recognize absolutely every time something does think, that it conforms to a set of given categories (or laws) of physics and chemistry that apply to and constrain all cognition. It would be foolish to think otherwise. Henceforward, I suggest that we think about enactivism's metaphysical claims as a qualification that operates in all that we know, and all that we can say about what we know. This is precisely the version of enactivism I have been advancing from the very beginning.

This shift to thinking about knowledge as a consequence of system-environment coupling is a core feature of the second-order cybernetic systems theory developed during the 1970s by Maturana, Varela, Bateson, Von Foerster and others. Bruce Clarke says, "To shift epistemology to an explicitly recursive system/environment paradigm forces a cascade of repercussions. This cognitive regime bars any traditional form of empirical or realist representationalism, any simplistic notion of knowledge as a mechanics of linear inputs and outputs…" (Clarke, "Heinz Von Foerster's Demons" 52). By "empirical or realist representationalism," Clarke means mind-independent forms of knowledge. There is no such thing as knowledge outside of this relationship between cognitive system and environment. Clarke develops this point in *Gaian Systems* as follows: "The prior epistemological distinction between subjects and objects becomes relative to an observer's choice…the cognitive paradox of self-reference cuts the ground out from under the pretense of unmediated objectivity" (Clarke 126). In other words, "Objectivity is surpassed by participation" (Clarke 127).

The autonomy perspective described by autopoiesis and later developed by Thompson in *Mind in Life* in more explicitly *phenomenological* terms asserts that in order for a world to be objectively described, it must first be subjectively given. In all cases, this subjectivity emerges first and foremost in the "lived body." Third-personal description depends on the situated perspective of first-personal sense-making. The epistemological shift described by Clarke recapitulates in systems (cybernetic) terminology one of the basic insights of phenomenology: that scientific practice and its pretense to "objective" knowledge emerges from a more fundamental subjective participation in the world (i.e., structural coupling).

Phenomenology, when bracketed from metaphysical or transcendental considerations, acknowledges an objective world, that is, a world that exists objectively for all, but it rejects that our access to that world is objective. Existential phenomenologist William Liujpen puts the matter thus: "…the objectivity of the world should not be interpreted in an objectivist manner…Our very thinking of an objective reality always *already* involves us in that reality and precludes any 'knowledge' prior to that involvement" (Luijpen 30). For phenomenology, the sense-making acts of the system (noesis) and its sense objects (noema) are irreducibly coupled. This perspective does not have to stand at odds with objective, mind-independent modes of description such as the scientific method. It simply asserts that such descriptions depend on a more fundamental autonomous observer. Accordingly, the objective pretensions of science are quite aptly described as a *method*—a heuristic for limiting human error.

Epistemological correlationism is another way of broaching the systems theoretical concept of *structural coupling.* Maturana and Varela say, "We speak of structural coupling whenever there is a history of recurrent interactions leading to the structural congruence between two (or more) systems" (*The Tree of Knowledge* 75). "System" and "environment" (like noesis/noema) are mutually constitutive: one cannot conceive of one without the other. In the context of autopoiesis, structural coupling describes the diachronic processes by which a self-organizing system delimits and individuates itself from the environment. We know that this process of differentiation and self-making are its essential cognitive acts. Therefore, "In describing autopoietic unity [sic] as having a particular structure…the interactions (as long as they are recurrent) between unity and environment will consist of reciprocal perturbations…"(75). And crucially, from the standpoint of epistemology, "…In these interactions, the structure of the environment only triggers structural changes in the autopoietic unities (it does not specify or direct them), and vice versa for the environment" (75). If there existed a simple one-to-one correspondence between states outside and inside of the system, then we would have a simple case of Shannon information—correlated physical differences that are embodied in some substrate. Autopoietic unities do not register difference in this one-to-one fashion; they *respond* to perturbations from without, subordinating them to the maintenance of their own organization. This is the essence of autonomy ("Autopoiesis…" 80).

In other words, the autonomy of biological life inaugurates both ontological and epistemological emergence. There is ontological emergence because autopoietic entities represent a distinct class of "cognitive" (i.e., teleodynamic) system that is its own locus of causal processes: it has the power to effect and direct change. There is epistemological emergence because this causal efficacy enacts all the knowable (perceivable) features of the objects it brings into relationship with itself.

We can see the essentials of this epistemological paradigm at work in "Autopoiesis: The Organization of the Living." We know that cognition is coextensive with autopoiesis, that a cognitive domain is "The domain of all the interactions in which an autopoietic system can enter without loss of identity" (119), and that due to its operational and physical closure, we, as outside observers, cannot reproduce this cognitive domain (92). Maturana and Varela tease out the epistemological implications of this as follows: "…the actual knowledge of the organism (its conduct repertoire)…is necessarily always a reflection of ontogeny [*sic*] of the knower…Intrinsically, then, no absolute knowledge is possible, and the validation of all possible relative knowledge is attained through successful autopoiesis" (119). In other words, there is no "knowledge" outside this cognitive domain which the autopoietic entity enacts via its self-making processes. These passages from *Autopoiesis* represent an important moment in the genesis of 4E cognition more generally. By defining knowledge as a "conduct repertoire" and by invoking the "behavioral capacity of the organism" to describe this knowledge, Maturana and Varela assert the primacy of embodied cognition. Accordingly, all propositional or "representational" processes are simply more complex instantiations of this basic mode of structural coupling with the environment, in which the autopoietic unity compensates for perturbations from without. As outside observers, we see the system respond in predictable ways to environmental stimuli and apply the category "representation" to this process, when in reality these are simply complex endogenous mechanisms that have evolved to compensate for perturbations.

These observations by Maturana and Varela in "Autopoiesis…" also recapitulate the autonomy/heteronomy distinction introduced in chapter two. Heteronomy describes the set of observations that can be made from without, in which an observer can objectively correlate changes inside and outside the autopoietic unity, but this will always be radically other from the autonomy perspective enacted by the cognitive system itself. The system can only compensate for outside perturbations (and so maintain its autopoiesis) but it cannot step outside this structural coupling and observe itself objectively from without. In his essay "Autopoiesis and a Biology of Intentionality" Varela states the matter as follows:

> I suppose others would prefer to introduce the word "information" instead. Well, there are reasons why I believe this even more problematic. Although it is clear that we describe an X that perturbs from the organism's exteriority, X is not information. In fact, for the organism only is a *that*, a *some*thing, a basic stuff to in-form from its own perspective. In physical terms there is stuff, but it is for nobody. Once there is body—even in this minimal form—it becomes in-formed for a self…Such in-formation is never a phantom signification or information bits, waiting to be harvested by a system. (8)

Whereas an outside observer can pinpoint a specific feature of the environment and correlate this feature with a specific change in the autopoietic unity, for the autopoietic system no such one-to-one mapping is possible. This is precisely what Clarke means when he says that "objectivity is surpassed by participation" (Clarke 127). It also explains why covariant or Shannon information is at odds with the autonomy perspective. This is because "The objectivist notion of information presupposes a heteronomy perspective in which an observer or designer stands outside the system and states what is to count as information" (52). In contrast, "Information looks different from an autonomy perspective. Here the system, on the basis of its operationally closed dynamics and mode of structural coupling with the environment, helps determine what information is or can be" (Thompson 52).[54]

[54] These statements go a long way towards explaining why Thompson needed to divide the issue into separate scientific/enactivists tasks in the first place. If autonomous organization inherently sequesters its own situated or "observational" perspective (Deacon 6), then how can science hope to articulate it in objective, third-personal terms? This is precisely how Deacon intervened in the problematic: to track the

In chapter two we encountered Daniel Dennett's critique of Thompson, and what he viewed as a facile distinction between heteronomy and autonomy. The counterexample he used was an AI that can be designed to "muck about in the world and devise their own categories" (Dennett 30), adding that, "As I and others have argued, all meaning in organisms is constructed by self-regarding processes that gerrymander the 'given' categories of physics to suit their purposes" (Dennett 30). This is essentially what Deacon's emergent dynamics laid out for us in precise thermodynamic detail. Cognitive systems are subject to the "heteronomous" regime of physics and chemistry and the ways in which these given categories constrain the emergent dynamics of autonomy and evolution. But I think we are now in a better position to see how Dennett overlooks the full phenomenological implications that follow from the emergence into autonomy. For example, when he says that "as an observer, one can in principle identify the categories adopted/constructed by an organism objectively — such as the categories of the sweet and the cute, the sexy and the ugly — once one understands in detail the evolutionary/ecological predicament of the organism in question" (Dennett 30). If we substitute "adopted categories" with the notion of "preserved constraints" then we could say this is precisely what Deacon's theory enables us to do, *nevertheless*, Thompson's point is that we cannot occupy another's cognitive perspective. It is operationally closed, sequestered, and privileged.

---

creation, propagation, and preservation of order/pattern in an autonomous system without arbitrarily or unknowingly applying one's own subjective standards.

Despite the fact that Maturana and Varela share the epistemological implications that follow from a recursive system/environment coupling, their use of the term "phenomenology" is nontypical. They say, "…[a phenomenological domain] is defined by the properties of the unity or unities that constitute it…Thus, whenever a unity is defined, or a class or classes of unities are established which can undergo transformation or interactions, a phenomenological domain is defined" (*Autopoiesis* 116). Throughout *Autopoiesis* "phenomenology" is not meant to invoke the interiority or subjective features of cognition but all interactive processes that couple biological systems to their environments. As with their use of the machine metaphor, I would attribute this nontypical, non-philosophical use of "phenomenology" to their mechanistic pretensions. In their desire to present a scientifically valid theory that avoids the charge of vitalism, Maturana and Varela bracket from consideration the philosophical problems posed by phenomenology.

In contrast, Thompson invokes phenomenology in its more traditional, Husserlian sense, as a way to signal the rudiments of interiority and experience endemic to the processes of cognition: in a word, *sentience*. Therefore, I want to turn from these rather abstract, epistemological considerations and start to unpack phenomenology as living practice. For Thompson, the observer's practical exercise of phenomenology is the ultimate point of departure for understanding the sentience that underlies basic cognition, and which Maturana and Varela left undeveloped (or else couched in the calculating and impersonal language of systems theory and cybernetics). By using phenomenology, Thompson makes a starting place of his own mental and bodily affordances as the means by which to see these features in the rest of living organization.

Phenomenology:

The process of "bringing forth" a world is where enactivism and phenomenology converge. To paraphrase a basic phenomenological idea: the world of experience is the precondition for the world of self-reflective understanding. Before it can be taken up by philosophical or scientific thought, the world must be given over to/appear to consciousness. Phenomenology (from the Greek *phainomenon* + *logos*) means the study of appearances, or more specifically, how things appear in consciousness.

Modern phenomenological practice originated with Edmund Husserl who argued that the defining feature of consciousness is its *intentionality*. It is directed beyond itself—correlated with that which is other than itself. Luijpen puts the matter thus: "[consciousness is present] to a reality which this consciousness itself is not. This presence and accessibility to a reality is not seen as a subsequent elaboration of a pre-existing, originally isolated consciousness; consciousness is this accessibility to reality" (Luijpen 27). In other words, consciousness is not a reified *thing*, but a structure of openness and accessibility to the real. We are always *conscious of* something, and phenomenology explores the nature of this connection—of how phenomena are given in experience.

Husserl, for example, in his efforts to move beyond stark subject/object divisions in philosophical thinking, articulated a cognitive continuum with two poles: noema and noesis. Whereas the Cartesian cogito avers an ontological distinction between the *res cogitans* and the *res extensa*—between an immaterial thinking thing and a material extended thing—the noesis/noema distinction is a way of differentiating between the intentional interpretive acts of consciousness (noesis), or the "I-pole" of consciousness, and the objects of sense towards which these acts are directed (noema), or the "object-pole." This schema does not simply reproduce the subject/object division in a more artful form because the noema is not the object, but rather how the object *appears* to the percipient through its acts of perception (noesis).

As part of this project, Husserl developed a meditative practice of bracketing, called the epoché, in which one strips away one's preconceptions in order to attend to the structure of consciousness and its manifold expressions. By situating consciousness and its world as poles on a continuum, Husserl tried to show that there was something more fundamental at work in experience prior to its objectification, namely, the lifeworld (*Lebenswelt*) and its horizons. Before the world becomes concretized, categorized, and thematized as object, there is the world of lived experience. This is the position that underlies Husserl's maxim "back to the things themselves" which, "formulates in a nutshell this intention to return to the world of original experience and the wealth of meanings which can be found there. This return must be made without any prior theorizing about experience and without *a priori* elimination of any realms of meaning whatsoever" (Luijpen 15).

Dermot Moran begins his survey of phenomenology by distinguishing it as "a practice rather than a system" (Moran 4). Rather than building a system of concepts and beliefs, phenomenology can also be understood as a doing—a *mode* of theorizing and engaging with the world which cannot be reduced to abstract, disinterested *theoria*. The two most important phenomenologists of the early twentieth century were Edmund Husserl and Martin Heidegger. Moran says, “Both Husserl and Heidegger rejected the traditional representationalist account on knowledge, the Lockean way of ideas, which explained knowledge in terms of an inner mental representation or copy of what exists outside the mind” (Moran 5). In this way, we can see how phenomenology prefigures the embodied dynamicist critique of cognitivism, namely, that cognition/consciousness is not a process of representing a fixed outside domain to a sequestered inner world of thought, but a process that continually brings forth (enacts) self and world.

In chapter one we saw how phenomenological and embodied methodologies attempt to confront dualistic and reductionist tendencies in cognitive science. By placing self/world, system/environment in the context of ongoing dynamical organization, enactivism tries to move beyond many of the myopic binaries and constraining category distinctions that tend to arrest our thinking about cognition in the totalizing gaze of third personal description—for example, the sensation/perception, event/experience, optical/biological binaries we encountered in chapter one. In chapter six, we will return to the realm of cognitive science via neurophenomenology and the methodological amendments it proposes. But the challenge we confront in this chapter involves extending phenomenology in the opposite direction, down into the domain of the simple cognitive systems, bacteria.

I am no practicing phenomenologist: I have not tried to produce the kind of explorational, first-person descriptions of time, place, space, movement, objects etc. that one finds in the likes of Heidegger, Levinas, or even Alva Noë. But I believe that the foregoing chapters *do* function as a sustained meditation on the nature of "cognition," a process of exfoliating many of the presuppositions that come along with notions such as consciousness, subjectivity, and mind. In other words, it has been a process of returning to "the thing itself" albeit by more formal, expository means. Therefore, I think we have already done much of the meditative bracketing asked of us by phenomenology and are now in a position to best understand the cognitive and sentient features of bacterial life, and in doing so, insert yet one more dagger in the anthropocentric prejudices that have elevated human cognition to the top of the evolutionary heap.

The Bacterial Lifeworld and its Sense-Making:

In this section we will examine the lifeworld of bacteria and autopoietic enactivism's phenomenological understanding of a bacterium's sense-making processes. What distinguishes Thompson's enactivism from its strictly autopoietic underpinnings is his willingness to read meaning and norms into these processes.

To begin, we can state that science already has a well-defined name for what I have thus far referred to as "sense-making": c*hemotaxis*. Chemotaxis denotes movement towards chemical concentrations. In the case of bacteria, flagella—tiny rotating hairs on the cell membrane—will spin together in rapid cycles until they detect a change in chemical concentrations, at which point a single flagellum will reverse direction, tumble the bacteria about, and initiate a new course (Bray 6). The most well-studied example is

E. coli, which can detect roughly fifty distinct chemicals—attractants like sugars and amino acids, and repellents like acids and heavy metals (6). Here then we seem to have an unambiguous, empirically vetted example of a single-celled microorganism engaging in sense-making behaviors. Thompson's intervention is to argue that "Such conduct [sense-making] is oriented toward and subject to signification and valence. Signification and valence do not preexist 'out there' but are enacted or constituted by the living being" (Thompson, "Life and Mind" 83).

In my introduction to this chapter, I cited a nearly identical passage from *Mind in Life* (2007) in which Thompson uses the word "significance" and valence. "Significance" suggests a more generic, and thus cautious/hedged, meaning: "**2.** The quality of being worthy of attention; importance, consequence" (OED, "significance"), whereas in the 2009 essay quoted above "signification" is used in its place, a change in terminology which possesses a distinctly *interpretive* connotation and which confidently ratchets up the phenomenological rhetoric. Here are just a few representative definitions from its OED entry: "**1.a.** That which is signified by something; meaning, import, implication."; "**2.a.** The fact or property of having or expressing meaning."; "**2.b.** *Linguistics* and *Semiotics*. The process of signifying; the production of signs."; "**4.a.** An indication of an idea, emotion, etc.; an intimation of a future event." Given the scope of these entries, signification can encompass everything from the indication of basic intuitions and emotions to the expression or representation of ideas and concepts. I will pursue the notion of "signification" for the remainder of this section, as I think it furnishes us with a more challenging and suggestive framework by which to understand chemotaxis as sense-making.

We know that enactivism makes strong claims *against* representationalism, understood as an isomorphic mapping between inside and outside states. The reason for this is that such a description only makes sense from a heteronomy perspective, as one standing outside the autopoietic system and observing it in some encompassing context. The autonomy perspective entailed by autopoiesis means that there are no objective features of the environment, only the endogenous and nonlinear activities that compensate for perturbations from without (Thompson 53). Therefore, "signification" for Thompson cannot mean "representation" in this isomorphic (cognitivist) sense of the word. "Indication" seems the best way to code Thompson's use of "signification" because it preserves the notion of structural coupling between autopoietic system and environment, but without interposing between the two some kind of objective, mediating process. Therefore, we might simply restate "signification" as "to indicate." When signification is read in concert with "valence" we get a still more comprehensive picture of what Thompson intends. Valence is a technical term from chemistry that denotes the capacity of an atom to combine with other atoms, but in the context of the humanities—specifically psychology—it means, "Emotional force or significance, *spec.* the feeling of attraction or repulsion with which an individual invests an object or event" (OED, "Valence"). Thus, we can understand "valence" as a somewhat more jargony way of describing the constitutive norms enacted by the organism.

We can clearly see signification (indication) and valence (norms) at work in Bray's brief gloss of chemotaxis. The bacterium indicates to itself the presence of environmental features i.e., sugars or heavy metals, and in response it acts with either attraction or repulsion depending on whether those features are either helpful (and thus will maintain autopoiesis) or are harmful (and thus will disrupt it). Taken together, *signification and valence* in Thompson's account amounts to behavioral significance. The meaning that gets enacted in an organism's sense-making is behavioral and relational, not conceptual or representational. Thompson says:

> That sucrose is a nutrient isn't intrinsic to the structure of the sucrose molecule; it's a relational feature, linked to the bacterium's metabolism. Sucrose has significance and value as food, but only in the milieu that the organism itself brings into existence. Francisco [Varela] summarized this idea by saying that thanks to the organism's autonomy, its world or niche has a 'surplus of significance' compared with the physicochemical environment. (Thompson, "Life and Mind" 82)

In enaction, the sugar that is responsible for the maintenance of the bacterium's metabolic processes becomes something other than a simple sugar molecule. It becomes a reflection of the needs of the bacterium—it becomes a *nutrient*. This passage demonstrates a clear break from a materialist paradigm because it explains its object in non-intrisinc (i.e., relational) terms. Scientifically speaking, we encounter problems when we conceive of this relationship as a "surplus" because it implies something vitalistic or epiphenomenal. "Surplus" erroneously suggests that the autonomy of the system somehow *adds* something to this system-environment coupling, as though autonomy waves a magic wand and creates meaning *ex nihilo*. Despite the fact that this passage by Thompson fails as a materialist description, it succeeds admirably as a phenomenological one. The sense-making of an autopoietic unity is always already meaningful, significative, (or ententional) in nature because "every sensorimotor interaction and every discernible feature of the environment embodies or reflects the bacterial perspective" (Thompson, "Life and Mind" 82). Furthermore, Deacon's emergent dynamics and his explication of constraint gives us a way of incorporating this relational (non-intrinsic) description back into a materialist paradigm: this relationship between bacterial system and its environment can be understood as one of preserved, constitutive absences.

This relationship (i.e., sugar-as-nutrient) enacted by the bacterium also moves us firmly beyond the bounds of a heteronomous description of Shannon (weak) information and closer to semantic (content) information[55] because this, "…information is context-dependent and agent relative; it belongs to the coupling of a system and its environment. What counts as information is determined by the history, structure, and needs of the system acting in its environment" (Thompson 52).[56] Here the qualifiers "context-dependent" and "agent relative" recapitulate the epistemological shift outlined by Clarke:

[55] I.e., "the kind of contentful information—the message—that some communications convey" (Hutto 55).

[56] To clarify, this passage occurs during Thompson's exposition of DST, and thus the notion of "information" he impugns here is the genocentric/neo-Darwinist version in which genes are assumed to carry context-independent information about the organism. Nevertheless, I think this description is apposite to the informational features of the environment encountered in a bacterium's sense-making activities, and thus illustrative for analyzing its phenomenology.

there is no such thing as disinterested informational content awaiting apprehension by a cognitive system.

A bacterium has a "history" because not only has it sustained itself over time, recreating itself from moment to moment, and locus to locus, but its sense-making requires that it keep a record of its past interactions. Dennis Bray calls this the bacterium's "short term memory," which he is keen to point out is a:

> …colloquial, nonspecialist way, referring to how a swimming bacterium carries with it an impression of selected features of its surroundings encountered in the past few seconds. This continually updated record is crucial for chemotaxis, because without it the bacterium would not be able to tell whether it was moving towards or away from a more favorable environment. (Bray 7)

Putting aside Bray's interesting recourse to the language of mentality (however hedged it may be), what we have here is one possible expression of "history." Without some way of maintaining an impression of past events, the bacterium would have no way of interpreting whether it was swimming towards or away from favorable or deleterious conditions. As Bray tells us, "This trace or record corresponds to what is termed adaptation in behavioral experiments, and it represents a sort of knowledge acquired by a cell" (Bray 8).

But it also strikes me that *history* and *structure* can be explicated in an interlocking fashion from an evolutionary perspective. The "structure" of the autopoietic system refers to the physical components utilized in the organizational pattern, in effect, the physical embodiment of the system. Structure assumes two further aspects which describe an organism's developmental trajectory. The first is *ontogeny* which is the structural development of an individual organism—"an integrated process of development towards an adult state, through which certain structures are attained that allow the organism to perform certain functions…" ("Autopoiesis…" 85). The second is *phylogeny* which is "the history of adaptive transformations" when viewed through the reproductive processes of the species (85). Undoubtedly, this downward pressure from the genetic inheritances of phylogeny onto the ontogenetic processes of development and self-maintenance plays an important role in determining certain biological functions and setting up constitutive norms. The fact that an individual bacterium has inherited some genetic variations which enhance its ability to sense chemicals in its environment is an example of an evolutionary process influencing an individual's sense-making processes.[57] Despite this downward pressure from species to individual, we need to recognize that these evolutionary constraints still ultimately depend on the structural coupling of an autonomous system.

Finally, a cognitive system demonstrates "needs," because in order to sustain itself it must constantly integrate energy from outside its boundary. The system is sustained by ongoing self-production which is predicated on energetic consumption. In effect, such systems are in a perpetual state of privation and must *act* accordingly. It is this dimension

[57] We encountered an analogous example in chapter one: face pareidolia in humans. As social creatures whose evolutionary fitness depends on our capacity to perceive and interpret facial nuances, our sensorium is particularly susceptible to seeing faces in inanimate objects. While this is an evolutionary inheritance of the species, it can also directly influence an individual's phenomenology

of need—privation—lack—that lays the foundation for self-making. From the perspective of the bacterium the sugar molecule is not an objective feature of the environment but its defining *raison d'être*—it is a *nutrient* and thus imbued with all the valence and meaning therein.

Deacon vs. Thompson: Two Sides of the Same Coin

We can see in the foregoing passages how sense-making—an autopoietic unity's essential cognitive process—is intimately connected with the notion of making sense, i.e., making *meaning* from its environment. Thompson's use of the term "signification"—understood as both *indication*, in which a cognitive agent evinces a non-intrinsic and yet constitutive relationship with something other than itself, and *valence*, in which a cognitive agent acts according to norms that advance or deter its integrity as a unified being—both evince the general property of ententionaity or *aboutness* which Deacon explicated for us in the previous chapter. This meaning enacted by a bacterium is foremost behavioral and embodied. I also firmly believe that this phenomenological analysis of bacterial chemotaxis represents a clear advance over the machinic metaphors and cybernetic descriptions that predominated in the original theory of autopoiesis.

According to Maturana and Varela's rather rigid cybernetic paradigm, "cognition" denotes the self-referential and homeostatic processes that conserve the identity of the system through energetic and material flux. For Deacon, these first and second-order cybernetic descriptions failed to fully account for the ententional features of life and mind. Thompson, for his part, refers to these bare-bones cybernetic mechanisms as "minimal autopoiesis" (Thompson 147). He argues that minimal autopoiesis can describe the conservation of the system's identity (organization), but that it is unable or unwilling to broach the issue of normativity or need (147). Thompson (following the development of Varela's own thinking with regards to normativity and purposiveness) declares that "minimal autopoiesis is necessary but not sufficient for sense-making (or cognition)" (Thompson 149). Throughout *Mind in Life* Thompson expands autopoietic enactivism beyond these basic cybernetic descriptions of cognition so that it entails "behavior or conduct in relation to meaning and norms that the system itself enacts or brings forth on the basis of its autonomy" (Thompson 159).[58]

According to Thompson, the sense-making enacted via the system's autonomy, "changes the physico-chemical world into an environment of significance and valence, creating an *Umwelt* for the system. Sense-making…is none other than intentionality in its minimal and original biological form" (Thompson 147). Thus, sense-making is to enactivism what "ententionality" is to emergent dynamics. It is meant to capture the intentional, purposive, and end-directed features of life and mind. We can now clearly see that Thompson, in fact, shares Deacon's critique of autopoiesis. Both agree that it failed to fully account for the purposive (and normative) dimensions at work in self-referential, autonomous systems. We can also see how Deacon's and Thompson's theories truly are two sides of the same coin. Whereas Deacon's emergent dynamics is able to demonstrate a more general *causal* shift into purposive, end-directed organization via the language of

[58] See Weber, Andreas, and Francisco J. Varela. "Life After Kant: Natural Purposes and the Autopoietic Foundations of Biological Individuality." *Phenomenology and the Cognitive Sciences*, vol. 1, no. 2, 2002, pp. 97–125, https://doi.org/10.1023/a:1020368120174.

thermodynamics and dynamical systems theory; Thompson's enactivism pursues this same emergent shift via a phenomenological idiom of significance, valence, normativity, and need.

Thompson defines sentience as "a feeling of being alive" (161). I think the foregoing analysis clearly demonstrates how bacteria—the most basic of cognitive beings—behave in ways that correspond to this definition, a conclusion that is made all the more forceful and convincing if we take that phenomenological first step that asks us to locate those same dynamics in ourselves. What Thompson approaches through the lens of phenomenology, Deacon situates in the language of emergent dynamics. He says:

> Because organisms are teleodynamic systems, they do not merely react mechanically and thermodynamically to perturbation, but generally are organized to initiate a change in their internal dynamics to actively compensate for extrinsic modifications or internal deficits. Feeling is in this most basic sense active, not passive. (Deacon 487)

This passage perfectly highlights the shared vision of life and mind that exists between Deacon and Thompson. In this passage Deacon describes the autopoietic notion of *structural coupling* and Thompson's view of sentience in an interlocking and immutable fashion, simply refined into his own conceptual language of emergent dynamics. Organisms do not "merely react" to their environments as though they were complex mechanisms. They *respond* to their environment by initiating "a change in their internal dynamics." They do not represent this environment in a one-to-one mapping or mechanistic fashion (as the heteronomy perspective or Shannon-information paradigm would have us think). Instead, they "compensate for extrinsic modifications." As Varela tells us in "Autopoiesis and a Biology of Intentionality," for the organism there is only "a basic stuff to in-form from its own perspective" (8).

In the following section, I will turn to a foil of Thompson's phenomenological position as a way to both measure its credibility and demonstrate what is lost when we cleave unnecessarily to the strictures of a materialist enterprise.

***Wetware*: Are Single Cells Sentient or Biological Robots?**

In his 2009 book *Wetware*, molecular biologist and neuroscientist Dennis Bray makes the case for why cells, and more specifically the protein networks that constitute a cell's metabolism, perform computations and logical operations. The opening statement of the book reads, "An individual cell, in my view, is a system that possesses the basic ingredients of life but lacks sentience. It is a robot made of biological materials" (ix). The basis for this claim is the observation that individual cells need to be able to process information and "be capable of some sort of logical analysis" in order to survive (Bray 64). Bray's primary point of departure is the information-processing evinced in the feedback and control mechanisms that enable bacterial chemotaxis. He says, "The molecular mechanism of *E. coli* chemotaxis is a superb illustration of cellular information processing…" (6). My contention in this section is that his computational approach to cellular dynamics colors his entire biological outlook, and it places him in stark contrast to the perspectives outlined by Thompson and Deacon.

Therefore, we have a common object of analysis—bacterial chemotaxis—but one that will be explicated according to an entirely different paradigm. In contrast to Thompson's phenomenological approach (as well as Deacon's counterintuitive dynamical systems approach) Bray gives us a view of cellular chemotaxis from the perspective neuroscientific, evolutionary, and molecular biological "main street." As such, I offer up Bray as a kind of cautionary tale. He strikes me as a worthy representative of the current state of science and its materialist approach to understanding of life and mind, which is interested in interrogating its extant explanatory gaps like never before, but which is not always comfortable or confident in asserting their equivalence.

Over the course of the book Bray adduces a number of distinct biochemical processes inside cells that demonstrate computational and logical operations, but they are all ultimately grounded in the *allosteric properties of proteins*: the fact that they can "recognize more than one molecular partner—that is, that they have two binding sites" (61). To explain the importance of this property and how it relates to the process of cellular computation, Bray asks us to imagine ourselves as an enzyme—a biological catalyst that speeds up the rate of a specific chemical reaction in a cell. Bray says, "From the standpoint of cell chemistry you are a conduit, a one-way pipe, taking up molecules of A and spitting out molecules of A^1" (66). Because of allostery, "You, the enzyme, can adopt two different shapes" (66). One is the *on* state in which A becomes A^1. The other is the *off* state, in which no catalysis is occurring. These two states are controlled by "the binding of a second molecule, call it B. If B is absent, you are off, inactive: the pipeline conversion to A^1 cannot proceed. But if you meet and bind a B molecule, you flip to the on state and the reaction can occur. In terms of logical operations you are performing the following: IF (A) and (B) then (A^1)" (Bray 66). He continues, "In biochemical terms, substance B causes a concerted change in your structure. This is allostery in action: a regulatory molecule binds to an enzyme…and causes a change in shape (switch to an active form)" (67). Bray likens the allostery of proteins to the transistors that make up an integrated circuit: "Each can perform a logical operation in which it, in effect, reads in the concentration of one small molecule and reads out another" (67).

The allostery of catalytic enzymes in the cell membrane are what account for the *E. coli* bacterium's capacity to sense helpful and harmful chemicals and initiate chemotaxis. Whereas animals have nervous systems with networked nerve cells and a brain, a single-celled bacterium has the allosteric properties of its proteins. But unlike the silicon "hardware" of computer chips, in which transistors turn an electrical current on and off in a mechanical fashion, cellular computation happens via thermal diffusion (Bray 56). In basic cells there are already thousands of molecules—sugars, amino acids, lipids, nucleotides—that are modified in this allosteric fashion (68), all of which are "linked functionally because they share a common pool of molecules" (69). In the cytoplasm of the cell molecules are "brought together…by the unending riot of thermal motion" (58) and undergo chemical reactions, not mechanical ones.[59] In sum, the allosteric properties of

[59] Although as Bray tells us, "The distinction between chemistry and mechanics is a human invention and not one that concerns a cell. At the atomic level, all movements entail a chemical change and all chemical changes create movements. The difference is one of degree rather than kind" (93).

proteins constitute the essential "wetware" of cellular organization. It is the mechanism of cellular computation and its information-processing.

Bray argues that the information-processing of these protein networks is representational (126). The key to this claim comes in chapter six where he makes the case for why protein networks are analogous to artificial neural networks. Reflecting on neural networks he says, "Once a network has become trained, its weight and activities acquire particular values that make the network function correctly. You might say that the network has then 'captured a representation of input patterns,' or more provocatively 'acquired knowledge of its environment'" (113). Here we see the more generic notion of representation as mediating process that we encountered in chapter one via Cain. The patterns of activation at the input layer of the network influence the patterns of activation at the output layer, and in a neural net "there is a progressive integration of perceptually significant features as input segues into output" (115). In other words, it maps the outside world, but it does so in a *particular* way because the training towards an ideal output means that the information being processed is "perceptually significant." This qualifier of *perceptual significance* serves as the truth-conditional or normative feature that, on its surface, would elevate this process from basic Shannon information to something approaching semantic representation in which there exists a meaningful relationship with the environment based on history, structure, and needs. It suggests a view of the bacterium understood as an individual, and not, as his opening remarks suggests, as a mere biological robot.

Throughout chapter six, Bray draws minute parallels between protein networks and neural networks—the details of which are not necessary here—only his conclusions. He says, "In general, you can say that signaling proteins, working individually or as a small group, resemble the hidden units of a neural network. That is, they have a semantic content, or meaning" (126). Here again the implication is that semantic content hinges on the way in which information gets coded according to the functions (i.e., the "output") of the protein or ensemble of proteins. In an even more interesting moment of conceptual overlap Bray compares this process of protein network "representation" to the process of visual perception. Just as vision breaks down features of the visual scene and recombines it as an electrical signal that can be processed by the brain, Bray says, "Something like this happens in a single cell. The raw stimuli it experiences are also fragmented, albeit into chemical rather than electrical signals…I discussed in Chapter 6 the semantic content of proteins: how they "mean" certain things to the cell" (239).

In his explication of protein networks, Bray appeals to the endogenous demands of the cell's metabolism and to biological functions (e.g., chemotaxis) which serve as the normative, or truth-conditional features that elevate these otherwise straightforward Shannon informational relations to the status of semantic information, i.e., information which *means* something to the *individual* cell, what Thompson has already explicated for us in terms of the interlocking notions of history, structure and needs. Bray's thesis is that the protein networks in cells are an example of information-processing, and we can see that in order to flesh out this thesis he makes exhaustive use of neuroscientific findings and its associated mentalistic language.

For example, in the previous section I quoted Bray and how he likens the bacterium's ability to persist towards favorable conditions as a kind of "short-term memory" (7). They have the capacity to maintain a "trace or record" of their immediate environment—an essential capacity in the context of chemotaxis because it signals to the bacterium whether it is swimming towards or away from helpful or harmful conditions. Bray says that this "represents a sort of knowledge acquired by a cell" (8). But at this point he makes a categorical distinction between "knowledge" and "learning." The former is the result of evolutionary inheritances that have been hard-wired into the DNA of bacteria. According to this view, *E. coli* chemotaxis is nothing more than hardwired biological programming. In contrast, "The storage of memories by higher animals…[have] Complex mechanisms [that] allow the strengths of synapses to change with experience. They are no longer specified solely by the DNA but also influenced by recent events. This is 'learning'" (Bray 9). In other words, bacterial memory is "ritualistic" "predictable" and "stereotypical" (Bray 8-9), in the words of Deacon, it merely *reacts* to its environment, whereas kingdom Animalia, thanks to its networked nerve cells and the plasticity of its synapses, *responds* to its environment in a categorically different way.

Yet at the same time we have seen how much Bray relies on the model of a neural network and its endogenous generation of norms in order to explicate the information-processing at work in allostery. In the above passages we are told that *E. coli* behavior and its evolutionary "knowledge" is categorically different from sentient cognition and its "learning," and yet throughout the rest of the book he will repeatedly invoke neuroscience and neural networks to define cellular information-processing mechanisms.

The inconsistencies I am pointing out here are meant to signal a methodological ambivalence and anxiety that is not present in Thompson or Deacon. In the opening statement of the book Bray calls individual cells biological robots and yet on the very next page he declares that “Living cells have an unlimited capacity to detect and respond to their surroundings…Every cell in your body carries with it an abstraction of its local surroundings in constellations of atoms. A basic knowledge of and *response to* the environment are integral parts of every living cell’s makeup” (Bray x, emphasis added). These two statements seem strangely at odds with one another, both tonally and conceptually. How is it that cells can have knowledge without mentation? How is it they can they have life without sentience? It also sets up a pattern that will persist for the entirety of the book. In one breath he will declare biological processes the result of “blind circuitry” (24), and in the next invoke their systemic and organizational features as grounds for questioning whether it is possible that microorganisms possess, “in an extremely reduced and primitive form, some of the mechanisms that mediate the sense of the environment and self in humans?” (25).

Despite the complex information processing evinced by single cells such as bacteria, or the much more complex behavior he describes in other microorganisms such as protozoa and amoeba, Bray begins his book by declaring in an unequivocal fashion that these are computational processes devoid of sentience, and he never explicitly walks back the categorical nature of this claim. He never offers the reader a rigorous definition of sentience, an interesting omission given his assertion that “Words like memory, awareness, and information are easy to use but require careful definition to avoid misunderstanding”

(7). The closest he comes to defining sentience is on page 25 where he says, "Or might that watery slurry contain an ember of emotion, a prototype of sentience?" Throughout the book he courts the sentient and mental features of brains as well as the discourse of neuroscience in order to help explain the computational features of simple cells without ever broaching what such a transition into mentation means or where, specifically, such a transition might occur.

This elision has important consequences, one which I would argue reproduces the pattern we have seen in other avowedly materialist approaches (e.g., cognitivism). Bray declares that single-celled organisms and all the complex behaviors they evince are nothing more than the result of "blind circuitry" (24), but he is only able to differentiate these biological robots from "higher" organisms by invoking their complexity. He says, "Single cells, therefore, emphatically do not have feelings or humanlike consciousness. They are just too small and simple. They not only lack a brain with a cortex (which is surely required before any subjective experience that we would recognize is possible), but they have no nerve cells at all!" (24).[60] This passage gets to the very heart of the problem mentioned above—i.e., courting neuroscience and the cognitive features of brains and nervous systems without actually broaching where cognition, sentience, or subjectivity begin, or what those sticky terms actually *mean*. If the "learning" that goes on in nervous systems is categorically different from the "knowledge" baked into single cells by evolution,

[60] This feels a bit like a red herring. I have yet to encounter a serious contemporary theorist who claims that single cells possess consciousness in the sense of subjective self-awareness. Autopoiesis is one of the most well-known theories that deals with the cognitive capacities of single-celled life, and nowhere do Maturana and Varela ascribe to cells subjective mental states, nor, as we've seen, does Thompson.

then the nature of this difference, for Bray, seems to hinge largely on the relative scale and complexity of the former in comparison to the latter. Here I am reminded of the epigraph from chapter two, in which Maturana and Varela say that "The beauty of life is not a gift of its inaccessibility to our understanding" ("Autopoiesis…" 83). Bray distinguishes between bacterial computation and its "knowledge" on the one hand, and subjective experience and its "learning" on the other, by offering up the awesome complexity of the latter as reason enough to treat bacteria as robots.

I do not think that these problems can be resolved by a mere disambiguation of the word "sentience" for, as we have just seen, Bray also refers to sentience in identical terms to that outlined by Thompson and Deacon, as even the most rudimentary of feelings—an "ember of emotion" (Bray 25). It seems to me that Bray has adopted an interpretive fore-structure that forecloses his ability to see the sentient features of bacterial life. He spends hundreds of pages explicating cells in all their dizzying complexity, but he cannot overcome the materialism and reductionism of his discipline.

I have engaged in this close reading of Bray because I am convinced that excising even the most rudimentary forms of internal experience for bacteria (calling them robots) creates far more problems than it solves. If we explicitly define sentience into life's simplest units—bacteria—not only do we unify, implicitly, life and mind; but we also avoid the problems posed by epistemological emergence as we climb the scale of systemic complexity. Theories such as autogenesis and autopoiesis address these issues from the very start and from the ground up and the result is that they do not find themselves in the kinds of ambivalent and equivocal positions that I believe Bray sometimes does.

Thus, despite the many affordances offered by his computationalist approach, Bray is beholden to a materialist perspective that can, at times, be difficult to parse. He has a clear commitment to understanding the interrelationship between life and mind, but his many equivocal moments of materialism can obscure the very connections he is trying to articulate. His thesis depends on an analogy between the mechanical computation of an integrated circuit and the chemical information-processing of proteins and without question this is an extremely productive and thought-provoking way of expounding cellular processes (I should say that I am a *huge* fan of his book). But after hundreds of pages and innumerable argumentative hedges that try to accommodate for vast differences in complexity, scale, and chemistry, Bray provides no compelling reason to excise sentience from the dynamics of a single cell. Indeed, it is precisely the effect of all these innumerable hedges—and the differences in complexity, scale, and chemistry that they highlight—that end up convincing me of the antithesis.

Despite the (at times) categorical way in which Bray denies cells this faculty of mentation or sentience, I do think *Wetware* provides autopoietic enactivism with an interesting vision of how information-processing, computational processes, and even a hedged notion of representation can become integrated with the autonomous sense-making that defines autopoiesis. Such an addition of computational or representational processes to the dynamics of cellular life need in no way compromise the epistemological implications of structural coupling, nor the basic insight that follows from this—that *embodiment precedes representation*—so long as these "representational" or "computational" processes are subordinated to the paradigm of autonomy. In other words, we can read allostery and its logical processes as another means by which an autopoietic unity compensates for perturbations from without.

The crux of Thompson's enactivism is to read—not just basic cognition—but the rudiments of sentience into the autopoiesis of simple cells. Thompson's working definition is the "feeling of being alive," which over the course of his analysis comes to involve the history, structure, and needs of an individual cell striving for its self-maintenance, and which inaugurate the rudiments of signification and valence. Ultimately, I think this is a much more holistic way of understanding the cognitive lifeworld of simple cells.

In the following section, I want to turn to a different version of enactivism, one which eschews the phenomenological paradigm of experience, inwardness, and sentience, and focuses more exclusively on the behavioral competencies (i.e., the know-how) that defines the enactivist position.

Representational "Habits": Can Single Cells Be Said to 'Represent' Things?

I want to end this chapter by offering up one final example of how we might integrate the language of "representation"—and all the semantic, intentional, and mentalistic connotations that come with it—into the basic autonomous, sense-making dynamics described by Thompson's autopoietic enactivism. The goal of proposing such a synthesis is to lend as much credence and argumentative weight as possible to the idea that bacteria are *sentient*—that they possess an interiority that is inherently meaningful and significative.

The structural coupling of an autonomous systems places very sharp qualifications on any use of the word "representation." As I argued in the previous section, Dennis Bray provides enactivism with one vision of how single-celled "information processing" can be understood along representational lines. We can speak cogently about the "representational" and information-processing capacities of bacteria, but in doing so, we cannot lose sight of the sense-making of autonomous individuals. Autonomy needs to be the starting point for any understanding of "representation," whether in the context of bacteria or, as we will see in the following chapter, at the level of human cognition and spoken language.

Nevertheless, the language of "representation" is often treated with suspicion because of the disembodiment that the word implies. For example, in their 2013 book *Radicalizing Enactivism*, Daniel Hutto and Erik Myin call their approach "radical" because it extends the sensorimotor and behavioral aspects underpinning enactivism to their logical and argumentative extreme, maintaining an *utterly* anti-representational view of cognition.[61] Much of Hutto and Myin's book is dedicated to examining and undermining attempts to naturalize "content" understood along the lines of a discrete and reproducible concept or token manipulation that—as Thompson would say—can "rise above the dynamic fray" (Thompson 56). Hutto and Myin's approach, while sometimes falling victim to disciplinary jargon and esoteric lines of inquiry, is based on the very reasonable

[61] As Giannotta tells us, it is now common to distinguish between at least three versions of enactivism: Thompson's autopoietic enactivism, Noë and O'Regan's dynamic sensorimotor approach, and Hutto and Myin's radical enactivism, all of which "diverge in how they account for the interaction between the embodied mind and the environment" (Pace Giannotta 210).

assumption that basic minds such as bacteria cannot be said to contain or manipulate content in the same way that organisms with nervous systems do. Thus, radical enactivism hangs its argumentative hat on this equivocal notion of "content."

Accordingly, radical enactivists argue that we must jettison all representational and semantic talk in favor of the "more austere teleosemiotic talk of informationally-sensitive responses to natural signs" (Hutto 59). Like Thompson's recourse to "autonomy," this move helps us to focus on the organismic whole and avoid the perennial mechanistic temptation to explain cognition by some inherently psychological or atomized mental state.

On the one hand, I am partial to Hutto's framing because the language of "semiotics" helps extend the properties of significance and valence to all corners of the natural world, wresting it from linguistic—and by extension, anthropocentric—overdetermination. Furthermore, the appellation "teleosemiotics" and its emphasis on "natural signs" places it in a rich genealogy of ecological and embodied work, such as Jakob von Uexküll and his analysis of the "animal's perceptual life-world" or *Umwelt* (Sagan 2), as well as Charles Sanders Peirce, whose pragmatism and "habit-based theory of signs" (Legg and Black 2267) occupies an interesting middle ground between analytic philosophy and phenomenology, one which we will turn to in the following chapter. However, I am not swayed by the argument pursued throughout Hutto and Myin's book that *conceptually explicit content* (what Hutto and Myin cast in terms of "intension")[62] is

[62] In semantics, meaning is thought to contain two aspects intension/extension (Putnam 699). Intension refers to the internal content of a word, that is, its conceptual underpinnings (710). In contrast extension

the only way of understanding or naturalizing a representation, nor do I share their kneejerk animus towards the language of representation per se.[63]

In this somewhat fraught context within enactivist theory, Deacon provides us with a convincing set of reasons for why the preservation of constraints both within and across emergent levels of a dynamical system can be understood as a kind of representational content. We know from chapter three that constraint serves as a mind-independent way of expressing orderly arrangements. Constraints do not appeal to some ideal form that stands outside the physical (e.g., spirality per se)—a doctrine known as metaphysical realism (Deacon 184); nor does it regard observable regularity of forms in terms of universals, that is, as epiphenomenon of inductive reasoning— a doctrine known as nominalism (Deacon 185). Constraints—the preservation of system states *not* realized—expresses organizational pattern in a way that avoids these two pitfalls, and because constraints always arise in the context of a system's capacity to do physical work it means that these organizational forms are always immanent in matter. They are always embedded in the material interactions of dynamical systems and how they couple with their environments.

---

refers to all the things in the world that correspond to that word, and an extensional description would be a list of all members of that class. The point here is that intensionality refers to some kind of *conceptually explicit* representation: a *this* belonging to a *that*. Catherine Legg sums up Hutto and Myin's position regarding intension as follows: "content is not the sort of thing that can be physically instantiated. Why not? Because fully-fledged concepts instantiate intensionality—the ability to present the same information under different 'guises' (Hutto & Myin, 2013, pp. 79–80, 184n2)" (Legg 14754).

[63] Here I would agree with Catherine Legg when she argues that Hutto and Myin's rejection of "indication relations,"—which posits a triad: (1) an indicator, (2) an indicated object or state of affairs, and (3) an interpreter—is, from a Peircian perspective, "a pity" (Legg 14755).

Deacon says that by abandoning "descriptive notions of form" in favor of constraints "two things are gained" (Deacon 194). The first is *individuation* because "possible features *not* expressed and degrees of freedom *not* exhibited are specific individual facts" (Deacon 194). The second is *extension*. Deacon says:

> constraints can have definite extension across both space and time, and across whole ensembles of elements and interactions. But although the specific absences that constitute a constraint do not suffer the epiphenomenality of descriptive notions of organization, they are nevertheless explicitly not anything that is present. (Deacon 194-195)

Together, these passages situate Deacon's concept of constraint within the aforementioned analytic and linguistic paradigm of semantic/informational content rejected by radical enactivism. Constraint allows for individuation as well as membership of a class—in other words, both extension across myriad contexts, and intension (a reproducible content that inheres in the representation).

We can concretize this process in the context of autogenesis. The fact that autogens can sequester and reproduce their constitutive constraints across many different types of conditions, both favorable and unfavorable, suggests at least a partial "de-coupling of what is dynamically possible from immediately present dynamic probabilities" (Deacon 458). As Deacon says in his 2021 article in biosemiotics, "…by being able to re-present its own boundary conditions in new instantiations it [the autogen] intrinsically represents and reproduces its own conditions of existence" (Deacon, "How Molecules Became Signs" 547). The argument here is that constraints preserve specific features of an autogen's environmental milieu (intension), and that these features can be maintained (re-presented) across many different contexts and conditions (extension).

Calling this process of re-representation "representation," however, means that we need to acknowledge/account for one crucial problem, and that is the apparent absence of an explicit semiotic vehicle or sign at work in the process of autogenesis. In basic autogenesis we are not yet presented with "a teleodynamic process [that] can offload certain critical constraints that it must preserve onto *a separate physical substrate*, and in so doing endow this substrate with semiotic functionality—in other words, aboutness" (Deacon 454, emphasis added).[64] The paradigm case for this being genetics. Genes are physical substrates—specific sequences of DNA molecules—that preserve constraints (features of the system and the environment) by coding for the construction of proteins.

But in an article for *Biosemiotics*, Deacon walks back the categorical nature of this claim. There he says, "The point of this model system [the autogen] is to establish what can be considered the ground of interpretive competence. In this respect it is effectively a 'zeroth' level of semiotic process. As such it 'interprets' the most basic semiotic

[64] I want to point out that in this section of *Incomplete Nature*, Deacon has turned his attention to the information-based reproduction strategies of DNA and to evolution. In this context he says, "The shift from simple autogen replication to information-based reproduction…The capacity to offload, store, conserve, transmit, and manipulate information about the relationship between components in a teleodynamic system and its potential environmental contexts is the ultimate ententional revolution. It marks the beginning of semiosis as we *normally* conceive of it, and with it a vast virtual representational universe of possibilities, because it marks a fundamental decoupling of what is dynamically possible from immediately present dynamic probabilities—the point at which the merely probable becomes subordinate to representational possibility. This is the source of the explosive profligacy of biological evolution" (Deacon 458, emphasis added). While this quotation may seem to go against his conclusions in "How Molecules Became Signs" I want to flag the qualifier "normally." If information-based reproduction is how semiosis is *normally* conceived (by teleosemantics, neo-Darwinism etc.), then I would argue that simple autogenesis is semiosis abnormally conceived. Just because it is nontypical, does not mean it is illogical or invalid.

distinction; i.e. between self and non self" (Deacon 547). In other words, *the explicit sign vehicle of the autogen is its own boundary, its own body.* Autogen's instrumentalize their own boundary by making it a repository of core constraints. In effect, the creation of a boundary is also the creation of a teleodynamic system's first "sign vehicle"—hence the idea of a "zeroth" law of semiosis.[65]

Given the profound overlap between Thompson's and Deacon's approaches to life and mind, this "zeroth" law of semiosis strikes me as the knockdown argument in favor of placing "representation" back into the paradigm of autopoietic enactivism. In other words, we can use the autogen as a stand in for thinking about bacterial chemotaxis. These arguments regarding the semiotic (or re-presentational) capacity of autogens demonstrate another way of thinking about representation and its content. It also inverts Varela's statement (and Thompson's endorsement of that statement) that autonomous systems create a "surplus of significance" compared to their physico-chemical environment (Thompson, "Mind in Life" 82). Deacon shows us how this "surplus" is really an absence—a preservation of orderly habits via constraint preservation. According to Deacon's analysis, the *E. coli* bacterium maintains its pattern of organization metabolically by representing to itself its own cell membrane. The cell membrane is a sign vehicle that serves as a repository for key metabolic constraints that preserve the identity of the system and distinguish it from its environment.

---

[65] This language of a "zeroth law" is meant to invoke the zeroth law of thermodynamics (so named because it was formulated only after the first, second and third laws had become popularized and sedimented in culture and scientific practice).

Conclusion:

The defining feature of Thompson's enactivism is the recognition that "inwardness or interiority [of the autopoietic selfhood] is disclosable to us because we ourselves are living beings who experience our own bodily selfhood firsthand" (Thompson 163). Thompson's enactivism asks us to begin any investigation into cognition *phenomenologically*: in order to perceive "cognition" in others, one must first intuit its features in one's own purposive and embodied lifeworld, and use this as a way to measure the degree of interiority or inwardness in other beings.

In keeping with Maturana and Varela, Thompson understands the sense-making processes of chemotaxis as "cognition in its minimal biological form" (Thompson 159). But he takes this one step further, adding to this basic cybernetic picture a phenomenological domain characterized by valence, significance, and norms that that system itself enacts. As Thompson tells us, "Sense-making…is none other than *intentionality* in its minimal and original biological form" (Thompson 147, emphasis added), going so far as to call this sense-making a form of selfhood (162). These passages clearly demonstrate Thompson's synthesis of autopoiesis and phenomenology, one which expands the domain of "intentionality" beyond its original connection to "consciousness," and down into the basic biological fundament of single cells—what I have broached throughout this chapter under the banner of *sentience*. Thus, when we speak about "intention," "agency," or "interiority," we need not limit ourselves to organisms with nervous systems, nor attach to these ideas the highly complex cognitive processes contained in words like "consciousness" or "subjectivity." Instead, we can think in terms of sentience—an individualized capacity for meaning, sense, and significance. Sentience denotes a basic expressivity that calls us out of ourselves and asks us to respond. In other words, in sentience we can perceive the origins of ethics—the origins of a face.

Throughout this dissertation, we have seen how much of our thinking about cognition and mentality is tied up in the language of representation. It is the notion that minds map their worlds and that they do so in a particular way, that such representations *mean* something particular for that mind. In the words of Thompson, that it creates signification and valence for individual cognitive beings. Throughout this chapter I have used the word sentience to denote this inherently affective and meaningful state of being. I have tried to show a number of ways we might productively think about representation and sentience in the context of our cognitive limit case—a single celled bacterium—on the assumption that, if we can show the rudiments of representation at work in these simple creatures, then we are already a long way towards disabusing ourselves of the kneejerk anthropocentric prejudice that there are simply certain forms of living organization that do not deserve our consideration and care. According to the perspective of autopoietic enactivism and the phenomenology that informs it, excising sentience from living organization is inherently contradictory, and as we have seen through our discussion of Bray, often creates more problems than it solves.

In the following chapter, we will begin our ascent from these cellular depths and continue our discussion of phenomenology, systems theory, and embodied meaning-making in the context of continental philosophy and literary studies.

Chapter Five:

Linguistic Systems and Redefining the "Center": from Structuralism to Pragmatism

I began this project by pointing out the anthropocentrism that results when we make spoken language a defining feature of subjectivity. In the intervening chapters I have tried to make the case for "cognition" as an alternative organizing principle for understanding mentality, interiority and selfhood, one that can mediate between myriad physical, thermodynamic, informatic, biological and phenomenological processes.

This chapter will mark the beginning of our ascent back into the world of spoken language. I will examine the linguistic turn that occurred in the humanities in the early twentieth century and focus on two of its watershed texts: Ferdinand de Saussure's *Course in General Linguistics*, and Jacques Derrida's 1966 lecture "Structure, Sign, and Play." I situate structuralist linguistics and its offspring as part of a more general systems-theoretical swing of the pendulum emerging in the early twentieth century, a move away from viewing the world in terms of isolated objects and parts and towards novel understandings of interconnection and wholeness. Despite these positive systems theoretical evolutions it expressed, I will show structuralism became increasingly disconnected from the material processes in which its virtual and organizational interrelationships are embedded. Thus, I will argue (a) that systems thinking formed an explicit and essential part of the linguistic turn, (b) that the wholly virtual (i.e., organizational and referential) understanding of language at the core of structuralist linguistics laid the foundation for social constructionism and ideology to dominate theories of subject formation—to the exclusion of embodied and material processes, and (c) that enactivism allows us to reassert the autonomy of the subject on embodied grounds, prior to the social and ideological construction of the subject. I will end by examining an alternative genealogy of semiotics founded in the work of Charles Sanders Peirce, in which embodiment and autonomy play explicit and constitutive roles. Peirce's semiotics—in contrast to the synchronic and binary (signifier/signified) notion of the sign in structuralism—is constructed on a triadic and processual conception of the sign that grounds signification in the body and its case-specific contexts. It is a theory that can unite enactivism with linguistics and interpretation, and as such offers literary studies another way out of post/structuralist ideology critique.

Structuralism:

Let us start the ascent by considering once again the systems theoretical distinction we encountered in chapter two, between autonomous meaning-making and heteronomous information-processing. As we saw in the previous chapter, a key feature of the so-called "enactivist task" is to show how autonomy inaugurates an epistemological shift and to read this shift into the very foundations of biological life. We can see an analogous epistemological shift in the continental philosophical tradition in the transition from structuralist linguistics to poststructuralist theories of language, but instead of individual speakers possessing autonomy, poststructuralism shifted this property to the level of the signifying system as a whole. In order to fully appreciate this shift, I want to begin by tracing the incipient systems theory at the heart of Saussure's *Course in General Linguistics*.

Saussure is widely credited as the founder of modern linguistics. He clarified its object and aims as a science by defining a fundamental distinction between speech (parole) and language (langue). In this schema *speech* is treated as a heterogeneous assemblage of objects and properties that straddle physical (e.g., sound waves), physiological (e.g., vocal and auditory organs), and psychological spheres (e.g., the fact that we internalize speech as signifiers/signifieds). In contrast *language* is "a social product of the faculty of speech," meaning the totality of conventions—grammatical and cultural—that regulate that faculty and enable communication—or what Saussure calls the speech circuit (11). Speech "belongs both to the individual and to society; we cannot put it into any category of human facts, for we cannot discover its unity" (9). In contrast, "[language] is a self-contained whole and a principle of classification. As soon as we give language first place among the facts of speech, we introduce a natural order into a mass that lends itself no other classification" (9). The rest of the *Course* is Saussure's attempt to come to terms with the myriad implications of this distinction.

According to this view, the meanings of words result from their position in a system of differences, together with the rules that govern their legitimation. This means that language does not, *cannot* function as a nomenclature because that would imply that objects of reference somehow preexist their appearance in the larger system of signs and their use by speakers. In chapter one I mentioned Saussure as an example of a representational theory of thought because of his two-part division of the linguistic sign into (1) a percept called the *signifier*, or sound-image, and (2) its associated concept called the *signified*. There I argued that the signifier, understood as a sound-image, demonstrated several representationalist assumptions because it is both image-like and psychological (i.e., we can "speak" to ourselves) and in these ways the signifier comes to *stand for* the outside world in a largely transparent, representational or map-like way. Saussure says, "The linguistic sign is then a two-sided psychological entity…the two elements are intimately united, and each recalls the other…" (67). This has important implications for cognition. If signifieds and signifiers are two irreducible parts of the same semiotic vehicle, and both are psychological in nature, then once someone acquires a language thought becomes *thought through signifiers*, and it takes a tremendous amount of mental effort to try and cognize a "naked" thing, uncoupled from this matrix of signification established via the two-part linguistic sign. So while language is a system based on differences and interrelationships—i.e., signs standing in opposition to other signs—once acquired it establishes itself in the psyche and assumes a transparency and fixity in cognition that mirrors many representational theories of thought. Saussure says, "Without language, thought is a vague uncharted nebula. There are no pre-existing ideas, and nothing distinct before the appearance of language" (112). To be sure, Saussure doesn't elaborate on what he means by "thought," but given the linguistic sign's role in synthesizing sounds with concepts, we can assume he is speaking in broad terms about conceptually grounded cognition. Even though Saussure imagined the study of language to be one part of a much broader science of semiology (16), language is given pride of place in our cognitive or inner life. These are prejudices that would pervade all subsequent notions of subject formation.

In this deeper psychological context of conceptual and representational thought, *speech* is the social and material sign that unfolds and concatenates in time; *language* is the virtual system of differences that can be isolated and studied in the abstract (i.e., statically, or outside of time). This means that behind any given moment of speech we can posit a virtual superstructure that constitutes the entire network of relationships determining the values operative in a given utterance. This static arrangement he renders in the classical systems terminology "states,"—a discrete configuration of parts. Saussure calls the study of the static states of language *synchronic analysis*, and the evolutionary changes observed in speech *diachronic analysis*. The purpose of this division was not to cut linguistics into two mutually exclusive fields of analysis (although that is effectively what he did) but to preserve *langue* as a coherent and unified *system*. He continues, "This distinction has to be heeded by the linguist above all others, for language is a system of pure values which are determined by nothing except the momentary arrangement of its terms" (80).

Saussure is making a functionalist argument. Language, so conceived, is a virtual totality that is substrate independent. For example, when he says, "The vocal organs are as external to the language as are the electrical devices used in transmitting the Morse code to the code itself; and phonation, i.e., the execution of sound-images, in no way affects the system itself" (18). According to Saussure, such material features do not belong to the system proper because the changes they inaugurate never alter the systemic fact of language, i.e., the totality of organizing relations. As a consequence of this he says, "The first thing that strikes us when we study the facts of language is that their succession in time does not exist insofar as the speaker is concerned. He is confronted with a state. That is why the linguist who wishes to understand a state must discard all knowledge of everything that produced it and ignore diachrony" (81). This may, at first glance, seem rather strange because it gives pride of place not to the actual utterance encountered in time—what Saussure calls the "natural instinct" (10)—but to the virtual unity/totality of the system which stands outside of time. Saussure likens this to a game of chess: one does not need to know the history of a game's changes and transformations. The only thing relevant to the next move is the present system state. In the effectuation of meaning, utterance and the material vagaries of speech are superseded by the virtual totality of the language state and the coherence it confers on speakers and linguists alike. This substrate independence is writ large in one of the most quotable of lines from the *Course*: "in language there are only differences *without positive terms*" (120).

The takeaway for the field of linguistics is that with respect to synchrony and diachrony, "…they are not of equal importance. Here it is evident the synchronic viewpoint predominates, for it is the true and only reality to the community of speakers (see p. 81). The same is true of the linguist: if he takes the diachronic perspective, he no longer observes language but rather a series of events that modify it" (90). To this end he says, "never is the system modified directly. In itself it is unchangeable; only certain elements are altered without regard for the solidarity that binds them to the whole" (84). The virtual whole is an unalterable fact so long as there are speakers that can constitute the speech circuit. To shed further light on this Saussure draws an analogy to the solar system. If a planet orbiting the sun underwent changes in dimension or weight, "this isolated event would entail general consequences and would throw the whole system out of equilibrium" (85). The system itself persists, but what has changed is the "equilibrium point." In this analogy, the solar system is like the language system, the planets are like the speakers, and their orbits are the conventions that bind them in semiotic exchange. If the equilibrium point changes, this simply means that the conventions for legitimating meaning have changed, *not* the underlying fact of the relationship binding them. If a change in speech is functionally useful it may come to alter the behavior of the system, but the unity of the systemic whole (*langue*) will persist despite these changes.

Saussure's *langue* is clearly an emergent phenomenon. It is animated by the underlying actions of speakers, but it maintains autonomy with respect to these vicissitudes. Saussure boldly tells us that "This new perspective, inspired by historical linguistics, is unknown to traditional grammar, which could never acquire it by its own methods. Most philosophers of language are equally ignorant of it, and yet nothing is more important from the philosophical viewpoint" (85). Taken all together these changes in the landscape of linguistic description represent a fascinating moment in the evolution of systems thinking itself and were coeval with the systems theoretical developments occurring elsewhere in the sciences at the turn of the twentieth century in fields such as organismic biology, gestalt psychology, ecology, and physics (Capra and Luisi 64-68). Saussure's privileging of the patterns of organization (things such as grammar and social conventions), the relationships between parts over the parts themselves (that no sign has a positive or univocal value: all meaning is a relation between opposing signs), and the emergent and virtual character of whole (represented in the primacy of synchrony over diachrony) are all standard markers of a general systems theory that was being codified in the early twentieth century. Under the sway of structural linguistics, literary and cultural theorists examined how all manner of cultural products and practices could be analyzed as instances of an abstract structural (systemic) whole.

This new perspective which Saussure inaugurates is predicated on two important, and ultimately contradictory, perspectives: The heteronomy of the observer, and the autonomy of the system. Saussure adopts the position of a linguist and a scientist, one who can situate himself outside the system in question. But language is an ever-present fact of mental life and so this pretense to objectivity is haunted by the observer's own involvement in the system. Saussure's workaround is to subordinate diachrony to synchrony: the only way the linguist can maintain his heteronomy is at the expense of the historical moment of diachronic utterance. And by privileging synchrony and subordinating diachrony Saussure could only conceive of change as a series of discontinuous and static cross-sections or states. This represents an artificial elision of process.

At the same time, by giving the emergent whole pride of place in his analysis—by subordinating *parole* to the totalizing influence of *langue*—Saussure is acknowledging the autonomy of the signifying system in its relationship to speakers. Saussure's revolution to was to assert that the locus of causal agency is *not* individual speakers and their heterogenous (yet measurable) *parole,* but the higher-order virtual whole that is *la langue*. It could be argued that this systems-based shift in linguistics requires even more counterintuitive mental gymnastics than in other fields in which systems theory was emerging (such as biology or ecology) because with language the scientist is not given any readily identifiable system boundary which they can unequivocally stand in heteronomous relation to. Even in the most philosophically loaded case—i.e., the emergence of consciousness from out of the seed of cellular autopoiesis—the boundaries of the host organism are never in question. The observer is confronted with a well-defined cell membrane that distinguishes the system from its environment.

Given this mixture of heteronomous and autonomous modes, we need to recognize (a) that Saussure's emergentist revolution in linguistics is built on a fundamentally incoherent elision of time and (b) that the material conduit for this emergence is localizable in the *speech circuit,* that is, the circulation of signifiers/signifieds between the auditory, vocal, and mental faculties of two or more speakers of a natural language.

My perception in reading through the *Course* again is that the emphasis on heteronomy and on static states has the effect of treating speakers like "little linguists". It overlooks all the ways in which, phenomenologically speaking, language recedes into the very texture of experience, becomes transparent, and usurps the perceptual and cognitive capacities of speakers from within. Saussure clearly recognizes this psychological transparency early on, and yet he still makes statements to the effect that the synchronic viewpoint of the language state is the "true and only reality to the community of speakers" (90), which phenomenologically speaking makes little sense. Of course, his goal in the *Course* is to establish a scientific method, and so he cannot be faulted for bracketing out the phenomenological approach. I flag it here because this is precisely where enactivism intervenes.

Poststructuralism and Ideology:

The pith of the poststructuralist critique was to deconstruct this pretense to objectivity by turning the theoretical gaze back towards history and context and all the ways these get expressed in discourse; in a word, *ideology*. Once time is introduced (via context), and *langue* reconceived more firmly in terms of process, a heteronomous description of language states becomes incoherent.[66] These states are always undergoing change, subject to the vicissitudes of context and to what Derrida will call the "play" of the signifier which, because of the system's inherently dynamic nature ends up exceeding the intentions of speakers and communities of speakers, and most importantly, undermining the presumed objectivity of the observer. *In these ways, poststructuralism represents a dynamic systems critique of structuralism*.[67]

In order to illustrate this point, I have chosen to touch on one hugely important and representative example for literary theory.

Jacques Derrida's "Structure, Sign, and Play," first given as a lecture at a Johns Hopkins in 1966 and shortly after published in his book *Writing and Difference*, adopts a sweeping, transhistorical scope. In sixteen short pages he impugns the entirety of Western philosophy by analyzing the notion of *structure* on which these epistemic formations depend. Shortly after this lecture Derrida would go on to publish three books in quick

[66] And not just language states, but any kind of cultural, semiotic system that one wants to study synchronically because the signifying processes of the observer are always insinuating themselves with the signifying structures of the object of study.

[67] Again, I merely want to emphasize that all emergent systems are dynamic systems, and the only reason I am making this distinction here is that Saussure's program is predicated on an artificial elision of time.

succession, all dealing with decidedly post-structuralist themes, and which developed a host of influential and interlocking concepts such as arche-writing, trace, supplementarity, and différance. But throughout the past decade of my studies in literature, I have returned time and again to "Structure, Sign, and Play," because to my mind it contains the germ of all his subsequent work.

Specifically, this essay attempts to analyze "the structurality of structure" which "has always been neutralized or reduced" by Western philosophy and science by "giving it a center or of referring it to a point of presence, a fixed origin" (*Writing and Difference* 278). All epistemes—that is, all organized bodies of scientific, "certain," or veridical knowledge—depend on something that can ground, orient, and organize the totality of its structure. This is what Derrida refers to as a "center." Centers are unique features of the system because they cannot be subject to, or dependent on, any of the other structural elements as such dependence would undermine this totalized organizing function. This leads Derrida to assert that "The center is at the center of the totality, and yet, since the center does not belong to the totality (is not part of the totality), the totality has its center elsewhere. The center is not the center" (279). All epistemic discourse is centered discourse because "The concept of centered structure…represents coherence itself" (279). Centers, it would seem, are unavoidable.

The fact that a center does not depend on any other element means that it is stable enough to ground the functioning of all other pieces. Derrida says, "By orienting and organizing the coherence of the system, the center of a structure permits the play of its elements inside the total form" (279). In other words, the center has the dual purpose of animating the terms within its system—giving them meaning and coherence—and simultaneously limiting the extent of their reach—their so-called "play." Here "play" can refer to the natural tendencies of speakers to change and deform the inherited structures of language to meet their specific needs. It can also refer to the emergent properties of signification in which its elements begin to aggregate and concatenate in wholly unexpected ways. In either case, if the elements of a structure stray too far from their center they begin to lose the intelligibility and coherence which the center originally conferred. *This relationship between a "center" and the limits of play that it permits is the essence of a system boundary. I would argue that the "structurality of structure" which Derrida characterizes in terms of center and play articulates precisely the same dynamically enacted inside/outside distinctions we see in autopoiesis*.

Throughout the essay, Derrida's analysis and critique of "structurality" and "centered discourse" is directed primarily toward Western metaphysics because of its perennial and explicit appeals to *being as presence*, by which he means some concept which purports to stand outside the signifying system and ground all of its internal relations in a fixed and totalizable order. He says, "it could be shown that all the names related to fundamentals, to principles, or to the center have always designated an invariable presence—eidos, archē, telos, energeia, ousia (essence, existence, substance, subject) alētheia, transcendentality, consciousness, God, man, and so forth" (279-280). By "presence" he means a term which is not subject to the differential relations identified by Saussure in his analysis of linguistic structure. Saussure says that in a signifying system there are no positive terms. Here in the context of Western metaphysics, Derrida is pointing to totalizing concepts that—due to their organizing or centering function—purport to exist outside this system of differences. All these concepts are examples of "centers," concepts that maintain the coherence of metaphysical discourse by organizing all the other terms within their ambit and which are not themselves subject to the signifying structure. We also need to recognize here that this is *not* the result of some malign force or singular intention at work in "discourse." Centers are not established as a nefarious means of duping people. They arise endogenously in a given discursive arrangement—what we might call a subsystem of the larger *langue* (e.g., metaphysics)—as it begins to self-organize, evolve, and establish the terms of its coherence.

The so-called *event* "that has occurred in the history of the concept of structure" (278) which opens the essay, and which serves as Derrida's first tongue-in-cheek attack on structuralism, refers to the moment in which this "structurality of structure" was first subjected to philosophical scrutiny, and refers not to an actual event but to a series of philosophical interventions and transformations in the Western episteme that made the analysis of structure possible. Derrida cites Nietzsche, Freud, and Heidegger as three originators of this deconstructive, structurally self-conscious philosophical mode. In their body of work he sees a movement away from the center "thought in the form of a present-being" and the dawning recognition that "the center had no natural site, that it was not a fixed locus but a function, a sort of non-locus in which an infinite number of sign-substitutions came into play. This moment was that in which language invaded the universal problematic" (280). In other words, this "event" represents the beginning of the linguistic turn.

Despite these inroads into the thinking of structure, we cannot lose sight of the fact that "We have no language—no syntax and no lexicon—which is foreign to this history; we can pronounce not a single destructive proposition which has not already had to slip into the form, the logic and the implicit postulations of precisely what it seeks to contest" (280-281). Every attempt to break with a center inevitably asserts another because a center is the precondition for meaningful utterance, ordered discourse, and knowledge.

To better illustrate this point Derrida cites Saussure, whose two-part division of the sign into signifier/signified was both a decisive evolution in the self-conscious critique of structure, and a reproduction of the metaphysical problem. Saussure's conception of the "sign" marks the difference between the material and conceptual features of language, a move that unsettles a metaphysics of presence because it shows how concepts are not pre-existing entities but emerge from the constitutive differences that define a language system. Derrida continues, "But, as I suggested a moment ago, as soon as one seeks to demonstrate in this way that there is no transcendental or privileged signified and that the domain or play of signification henceforth has no limit, one must reject even the concept and word 'sign' itself—which is precisely what cannot be done" (281). The self-conscious critique of structurality embodied in Saussure's concept of "sign" itself depends for its coherence on a prior distinction between the sensible and the intelligible, between materiality and ideality. To invoke the notion of the sign is to invoke "this opposition and its system" (281), in effect, an appeal to another center. Nevertheless, the notion of the sign enables a novel form of critique, and therefore:

> …we cannot do without the concept of the sign, for we cannot give up this metaphysical complicity without also giving up the critique we are directing against this complicity, or without the risk of erasing difference in the self-identity of a signified reducing its signifier into itself or, amounting to the same thing, simply expelling its signifier outside itself. (281)

Thinking structurally confronts one with a double bind. It participates in this back-and-forth play of presence and absence: asserting a presence (a coherent critique) by means of an absence (a concept that is itself subject to differential terms); signaling an absence (a system of differential relations) by means of a presence (a totalizing concept). The structuralist concept "sign" assumes a centering function, but at the same time, one cannot abandon this Saussurean division because without it the center would inevitably reassert itself in a still more intractable form.

What are we to make of this odd state of affairs? Derrida is *not* rejecting the idea of a center per se. He says—in so many shape-shifting words—that the center is unavoidable, that it represents "coherence itself." His critique is levelled at those discourses (e.g., metaphysics and ethnography) that blithely and/or unconsciously assert them, for as he states, "the quality and fecundity of a discourse are perhaps measured by the critical rigor with which this relation to the history of metaphysics and to inherited concepts is thought" (282). In other words, he is not denying the epistemological necessity of a center. His aim is to produce a reader or thinker that is aware of this *dynamic*—and thus unstable—process of knowledge production, meaning-making, and interpretation.

The autopoiesis-like descriptions of signification in this essay—that is, the dynamic interrelationship between the "center" of a signifying system and the limits of play that define its boundary—further attest to the dynamic systems logic operative in Derrida's poststructuralism. Autopoiesis utilizes the language of "recursion" to denote persistence of a form/pattern despite material flux, while Derrida uses the notion of a "center" to denote the self-referential process of signification, *self-referential* because given a particular discursive arrangement—e.g., the "sign" of structuralism, or "substance" (*ousia*) for metaphysics—all signs can trace their meaning back to the center and the coherence it confers. All signs refer to this center, and this center refers only to itself.

Furthermore, the play of presence and absence which Derrida casts as a metaphysical problem has a clear analogue in autopoiesis. By drawing a boundary between itself and the environment the autopoietic entity establishes itself as a self-present entity. Yet this embodied presence is enacted only by compensating for perturbations from without, *not* as a function of objective access to the real (but as Varela stated in the previous chapter) by engaging with, "a basic stuff to in-form from its own perspective" (Varela, "Autopoiesis and a Biology of Intentionality" 8). It effectively ignores those environmental features which do not directly contribute to the maintenance of the system's autopoiesis. Because they are not directly relevant, they are effectively rendered "absent" to the system's cognitive processes.

We could trace many different iterations of this dynamic systems logic that is poststructuralism, as well as the many autopoiesis-like descriptions of formal self-reference that can be found therein. But the big three for literary theory have been

Jacques Lacan, Michel Foucault, and Jacques Derrida.[68] Lacan because he extended signifying processes down into the structure of the psyche and the unconscious, thus

---

[68] For example, Lacan's "The Instance of the Letter in the Unconscious," initially given as a lecture in 1957 and later published in the *Ecrits,* provides us with one of the most granular accounts of signification (i.e., the "letter") and its emergence in the unconscious and libidinal faculties of the psyche. There Lacan interrogates what he calls the "algorithm" of Saussurean linguistics S/s, with the capital 'S' (signifier) standing above the 's' (signified) because it is the articulated, differential units (in the case of spoken language, phonemes) that constitute the signifying system, *not* the ideas they supposedly stand for. The bar '/' represents the absolute difference between these two orders, which in Saussure's reckoning cannot be crossed. In this relatively orthodox structuralist context Lacan asserts that the "letter" can be understood as "the essentially localized structure of the signifier" (418), meaning any discernable unit that can serve the purposes of differentiation and articulation. He asserts that all signifying systems "are subject to the twofold condition of being reduced to ultimate differential elements and of combining the latter according to the laws of a closed order" (418).

It is this second aspect of signification that I want to linger on. Lacan continues: "The second property of the signifier, that of combining according to the laws of a closed order, affirms the necessity of the topological substratum, of which the term I ordinarily use, 'signifying chain,' gives an approximate idea: links by which a necklace firmly hooks onto a link of another necklace made of links" (418). By calling such laws a "closed order," and furthermore by linking it to a topological substratum (more on this in a moment) Lacan's description—like Derrida's—is redolent of autopoietic enactivism and its description of "operational closure." In Lacan's case, the rules that govern the concatenation of signifiers in a chain must remain invariant in order for signifieds to emerge. This is closure of a *process* or organizing relations akin to those sequestered metabolic processes occurring on the inside of a cell, except that in the case of signification we might view these as core features of grammar that persist despite incessant variation in speech (as Saussure does), or, as Lacan proposes in this essay, as the unconscious processes of condensation and displacement that underlie one's experience of dreams. It is worth mentioning here that this is one of the most explicit examples in the *Ecrits* of Lacan extending structuralist analysis into the unconscious. He insists that this is not a *new* idea, but one that was already operative in Freud's *Interpretation of Dreams*. Freud elaborated the structure of dreams in terms of "condensation" and "displacement." By using Roman Jakobson's ideas of metaphor and metonymy as representative of the differential (synchronic) and temporal (diachronic) dimensions to signification, Lacan argues that Freud's use of condensation and displacement in dream logic demonstrates how the unconscious is, in effect, *structured like a language*. There are many facets to Lacan's thought in the *Ecrits* because it represents a compilation of roughly thirty years of work. The oft-quoted idea that "the unconscious is structured like a language" (Lacan 737) is largely associated with Lacan of the 1950s. The Stanford Encyclopedia notes that "the Lacanian unconscious is structured like '*un langage*' and not '*une langue*.' Although both French words translate into English as 'language,' the former (*langage*) refers to logics and structures of syntax and semantics not necessarily specific to particular natural languages…" (Johnston). The important point here is that Lacan is using structuralism to highlight the formal characteristics of these unconscious processes.

But to return to the quotation from page 418. It is this operational closure which in turn "affirms the necessity of the topological substratum" which he calls signifying chains. This is also very suggestive because it demonstrates an interdependence between an abstract process and its constitutive "materiality." I place materiality in quotes because given his focus on the psyche and on the structure of desire, this substratum does not need to be manifestly material (as in sound waves) but some kind of spatially distributed phenomena. In Greek, *topos* denotes place, or space, and in mathematics we have seen that "topology" is concerned with the geometry of spatial relations, specifically, "with those properties of figures and surfaces which are independent of size and shape and are unchanged by any deformation that is continuous, neither creating new points nor fusing existing ones; hence, with those of abstract spaces that are invariant under homœomorphic transformations" (OED "topology"). For this reason, topology is a uniting psychosocial and psychosexual processes of development with the act of interpretation, and in which art plays an important role in both mediating and placating our desires. Foucault because his "genealogical" approach to history sought to analyze the underlying context and conditions from which discursive structures emerged, what Foucault called an "episteme." This approach united historical processes with interpretive processes and became foundational to the new historical approach in literary theory.

---

cornerstone of dynamical systems theories. It describes the constancy of patterns and configurations *in space*. According to Lacan, the very notion of a closed order requires topologically defined underpinnings. In this case *signifying chains*—sequences of "letters" (any articulable unit at work in the unconscious) whose concatenation facilitates the closure of a systems' governing laws. To be sure, the imprecision here would vex a neuroscientist although it would be interesting to try and map the notion of neural assemblies onto this idea of a topological substratum Lacan is developing here.

Together these revisions/extensions of Saussure's structuralism demonstrate that in the algorithm of the sign, the signifier is in fact primary. The "letters" by which the psyche is organized possess a causal efficacy that can determine and deforming the equilibrium point of the psyche. These signifiers will naturally (spontaneously) form "boundaries" that constitute a structure of desire, and that do so irrespective of the conscious intentionality of speakers. Like an autopoietic system, a signifying chain is sufficiently complex to transcend simple descriptions of openness or closure because its operational closure is dependent on a the energetic "openness" evinced by signifying chains. The very act of privileging the signifier, of placing it over the signified puts us, once again, on the path to a "dynamic" systems critique of structuralism because from this it follows that the "signifier in fact enters the signified" (417), that there is "an incessant sliding of the signified under the signifier" (419), such that what before was taken to be a straightforward process of reference from word to ideas, we find only an ongoing process of signification—patterned concatenations of signifiers which cannot be static or fixed referents. In effect, the signified is always displaced by other signifiers and signifying chains, such that signification itself comes to assume a kind of agency, that is, it evinces self-organizing and autonomous characteristics. Lacan says, "Whence we can say that it is in the chain of the signifier that meaning *insists*, but that none of the chain's elements *consists* in the signification it can provide at that very moment" (419). Because we can never perfectly "fix" (i.e., render static) the signified and so arrest the chain of signification, what signification attests to, first and foremost, is its own self-referential processes of differentiation and combination. This difference between insisting/consisting is the essence of the emergent nature of the process. If meaning *consisted* in the relation with signifieds, then meaning would be terminal. Insistence is the effect of a dynamically organized whole perpetuating itself.

I think that poststructuralist theory offers us many different versions of this autopoiesis-like dynamic systems critique of structuralism, all of which would involve pointing out the same logic under different guises. Interesting (for some of us at least) but not fundamentally advancing the point I have already made using Derrida's "Structure, Sign, and Play," and to a lesser degree, Lacan's essay here. But, in order to round out my claim regarding the "big three" as I have called them, I would also direct the reader to Foucault's chapter "Panopticism" from *Discipline and Punish*. In his explication of the panopticon, we see a system of power-knowledge in which the organizing relation—surveillance—is reproduced and perfected homogenously throughout the entire panoptic system. The tower at the center of the panopticon and the walls of the prison also play nicely into the center/boundary conceit operative in my readings of Lacan and Derrida.

Furthermore, his conception of "power-knowledge," in which power creates, elaborates, and amplifies itself suggests the positive feedback of a cybernetic system. Finally, Derrida, because his critique of Western philosophy and its metaphysics of presence effectively equated philosophy with interpretation, thereby placing literature on a par with philosophy—an equivalence that helped expand literature's remit and importance during this period.

There is one more theory of subject formation worth mentioning here: the sociological linguistics promulgated by the Bakhtin Circle, whose work emerged in the 1920s and 1930s in Russia and was not picked up in American academies until the 1970s and 80s ("Mikhail Bakhtin," *Norton Anthology* 1073). The Bakhtin Circle's critique of the "monologic utterance," Romanticism, and the theory of expression they called "individualistic subjectivism" represent an interesting middle ground in this genealogy of literary systems thinking.

Theorist and key member of the Bakhtin Circle Valentine Vološinov defines the monological utterance as a "purely individual act, the expression of an individual consciousness" (84), and individualistic subjectivism as the larger theory of subject formation that "took the monologic utterance as the ultimate reality and the point of departure for its thinking about language" (Vološinov 84). Individualist subjectivism maintains that expressible content can "somehow take shape and exist apart from expression" (84), the idea that because these thoughts, intentions, feelings, etc. occur *in me*, that I am necessarily their cause, origin, and author. The dialogic approach developed by the Bakhtin Circle turns away from the individual and towards the social body by arguing that "It is not experience that organizes expression, but the other way around—*expression organizes experience*" (85). For the Bakhtin Circle many forms of experiential structuration exist—not just linguistic. Their approach revolves around the sociologically formed subject, of which language forms an important but still relative part. Like structuralism, dialogism views language as a social fact, but unlike structuralism, language is, from the outset, a diachronic and material phenomenon that circulates in concrete ideological ways. Its existence does not manifest in abstract language *states*, but as a function of historically embedded and ever-evolving utterances.

I am very partial to Bakhtinian dialogism because it strikes me as an inherently *ecological* theory. The systems it analyzes are ideological systems in which individuals produce and consume utterances like members in a food web. By focusing on processes of production, consumption and transformation between self and society—for example the interaction and competition between local languages and dialects on the one hand, and scholarly or administrative languages on the other; or between certain genres over others; or what Bakhtin called the "centripetal" (Bakhtin xviii) ideological forces exhibited by certain discourses such as ethics, law, art, politics—the Bakhtin Circle trades an abstract system of reference in favor of a living system of exchange, in how these concrete features of utterance circulate in social formations and thereby create political and material realities.

Language Use via Linguistic Coupling:

"Structure, Sign, and Play" articulates precisely the same process of epistemological emergence outlined by Clarke in the previous chapter but in the context of Western metaphysics and continental philosophy. It is a paradigm of "recursive system/environment" coupling that precedes all realist or empirical modes of representation. Autopoiesis and second-order cybernetics say that there is no way to know a thing outside of this structural coupling; deconstructionism says that there is no way to know a thing, outside the play of presence and absence that animates signification.

Like a signifying system, a cognitive system forecloses totalization, both from without and within: Because its self-making processes depend only on salient features of its Umwelt, the cognitive system will never have objective access to the real. Because of its operational closure, an outside observer will never have totalized knowledge of the phenomenology of the cognitive system. Simply put, the boundaries of autonomous systems foreclose totalization.

Therefore, even though structuralist and poststructuralist philosophy shifted the locus of dynamical organization away from individual cognitive systems (i.e., speakers/subjects) and onto the emergent properties of the linguistic system, this does not substantively alter the epistemological implications outlined by Clarke. In both perspectives access to a "self-present" or objective/univocal meaning is vacated. In each case the system, be it cognitive or linguistic, enacts its own objects of reference. Embedded in this claim is the idea that both are dynamic and processual. Each system is composed of what Deacon would call "intrinsic constraints," their parts cannot be properly understood outside of this participation in a larger dynamic process and therefore represent a higher-order form of emergence in which "the part/whole distinction and the synchrony/diachrony distinction are intertwined" (Deacon 164).

I have tried to emphasize the differences between structuralist and poststructuralist analysis in order to highlight how Saussure's pretense to heteronomy in relation to linguistic structure reproduces the same descriptive problems we have encountered time and again in biological and neuroscientific contexts: In his or her relationship to mind or living organization, the observer always becomes entangled in the pattern of organization they are attempting to heteronomously describe. The observer projects (or hides, as the case may be) a teleological description or an intentional interpretive agency in their description in order to make the description cohere. In the case of Saussure's structuralism, Derrida's shows us how his binary conception of the sign is already part and parcel of the structure he is describing. The pith of Derrida's essay, and of all poststructuralist critique, is to throw a wrench into this pretense to objectivity. Nevertheless, we need to recognize that structuralism, poststructuralism, and deconstruction exist on a continuum insofar as they identify the signifying system as an emergent and autonomous entity.

To be sure, this systems-theoretical comparison between structuralism and cybernetics/autopoiesis depends on a certain level of abstraction. If we think about language as the "environment," (both the sheer fact of acquiring a natural language or specific discourses therein), and the speaker as the "cognitive system," then the aptness of this comparison comes more sharply into focus. In the same way that a bacterium metabolizes energy in its environment as part of its ongoing self-maintenance, the human mind metabolizes linguistic material and assimilates it into its idiolect, dialects, and larger repertoire of discourses. This is an abstraction because in this proposed equivalence, the role of other speakers—i.e., other cognitive systems—is not explicit but embedded in the very notion of an acquired language or "discourse."

Language does not speak itself. It is an emergent phenomenon that depends on the underlying intentionality of speakers who shape their utterances according to specific purposes, audiences, and contexts. These intentions animate an otherwise inanimate material. Its purposes and ends are supplied by the teleodynamic processes of speakers. If there were no such purposes or ends, however inchoate, then how could speakers begin to compose their sentences in logical and mutually intelligible ways? As Wittgenstein famously said, "the meaning of a word is its use in the language" (25). Speakers *enact* meaning according to this confluence of intentions, expectations, and contexts.[69]

Wittgenstein's dictum can be easily mapped onto the notion of structural coupling articulated by autopoiesis and all the counter-intuitive implications that the position

---

[69] To be sure, these are longstanding criticisms of structuralism. For one, using Saussure's structuralism as a way to critique or undermine logical positivism—in which language is only considered in terms of *reference* from word to object (empiricism) and *propositions* (logical validity). As Surette tells us, more than a decade before Derrida came on the scene Wittgenstein's *Philosophical Investigations* (1953) and J. L. Austin's "How to Do Things with Words" (1955) "called into question the Frege-Russell view that language was purely propositional. Both Wittgenstein and Austin drew attention to the fact that language use predominantly consists of 'speech acts,' that is, acts of communication, as opposed to pure acts of reference" (Surette 416). This leads Surette to conclude that "Derrida's attack on logical positivism in *Of Grammatology*, then, was a day too late as well as a dollar short" (Surette 416).

In addition to recapitulating this well-established critique of poststructuralism for us, Surette also makes several historical observations which attempt to rationalize this fetishization of signifying structures over other, more pragmatic approaches. Surette argues that postwar French intellectuals were disaffected and disillusioned by "liberal capitalism," and consequently latched onto Saussure's structuralism because of its anti-essentialist core. He says, "The adoption of Saussurean structuralism permitted French intellectuals to shift their polemical target from liberal capitalism to empirical science, understood as the philosophical underpinning of liberal capitalism…Stuck in an intellectual economy of opposition to domestic capitalism, they found that debunking empiricism's claims to truth could serve as a more palatable alternative to the promotion of the dictatorship of the proletariat" (Surette 417). This is an interesting observation, and one which I have heard articulated in various ways by public intellectuals as a way to discredit the current state of humanistic study in public universities. I flag it here because it occupies a recurrent part of the zeitgeist around poststructuralist discourse.

entails for notions like "representation" and "information." As we discussed in detail in the previous chapter, *structural coupling* means that the autopoietic system compensates for perturbations in its environment in an endogenous way. It maintains its autopoiesis by selectively engaging with environmental stimuli in the compensatory way. It does not *represent* this outside world by reproducing it inside the system in any kind of one-to-one map-like fashion.

In *The Tree of Knowledge* Maturana and Varela extend this notion of structural coupling into what they call "third-order structural coupling," (181), and later on refer to as "social coupling" (206), which both describe linguistic processes and spoken language but in the terms laid out by the systems theory and cybernetics on which their original theory of autopoiesis is built.[70] Third-order structural coupling describes the process by which organisms enter "intro structural coupling with other organisms" (180). In this process, the other organism is initially treated like any other (non-biotic) feature of the environment, that is, they are one more source of external perturbation. However, "It is *possible*…for

---

[70] I will admit that the terminology here is not always intuitive at first glance, for nowhere in their book *The Tree of Knowledge* do they describe "second-order structural coupling." They say that "The structural coupling with the medium as a condition of existence covers all possible cellular interactions. Therefore, it includes interactions with other cells as well" (77). In other words, the phrase "structural coupling" sufficiently accounts for/describes the epistemological emergence we analyzed in chapter four and which necessarily follows from the ontological emergence of an autonomous autopoietic unity. This epistemology rejects objective access to the real, as well as any conception of "representation" understood as a one-to-one mapping between inside and outside states. Thus, you may be wondering how they went from "structural coupling" to "third-order structural coupling" without developing anything they call "second-order structural coupling." In these same pages, Maturana and Varela go on to describe "metacellulars," which they also call a "second-order unity" and which together denote a cell that is capable of giving rise to lineages, reproducing, and "which will have a structural coupling and ontogeny adequate to its structure as a composite unity" (78). Ultimately, this leads Maturana and Varela to describe social/linguistic coupling as "third-order structural coupling," which denotes "the same mechanisms…discussed in relation to the makeup of second-order autopoietic unities" (181). Stated simply, "third-order" coupling arises from metacellulars which they call a "second-order unity." A lot of jargon which is not always productive or intuitive to navigate.

these interactions between organisms to acquire in the course of their ontogeny a *recurrent* nature. This will necessarily result in their consequent structural drifts: co-ontogenies with mutual involvement through their reciprocal structural coupling…" (180), which according to Maturana and Varela gives rise to a new phenomenological domain—social coupling, or what I would rephrase in more phenomenologically-loaded language as *the social lifeworld*. This may all sound a bit abstract, but the essential point is that recurrent patterns of perturbation between organisms give rise to well-defined social behaviors—everything from the basic sexual processes of gametes to the complex parenting behaviors of birds.

In this (autopoietic) context of "structural coupling," we can understand spoken language as simply the most complex evolution of social coupling. Maturana and Varela go on to describe what we could just as easily call "linguistic coupling" as follows:

> language as a phenomenon takes place in the recursion of linguistic interactions—linguistic coordinations of linquistic [sic] coordinations of actions. Therefore, the linguistic domain becomes part of the environment in which linguistic coordinations of actions take place, and language appears to an observer as a domain of descriptions of descriptions. (*The Tree of Knowledge* 211)

In other words, language is formal self-reference made explicit (communicable), a capacity that, like all structural coupling, first develops in relation to the individual and their embodied cognition. According to this perspective, what the observer perceives as processes of representation are really just complex elaborations of a basic structural coupling by which an organism compensates for perturbations and so maintains its autopoiesis. This is the idea of sense-making developed in the previous chapter, and which

I recast in terms of Deacon's emergent dynamics as constraint preservation. If we step out from this cybernetic language, we can see that the picture Maturana and Varela paint of language development in *The Tree of Knowledge*—as a gradual increase in the complexification of the social lifeworld—makes perfect sense from an evolutionary perspective. The fossil record clearly shows that early hominids were fundamentally social creatures. To this end Maturana and Varela say, "Male and female were attached to each other by a permanent and nonseasonal sexuality. Through conservation of food sharing and male participation in the care of the young, this led to a biology of cooperation and linguistic coordination of actions" (221).

Very clearly language cannot speak itself, and this means it must be enacted by individual organisms. At the same time, it is a shared, social phenomenon. It does not belong to the individual. We see echoes of this tension between the autonomy of the system and the autonomy of speakers at work in Saussure's description of linguistic evolution. He calls functional changes in the system (i.e., those that alter its equilibrium point) "fortuitous" changes of state (85) and he says that "In each [fortuitous] state the mind infiltrated a given substance and breathed life into it" (85). This is a key moment in his analysis of the virtual linguistic system because it acknowledges the teleodynamic underpinnings of language. As with evolution by natural selection, linguistic evolution is understood as a fortuitous, non-teleological process, but one which depends on the purposes and intentions of individual speakers.

Yet we can also explore the implications of this dynamic systems logic via a different level of abstraction, by shifting the locus of intentionality from speakers to the signifying system itself. This is a move that was inaugurated by Saussure's structuralism and intensified to its logical extreme by poststructuralists such as Derrida. But this shift in the locus of causal agency requires an artificial elision of the agency and intentionality imparted to signifying process by its speakers. It subsumes these individual intentions to the emergent properties of signification. In the previous paragraph I began by observing that "language does not speak itself"; this is patently true. Yet Derrida's description of the play of presence/absence highlights the fact that signification is always exceeding the intentionality of speakers. Catherine Belsey provides us with a lucid gloss here: "Language is a system which pre-exists the individual and in which the individual produces meaning. In learning its native language the child learns a set of differentiating concepts which identify not *given entities* but *socially constructed meanings*. Language in an important sense speaks us" (Belsey 41). Not only does language pre-date the subject and thus provide a ready-built repository of meanings, but the fact that it is enacted and co-determined in relation to other minds means that one's meaning is not reducible to one's intentions. It has to be recognized by an Other.

This is where the comparison between cognitive system/environment and speaker/signifying system starts to break down. Whereas the system/environment coupling of autopoiesis focuses on how singular systems enact meaningful environments, post/structuralism focuses on how meaning is enacted by the emergent processes of signification, which by necessity supervenes on a community of cognitive systems. In the poststructuralist paradigm language/discourse is the dynamic system metabolizing the utterances of its speakers, subsuming their agency and autonomy in a larger emergent process. In the post/structuralist approach *language* is the autonomous system, *not* speakers. The structuralist paradigm prioritizes and privileges the autonomy of language over and above the autonomy of speakers.

This is not to say that the enactivist approach to explaining cognitive content is "more correct" than structuralism and ilk. After all, they are both systems theories in which boundary and center can be defined in different ways. The structuralist approach is ideally suited to analyzing language, understood as a "social fact." Enactivism is ideally suited for understanding how the body and its know-how animate language in ways not discernable by structuralism's emphasis on the social fact. These differences only become problematic when one paradigm is asked to do too much. In the case of poststructuralism, signification became a totalizing perspective by which to understand cognitive processes and subject formation. It prioritized culture and the social fact over the autonomy of speakers—of which it was deeply suspicious.

Autopoietic enactivism and post/structuralism are *united* by the idea that subjectivity is a systems-level phenomenon: it is a configuration of differential parts that the cognitive agent enacts, producing a higher-order unity or systemic identity. [71] The

---

[71] In many ways, this perspective (i.e., its formal characteristics) is legitimated by contemporary neuroscience which makes fundamental distinctions between sensory "events" (what Lacan might call the inanimate materiality of the letter) and perceptual "experiences" (the unification of this materiality into autonomous "centers" of valence and meaning). In our discussion of Albright in chapter one, we saw how perceptual experience is understood as an interpretive process that disambiguates stimuli via different types of *context*. Just as speaker, audience, and immediate context constrain the process of linguistic interpretation, so too do spatial, temporal, and evolutionary contexts constrain perception, elevating the incoherent noise of sensory "events" to the coherence and orderliness of perceptual "experience." Thus, in terms of formal characteristics both processes are interpretive and so depend on context. The glaring difference between the two lies in the nature of their content: in neuroscience we are not dealing with socially mediated "signs" but endogenously generated percepts.

crucial difference is that structuralist approaches make the socially constructed semiotic system the essential feature of subjectivity, whereas autopoietic enactivism demonstrates the embodied and socially un-mediated dimensions to subject formation. Thus, my contention in this chapter is that autopoietic enactivism can obviate the linguistic overdetermination of structuralist, poststructuralist and deconstructionist epistemology by centering subjectivity foremost in the body. It offers up the body as a center, unmediated by the social fact. As I have tried to show, Thompson's enactivism (with a little help from Deacon) can effectively naturalize the autonomous observer self in its relation to biology and neuroscience, but I also believe we can use enactivism as a way to ground and naturalize the "subject" in relation to poststructuralism. This move can give literary theory another way out of the morass of language that resulted from the linguistic turn.

Belsey's earlier quotation—the idea that "language speaks us"—highlights for us the top-down causal processes at work in language and meaning-making. Post/structuralism brought these processes firmly into view under the banner of "ideology." This is the origin of what I cited in the introduction as the hermeneutics of suspicion. From this point of view, any mention of an individual's "autonomy" elicits paranoia because this word is wrapped up in structures of discursive power that recognize certain subjects to the exclusion of others. However, in the context of biology, neuroscience, and phenomenology autonomy is not a dirty word. Autopoietic enactivism offers us a reparative reading of the term. It recovers this concept via biological and phenomenological means and makes it meaningful once again for a larger humanistic audience. This, I contend, is another profound upshot of autopoietic enactivism for the larger discourse of the environmental humanities.

The kneejerk deconstructionist or poststructuralist reaction to such a move would be to impugn the notion of "autonomy" on which its descriptions of cognition rely, but this would be a misreading of autopoietic enactivism. Like the systems theory that is structuralism, or the dynamic systems theory that is poststructuralism, enactivism describes a form and not a content (i.e., a signified). Its recourse to "autonomy" simply names the ineluctability of a center albeit at a different scale of description. The fact that enactivism and poststructuralism share a basic formal understanding is significant because it lends credence to the enactivist approach from within poststructuralism's own ranks. The crucial difference is that autonomy names a center that cannot be uncoupled from the self-presence of a body. For enactivism, every conscious or unconscious process of perception depends on this invariant embodied underpinning. Because of this, we can indeed make the sheer fact of body a fixed point of reference that stands outside the social fact of language, it is *unmediated* by its conventions. Discursive centers are simply a higher-order expression of this foundational, teleodynamic sense-making.

Finally, I would return us Wittgenstein's dictum "meaning is use." It offers us a practical and commonsense way of centering the problematic of language. Meaning always arises in specific contexts. It is case-specific, agent-relative, and subject to the constraints of immediate purposes, audiences, and contexts. In the final section, I want to turn to an alternative paradigm of semiotics that will allow us to see language—*not* as a top-down process of subject-formation that tends towards the malign—but rather as an instrument of embodied agency.

Peirce's Habit-Based Theory of Signs:

One the one hand, structuralism and its offspring demonstrate how systems theory has been a formative part of literary studies and the linguistic turn. On the other hand, it is a systems theory devoid of naturalist grounds and material underpinnings. The epistemological shift that follows from a recursive system/environment coupling still obtains, we just need to center it in bodily processes and ~~find~~ articulate a semiotics that does not become consumed by its own virtuality, untethered from the real.

Charles Sanders Peirce developed a semiotic and philosophical tradition that runs parallel to that espoused by Saussure, and from which we can begin to integrate bodily autonomy with the representational capacities of signs. It offers us a pragmatic counterpoint to the abstract system-state psychologizing operative in Saussurean structuralism. But until recently, it was an approach that remained largely isolated in the field of analytic philosophy.

Peirce's semiotics is described as "semiotic realism," which coming on the heels of our analysis of Saussure may sound a like a contradiction in terms. After all, if signs are constituted in an abstract system, then how can they occupy independent causal/ontological status in the world? How can they be manifestations of the objectively real itself, and not mere correlations between an ideal system of thought and being?

Like structuralism, Peirce's notion of the sign mediates one's access to thought. Thought by this measure it is not self-present, auto-affection. But for structuralism, signs are dyadic (signifier/signified) units held together and made comprehensible only via static states of the whole system of values (*langue*). For Peirce, the sign is distributed in a triadic *process* of sign, interpretant and object. For structuralism, the virtual totality is primary. For Peirce, the utterance and its immediate context is primary. Peirce's theory works pragmatically and from the ground up, focusing on the habitual and experiential processes by which the meanings of signs get established over time. This extension across time also accords with the circularity that defines enaction, the temporal contexts outlined by Albright, and (as we will see in the following chapter) with neuroscience's understanding of habituation.

In his introduction, James Hoopes begins his exposition of Peirce's semiotics by situating it in the larger context of Western philosophy of mind, specifically Descartes and Locke (Hoopes 5-6). He says, "Peirce rejected both Cartesian dualism and the Lockean description of all thought as the experience or internal perception of ideas" (Hoopes 7). For Descartes, thought is understood as unmediated and intuitive access to one's interiority, an insight which Locke extended by arguing that "...every idea...[is] either a copy of a sensation or a reflection upon the mind's operations" (Hoopes 6).

In this context, Peirce uses the concept of the sign to intervene in this unmediated state of affairs between thought and ideas. Hoopes says, "The difference between an 'idea' and a 'sign' is at the heart of Peirce's semiotic" (Hoopes 7). He continues, "A 'sign,' as Peirce employed the term, is also a thought, but it differs from an "idea" in that its meaning is not self-evident. A sign receives its meaning by being interpreted by a subsequent thought or action" (Hoopes 7). To illustrate this Hoopes gives the example of a stop sign. We perceive its features—hexagonal shape, red color, a sequence of letters—but "It is only in relation to a subsequent thought—what Peirce called an interpretant—that the sign attains meaning. The meaning lies not in the perception but in the interpretation of the perception as a signal to stop or, better still, in the act of stopping" (Hoopes 7). For Peirce, this basic pattern obtains in *every* thought: "every thought is a sign without meaning until interpreted by a subsequent thought, an interpretant. Thus the meaning of every thought is established by a triadic relation, an interpretation of the thought as a sign of a determining object" (Hoopes 7). All thought can be understood on the basis of this processual triad.

The upshot to this view is that it unites all manner of cognitive phenomena under the same banner—everything from a feeling to a percept to an utterance—because they can all be rendered in the more general language of "signs." Accordingly, spoken, written, or conceptually explicit language (i.e., intension) is no longer the standard by which to judge semiotic activity because "verbal language is far too narrow a field from which to construct a general theory of signs, as becomes evident from the arbitrariness posited of the social world by those who suppose that it is constituted in language" (Hoopes 11). Here Hoopes is taking aim at Saussure's semiotic theory, whose focus on spoken language meant that the arbitrariness of the spoken signifier obscured all those types of signs that retain very real relations to the world—what Peirce called "indices." He says, "The meaning of the sign is not necessarily arbitrary but may be as logical as the thought that interprets it." He continues, "Signs may be arbitrary, but they can also be extremely logical. It all depends on the way in which interpretants are structured. If an interpretant predicates a real relation, then the sign is no longer wholly arbitrary. Furthermore, even arbitrary symbols—such as a stop sign—are not interpreted in arbitrary ways" (Hoopes 12)—one does so at their peril. This triadic relation avoids essentializing the sign and in effect keeps it within the paradigm of a dynamic systems critique.[72]

My argument is that Peirce's theory neatly describes the practical ways in which language is generated in the organism's cognitive lifeworld, and furthermore perfectly accords with the picture of structural coupling discussed earlier. Peirce's conception of the triadic sign returns our understanding of language and "representation" to those particular (diachronically embedded) speech acts which Saussure felt compelled to set aside in the name of scientific objectivity. Enactivism argues that m*aking sense* (informational

---

[72] There is also significant overlap between teleosemantics and Peirce's semiotics. For teleosemantics, "representations" are states that have a biological function to guide behavior in a particular way given the proper set of conditions (Papineau 97). According to teleosemantics, it is the consumers of a representation (not its producers) that determine a representations truth conditional content (Hutto 57, Papineau 97). Papineau continues, "Representation arises whenever a consumer *interprets* some state as signifying some circumstance, in the sense that it *acts* in a way appropriate to that circumstance" (98). One particularly lucid example that Papineau provides us is the different types of alarm calls produced by Vervet monkeys—one for eagles, one for leopards, and one for snakes—three predators which pose a direct threat to the monkeys in their environment. If "The truth condition of the signal depends on how the consumers behave in response to it, not on what causes producers to emit it" (98), then it does not matter if the stimulus that produced the call is an eagle, a buzzard, or a low flying airplane, all that matters is whether the consumers of the call respond in the appropriate manner (e.g., by looking up and receding into a tree). In this example, the representation is embodied in the distinct vocal call produced when a threat is perceived—it is the state (arrangements of parts) necessary to produce the representation's truth-conditional contents (flying-threat, walking threat, slithering threat) and whose biological function is to preserve the monkeys from harm. In the circuit of production and consumption, if this state elicits a clear set of truth-conditions in which this state obtains, then according to teleosemantics, this circuit can be understood as possessing a semantic *content*. Papineau says, "So here we have a simple explanation of representation. It isn't magic. It is just a matter of certain states having the biological function of instigating behavior that is appropriate to such-and-such conditions" (Papineau 98). Clearly this teleosemantic analysis is very amenable to Peirce's triad: The monkey's vocal call is the *sign* (i.e., something that stands for or represents) an *object*, in this case a bird of prey, and the *interpretant* is the behavior issued in response to this sign: run and hide.

sensitivity and knowing-how) is prior to *making meaning* (semantic content and knowing-that). Peirce's semiotics and its view of *thought as action* and *meaning as behavioral significance* (or, as we will shortly see, *habit*) is the ideal theory for connecting the embodied cognitive claims of enactivism to the representational aspects of language.[73]

The most salient feature of Peirce's semiotics for our purposes is the foundational role of *habit*. In his essay "How to Make Our Ideas Clear," Peirce describes habits as follows, "…the whole function of thought is to produce habits of action; and that whatever there is connected with a thought, but irrelevant to its purpose, is an accretion to it, but no part of it" (Peirce 168). In this passage Peirce does not mince words. Thought is subordinated to action and to the creation of habits of action. He says that anything of an ancillary nature that is dragged along with thought—that is, anything which does not ultimately serve this larger purpose of inculcating habits of action is merely an *accretion*. He continues:

> If there be a unity among our sensations which has no reference to how we shall act on a given occasion, as when we listen to a piece of music, why we do not call that thinking. To develop its meaning, we have, therefore, simply to determine what habits it produces, for what a thing means is simply what habits it involves. (Peirce 168)

---

[73] As Hoopes tells us, this approach can also help us transcend some of the disciplinary divides between the hard and human sciences: "Peirce's model of thought as feeling or action corrects the assumption of some "hard" social scientists that thinking is not as "real" and potentially influential as other, seemingly more tangible, elements in human life and society. Similarly, it corrects the view of some humanists that thinking is an utterly free process, unfettered from the relations in the natural world that constrain other kinds of human activity" (Hoopes 9). In this way, it also seems well-suited to the interdisciplinary goals of the environmental humanities.

This is an interesting example. Listening to music is certainly an aesthetic pleasure that we, as listeners, undoubtedly unify into an auditory perception. But the claim here is that such perception is not yet called "thought" or "thinking" until we employ it in some kind of habit of action. I suppose what he means here is that we do not technically begin "thinking" about (or we could say, actively perceiving) this raw aesthetic pleasure until we start to mobilize these perceptions for the sake of discerning the music's formal characteristics—things like melody, harmony, or rhythm. This would be the moment in which the perception takes on meaning. In this example we can see how, in order to acquire "meaning," audition needs to be put to the immediate use of forming these perceptions into a unified whole. It needs to get *enacted* in a larger set of purposes, contexts, and intentions that constitute the musical meaning. Peirce concludes by saying, "Thus, we come down to what is tangible and practical, as the root of every real distinction of thought, no matter how subtile it may be; and there is no distinction of meaning so fine as to consist in anything but a possible difference of practice" (Peirce 168). I think the perception of musical form and pattern is a good example of just how subtle these distinctions of meanings can be in our cognitive life.

As Legg and Black tell us:

> For Peirce, habit is the ur-ingredient of mental life…This constitutes a naturalistic approach, since habits are (i) open to empirical study (observable, not merely introspectable), (ii) shared by humans and all livings organisms, down to the most simple single-celled creatures, (iii) already subject to fundamental empirical work in biology. (Legg and Black 6-7)

And I would add, subject to empirical work in neuroscience as well, in which *habituation* is widely acknowledged to be the cornerstone of learning, on both the neuronal (e.g., Hebbian learning) and behavioral scale.

Fortunately, the work of translating Peirce's habit-based theory of signs into the language of systems theory is already at work in Terrence Deacon's emergent dynamics. In this context, he makes the case for why constraint can also be understood as a type of habit. He says:

> Translating this [constraint] into Peircian terms, if not all states of a process are realized, or if there is a bias in the probability of their occurrence, there is a habit…A habit can thus be described as the expression of a constraint. It is an attractor disposition—whether dynamically or statistically generated—that can arise irrespective of any outside mental assessment. (Deacon 202)

As Deacon's analysis has shown us, we can conceive of the cognitive dynamics of autopoiesis in more rigorously scientific terms as the preservation of constitutive constraints. Therefore, for an autopoietic (or autogenic) system, its suite of homeostatic processes (e.g., autocatalysis, self-assembly of macromolecules that aid in containment) can be understood as the cognitive "habits" that keep the system in teleodynamic organization. We can easily scale this notion of cognitive habit to the level of nervous systems via language, propositional knowledge, or the perception of redundant formal features in our environment (e.g., types and tokens). Deacon suggests we think of habit in the following way:

> The general properties by which we assess the similarities and differences among the diverse phenomena of our experience (e.g., pattern and order) are a consequence of the interaction between habits of our brains and habits of the world. So, putting this in constraint terminology, we can dispense with the notion that habit, pattern, and organization presuppose mental representation and comparison. The causal efficacy of these redundant features of the world is presupposed in the very notion of a mental category, not the other way around. (Deacon 202)

This passage demonstrates how constraint and habit can be used to naturalize what our linguistically-organized cognition codes as general types, universals, or concepts. Here Deacon argues that such mind-dependent conceptions of pattern and order can be grounded in this notion of habit/constraint—general tendencies of dynamical interaction that are preserved and transformed in ongoing relationship with one another. In this way, Deacon's articulation of constraints-cum-mental habits also strikes me as a reasonable description of naturalized *content*. It lays the groundwork for general types or classes of things, under which we can pick out specific instances or tokens of that type, and contrary to what Hutto and Myin argue in their radical enactivism, this kind of pragmatism would avoid begging the "content" question. As Deacon shows us, Peirce's habit-based theory of signs is capable of moving between enormous differences in scale—from the minute thermodynamic processes at work in the autogen to the ways in which neural assemblies habitually respond to the perception of order or pattern in the perception of a musical melody.

So how can we connect this habit-based theory of signs to the realm of spoken language? In the following chapter, I will look for these embodied and habit-based cognitive processes in literary narratives—arguably the most complex linguistic structures humans produce. But we can foreshadow some of that work here in more simple and granular terms.

A good example of connecting language and habit can be found in Andy Clark's 2008 work of extended cognition *Supersizing the Mind*. There Clark asks us to conceive of language as a "mind-transforming cognitive scaffolding: a persisting though never stationary, symbolic edifice" (Clark 44). The notion of a scaffold implicitly shifts the locus of agency back onto individual speakers. A scaffold is a temporary structure that supports a building process. It is not an end in itself but a means to an end. Conceived as a scaffold, language becomes a tool in a broader cognitive toolbox that is constructed for the purpose of realizing some concrete end—and not as ideology critique would have us believe—as the totalizing arena of subject formation.

Throughout *Supersizing the Mind* the ends that Clark has in view are pragmatic cognitive operations. His goal is to show that mental life is quite literally extended beyond the brain and distributed throughout one's environmental milieu. In this way, his approach very much in line with Peircian pragmatism. He is not concerned with truth, but with efficacy. By situating language as a cognitive scaffold he effectively sidesteps a whole swath of analytic concerns regarding the content of propositional attitudes, or cognitivist concerns that would situate these propositional attitudes as an internal formal process or syntax. Instead, the scaffold draws our attention to how thought is largely an embodied process that achieves pragmatic ends by establishing habits of mind.

Clark highlights three facets of this linguistic scaffold, each with their own pragmatic upshot. The first is our use of *labels*, which enables new "computational opportunities." By this he means that labeling objects enables the discovery of "increasingly abstract patterns in nature" (44). Labels reduce the cognitive (or if you like, computational) burden of the brain as it negotiates acts of perception. Once established, labels act like a new perceptible object that can help fix one's attention on those aspects of the environment most amenable to a given purpose or task. Ultimately, this allows one to more readily perceive new patterns and relationships in one's environment.

The second feature of the linguistic scaffold is structured *sentences*, which permit "otherwise unattainable expertise" (44). Whereas a label provides a focal point that can aid in conceptual grouping and pattern recognition, a structured sentence enables a dynamic and recursive process of behavioral revision and refinement. By enabling a diachronic engagement with sensorimotor skills, sentences allow for better self-control, and ultimately, expertise. Finally, our most complex linguistic structures allow for metacognitive reflection, enabling one to "control and guide the shape and contents of our own thinking" (44). This is clearly the realm in which literature operates.

All three levels mobilize the syntactic and semantic features of language for the sake of increasingly complex modes of self-control over one's actions and behaviors. Clark's linguistic scaffold is one way of conceiving of language in terms *habits*—in how the linguistically mediated process of thought can be instrumentalized for the sake of action at different scales of complexity. I also think that Clark is a good example of how 4E theory can help redefine the virtual and disembodied role of language that the humanities inherited from structuralist linguistics by situating it in a larger ecological and embodied milieu.

Conclusion:

Understanding language and representation in terms of embodied processes of habit is a practical way of setting aside ideology critique. In this chapter we have seen how the holistic, systems-based tendencies at the heart of post/structuralism made it a powerful framework for philosophy and literature and their revisioning of subjectivity. Yet we have also seen how its prioritizing of the autonomy of signifying system over the autonomy of speakers made it all too easy for post/structuralist analysis to become infatuated with its own navel-gazing systematicity. With the help of theorists like Peirce, Clark, Deacon, and Thompson, we can ground our linguistic faculties in embodied, ecologically embedded, and case-specific contexts.

Furthermore, this return to case-specificity strikes me as the precondition for all ethical experience. It is exceedingly difficult to see *individuals* (towards whom we have a concrete ethical obligation) when we are confronted instead by sprawling systems that conceal their individuality. One example that always comes quickest to my mind is our relationship to food and to factory-farmed meat. Therefore, if we ever hope to bring an ethical optics and its concomitant moral sensibilities to bear on these vast, and interconnected assemblages—whether it be factory farming, specific ecological communities (such as the Puget Sound watershed from which I write these words), or a socially constructed political category—then we must first learn to cultivate this optics in our relationships with individuals, and this always occurs within the confines of specific purposes, audiences, and contexts that *affect* us in our very sense of embodied self.

In the final chapter we will return to the realm of neuroscience and examine how 4E approaches are being integrated into contemporary literary studies. We will also more closely examine Alva Noë's enactivist account of perception. I will use his sensorimotor approach as a means of concretizing the concept of the "literary gestalt"—developed in the hermeneutic and reader-response traditions—on embodied cognitive grounds.

Chapter Six:

Embodied Hermeneutics: Towards a 4E Literary Formalism

> We know a world is an organism, not a machine. We also know that a genuinely created world must be independent of its creator; a planned world (a world that fully reveals its planning) is a dead world. It is only when our characters and events begin to disobey us that they begin to live.
>
> —John Fowles, *The French Lieutenant's Woman*

Enactivism and Literature:

As we ascend the scale of systemic complexity—from the autogen in chapter three, to bacteria in chapter four, to the context-specific interpretants that concluded chapter five—we arrive firmly at the level of human cognition once again. This chapter asks: how can we reimagine literature as a kind of know-how?

Reading literature is often viewed as a highly intellectualized, pictorial, and disembodied act. We imagine narrative worlds from words while seated, contemplative, and inert, often with the assumption that immersion into these narrative spaces is predicated on the reader's ability to conjure up the necessary mental imagery and to construct a vivid visual scene in the mind's eye. My goal is to disabuse us of this pictorial and imagistic conception of literary engagement and demonstrate how literature depends on the kind of sensorimotor know-how that has characterized cognition at every step of this dissertation. Specifically, I will focus on the notion of literary gestalts and argue that we can reconceive of this visual aesthetic in terms of contemporary 4E theory.

Enactivism's integration of phenomenological theories and methods overlaps with several well-established veins of literary theory and criticism—most notably, hermeneutics, reader-response, and cognitive narratology—all of which hold that the reader is the ultimate arbiter of a text's meaning because they endow the work with the sense, significations, and affective associations of their lived experience, in effect animating an otherwise static object. The enactivist approach adds to and improves upon these methods by grounding their theorizing of self, world, and the literary object in concrete perceptual, embodied, and sensorimotor processes.

Phenomenology Redux: Embodied Hermeneutics

Before we can turn our full attention to the aesthetics of literature, I want to trace the origins of this hermeneutical and phenomenological tradition and its embodied and context-specific (i.e., pragmatic) underpinnings.

Heidegger's most famous work *Being and Time* is a phenomenological analysis of Dasein, (literally in German, *being there*), a concept which effectively stands for human being. The expression *Dasein* captures something fundamental about phenomenological analysis more generally. The *Da* (there) of being (sein) signifies that human being is irrevocably bound to a world—what Heidegger famously called our *being-in-the world*. For Heidegger, Dasein is a special subject of phenomenological analysis because it is an entity for whom its own being is a fundamental concern. He says, "It is peculiar to this entity that with and through its Being, this Being is disclosed to it. *Understanding of Being is itself a definite characteristic of Dasein's Being*" (Heidegger 32). Dasein is unique because it adopts a relationship of understanding towards its own being. This leads Heidegger to declare that "the phenomenology of Dasein is a *hermeneutic* in the primordial signification of this word, where it designates this business of interpreting" (62). Dasein's being is constituted by acts of interpretation designed to render its own Being salient. Thus, what Husserl conceived of in terms of *consciousness*—a structure of openness towards the world—Heidegger recasts in even starker existential and interpretive (i.e., hermeneutical) terms.

For both Husserl and Heidegger, all acts of understanding require interpretation. We perceive individual things as discrete and meaningful because they are projected against a backdrop of intentions and possibilities. This is known as the "as-structure" of conscious experience. As Heidegger tells us, "The 'as' makes up the structure of the explicitness of something that is understood. It constitutes the interpretation" (Heidegger 189). He continues, "Whenever we see with this kind of sight, we already do so understandingly and interpretively. In the mere encountering of something, it is understood in terms of a totality of involvements…" (189). Thus, when we come to understand a discrete object in the world "*as* a table, a door, a carriage, or a bridge…" (189) we do so with an implicit understanding of the actions, intentions, or as Heidegger says, "the totality of involvements" in which these objects are intertwined. Objects appear to us as discrete, meaningful things because they are bound up in this as-structure that animates them in a space of possible actions, behaviors, uses, and intentions. This as-structure is what the language of classical hermeneutics renders in terms of parts and wholes (Grondin 399), with "parts" representing the individual objects of perception, and "wholes" the projected backdrop that enables the parts to become salient. Accordingly, the as-structure represents a recursive movement between part and whole in a schema that became known as *the hermeneutic circle*. An important consequence of this as-structure is that:

> In interpreting, we do not, so to speak, throw a 'signification' over some naked thing which is present-at-hand, we do not stick a value on it; but when something within-the-world is encountered as such, the thing in question already has an involvement which is disclosed in our understanding of the world, and this involvement is one which gets laid out by the interpretation. (Heidegger 191)

In other words, the things we encounter in the world are not disinterestedly awaiting a signification. We do not encounter a "naked" thing, devoid of meaning as though it were a blank canvas—a tabula rasa. Rather, every act of interpretation is *already* grounded in a set of involvements with the world.

Heidegger further develops this hermeneutical mode of being in terms of fore-structures, specifically, "Vorhabe" (fore-having), "Vorsicht" (fore-sight), and "Vorgriff" meaning (fore-grasp or fore-conception) (Heidegger 191n2). Crucially, these fore-structures are determined foremost by material facts and uses; by actions, behaviors, and intentions that are directly grounded in in the *body's* relationship to world. We can see this embodied dimension in several of Heidegger's most important neologisms, for example the concept "ready-to-hand." Objects in the world become ready-to-hand when Dasein encounters them and puts them to some immediate and situated use: the classic example of this in

Being and Time is a hammer. Readiness-to-hand constitutes Dasein's ordinary and everyday relationship with the world—our default setting if you like—because we habitually encounter people, places, and things via these interpretive fore-structures that are grounded in immediate uses.[74] The very verbiage of these neologisms underscores the embodied dimension to Heidegger's thinking: the world is quite literally *at hand*. If Dasein's mode of being is foremost hermeneutic, then this interpretive process and the meanings that flow from it are intimately tied to material interactions with objects and actors, and by extension, with the body.

The key insight here is that all processes of understanding, beginning with perception, are bound up in a set of involvements and interpretive fore-structures, and while interpretative acts *depend* on these involvements and the presuppositions that go with them, interpretation is also capable of making these various types of fore-knowledge explicit and thus open to further phenomenological scrutiny. As Jean Grondin tells us:

> Hermeneutical thinkers like Heidegger, Bultmann, Ricœur, and Gadamer view the hermeneutical circle more favorably since it constitutes for them an inescapable and positive element of understanding: as finite and historical beings, we understand because we are guided by anticipations, expectations, and questions. For them, the

[74] Heidegger distinguishes the ready-to-hand from the "present-at-hand," which is the contemplation of a thing once it has been removed—insofar as it is possible—from all such immediate contexts. Accordingly, "present-at-hand" represents an abstract and derivative mode of being because it takes a lot of mental gymnastics to contemplate a thing removed from its myriad fore-structures.

> key is not to escape the hermeneutical circle, but, following Heidegger's famous phrase, to enter into it in the right way. (Grondin 399)

But by the time we get to Hans Georg Gadamer's *Truth and Method* a few decades later, an important shift in this circular hermeneutical process has occurred. Gadamer says that Heidegger entered into historical hermeneutics "only in order to explicate the fore-structures of understanding for the purposes of ontology. Our question, by contrast, is how hermeneutics, once freed from the ontological obstructions of the scientific concept of objectivity, can do justice to the historicity of understanding" (Gadamer 268). Heidegger's project wanted to arrive at a pre-linguistic, phenomenologically derived meaning (or sense) of Being. His hermeneutics was always in the service of ontological inquiry and thus cleaved as close as possible to objects and actions without necessarily needing to interrogate their social construction/determination. What mattered for Heidegger was disrupting the transparency of these objects and their interpretive fore-structures in order to arrive at a more full and authentic sense of Dasein's Being. This is what Gadamer means by "the scientific concept of objectivity." It is a phenomenology of things, feelings (e.g., authenticity) and acts.

In contrast, Gadamer focuses on communication and socially constructed meaning, and this shift is reflected in the language used to describe the process: the embodied and immediate contexts that characterize "fore-structures" have been replaced by "horizons," which signals a move away from the immediacy of the ready-to-hand and towards the socially constructed and historically inherited dimensions to interpretation. It is a cultural rather than scientific process. Gadamer examines the transhistorical dimensions to interpretation—the shared repository of interpretive practices that become embedded in a culture's collective sense of "tradition," and in how tradition mediates and legitimates interpretation. Thus, for Gadamer, hermeneutics is primarily an epistemological, not ontological or phenomenological undertaking. Like Heidegger's various fore-structures, "horizons" of meaning are established via expectations, anticipations, and questions—a kind of projected whole—upon which the individual parts (in this case, texts) can reveal themselves as intelligible and discrete. But these are expectations, anticipations, and questions that have been conditioned by a shared, sociohistorical consciousness.

In addition to the language of "horizons" Gadamer's hermeneutics stress the importance of "prejudice," which is precisely the kind of interpretive fore-structure one acquires through tradition. He says, "If we want to do justice to man's finite, historical mode of being, it is necessary to fundamentally rehabilitate the concept of prejudice and acknowledge the fact that there are legitimate prejudices" (Gadamer 278).[75] The point I want to emphasize here is the increasingly *disembodied* nature of the discourse around

---

[75] In the chapter "The Elevation of the Historicity of Understanding" Gadamer alights on the notion of "prejudice" as his primary point of intervention precisely because its modern connotation has been so thoroughly determined by Enlightenment thought. He says, "And there is one prejudice of the Enlightenment that defines its essence: the fundamental prejudice of the Enlightenment is the prejudice against prejudice itself, which denies tradition its power" (Gadamer 272-273). Enlightenment epistemologies were united in their belief in the power of abstract and universal reason as the ultimate arbiter of meaning and truth. According to Gadamer, it is this *unique Enlightenment prejudice* in favor of impersonal, universal, and scientific knowledge that informs our contemporary understanding of the term "prejudice" as an inherently bad thing. As offspring of the Enlightenment tradition, we have been trained to forego making determinations about the world without prior, systematic investigation. Therefore, by resuscitating and elevating the notion of "prejudice" in his argument, Gadamer hopes to (a) make manifest the fact that the Enlightenment itself is nothing other than an historically determined tradition; (b) divest us of this Enlightenment prejudice towards prejudice; and (c) as a consequence of the aforementioned, once again open our understanding to the foundational role that sociohistorical horizons and pre-judgements have in determining our interpretations.

hermeneutics. Like "horizon," the language of "prejudice" demonstrates a turn away from material and embodied contexts. A prejudice is a pre-*judgement*, which grounds the interpretive act in logical and propositional content over and above embodiment.

Wolfgang Iser is another important thinker in this genealogy of hermeneutics, and someone who helped bring this tradition more firmly into the realm of literature and aesthetics. Iser, along with Hans-Robert Jauss, founded the Konstanz School of reader-response theory, which built on the hermeneutical methods established by Heidegger and Gadamer. Beginning in the late 1960s and early 1970s, reader-response theory quickly became a dominant mode of literary interpretation, particularly in Germany and the U.S., in part because it offered a much more intuitive and matter-of-fact approach to interpretation than the poststructuralism coming out of France.

In "The Reading Process," one of Iser's most well-known essays, we see another important step in the evolution. Unlike Gadamer, who focuses on the fore-structures (i.e., horizons) inaugurated by "tradition," Iser zooms down into the microscale of individual perception, evidenced in his appropriation of the concept "gestalt." But as we will see, this shift into the language of gestalts is a double-edged sword. On the one hand, it directly connects reading and interpretation to systems theory, cognitive science, and perception; on the other hand, this inheritance is intimately tied up with *visual* modes of perception. My contention is that this conceptual transposition established a representational, i.e., pictorial (or as Noë would say, snapshot-esque) view of the reading process, to the detriment of other embodied ways of thinking about interpretive acts.

The basic argument in Iser's essay has become a mainstay of reader-response theory and so-called "configurative" reading practices. In brief, it holds that the meaning of a work is a virtual (i.e., relational) phenomenon that arises from the transaction between text and reader (Iser 279). He says:

> The literary text activates our own faculties, enabling us to recreate the world it presents. The product of this creative activity is what we might call the virtual dimension of the text, which endows it with its reality. This virtual dimension is not the text itself, nor is it the imagination of the reader: it is the coming together of text and imagination. (Iser 284)

It holds that the text is not merely the haphazard interpretation of a reader, nor is it reducible to its objective features, i.e., the "words on the page." Instead, a text must be understood as a product that exists between these two centers of intentionality—conditioned by the words on the page but realized in the mind of the percipient. It is a transactional, dialectical, and configurative process *enacted* between text and reader.

Like Gadamer—who argued that in every interpretive act we are "pulled up short by the text" (Gadamer 270)—Iser contends that there are "gaps" that emerge in this process of configuration, between the expectations one projects onto the text and the connections we are ultimately able to construct from it (Iser 285).[76] For Iser, literary texts are characterized

---

[76] For example, in the last half of his essay, Iser alights on the "illusion-forming and illusion-breaking that makes reading essentially a recreative process" (Iser 294). This is the basic argument we've already seen outlined by Paul Armstrong in *Stories and the Brain*, which he codes in terms of "concordant discordance" (Armstrong 12). The building up of illusions and expectations in the process of reading is what permits immersion in, and identification with, the text (concordance). Whereas the breaking up of these illusions can lead to the defamiliarizing effects long celebrated in modernist and postmodernist literature (discordance).

by a preponderance of these gaps. Literature tries to multiply and confound expectations and, in this way, enrich the imaginative processes that give rise to one's realization of the text. In contrast, expository texts are characterized by the straightforward manner in which they establish expectations and then fulfill them, thereby instructing readers in their controlling ideas and themes. Iser says, "These gaps have a different effect on the process of anticipation and retrospection, and thus on the 'gestalt' of the virtual dimension, for they may be filled in different ways" (Iser 285). For Iser, the imagination is the engine that drives the creation and realization of the text, and it does this by forming *gestalts*.

A gestalt is an irreducible perceptual pattern (Capra and Luisi 66), one whose properties cannot be explained by analyzing or decomposing it to its underlying parts. It was first used by Austrian philosopher Christian von Ehrenfels in his 1890 essay "On Gestalt Qualities." Here, Ehrenfels asserts the existence of gestalts, "based on the observation that humans can recognize two melodies as identical even when no two corresponding notes in them have the same frequency" (Wagemans et al. 1175). In the case of musical melodies, the essential property is the relationships of difference—the musical intervals—between different pitches, and not their precise frequency as measured in Herz. The degrees of difference in a musical melody remain the same regardless of whether it is rendered in an A scale or a G scale, for instance. Ehrenfels is often credited with introducing the dictum: *the whole is greater than the sum of its parts*, "a statement that would become a key formula for systems thinkers later on" (Capra and Luisi 66).

Before long Ehrenfels' "gestalt" was given more rigorous empirical validity. In 1912, German psychologist Max Wertheimer discovered the phi phenomenon which was a special case of apparent motion (Wagemans et al. 1173). As Wagemans et al. tell us, "According to the conventional view of apparent motion, we see an object at several successive positions and motion is then added subjectively" (1173). In contrast:

> The phi phenomenon was the perception of a pure process, a transition that could not be composed from more primitive percepts of a single object at two locations. In other words, perceived motion was not added subjectively after the sensory registration of two spatiotemporal events but had its own phenomenological characteristics and ontological status. From this phenomenon, Wertheimer concluded that structured wholes or Gestalten, rather than sensations, are the primary units of mental life. (Wagemans et al. 1173)

Perhaps the most pervasive example of phi motion is film. In moving pictures, there is no singular object that occupies different spatiotemporal locations. Instead, you have a series of stills that when flashed at the appropriate speed creates the illusion of continuous movement.

Wertheimer, along with his colleagues Wolfgang Köhler and Kurt Koffka described a number of gestalt principles, which they dubbed "laws". The most foundational was the *law of prägnanz*: when given a complex array of stimuli our brains will render them as simply as possible. There is the *law of similarity*: things that are similar are grouped together. The *law of proximity*: things that are close appear related. The *law of closure*: our brains fill in what is lacking (e.g., the Kanizsa triangle). The list goes on. The early history of gestalt theory is complex and fascinating, but beyond the remit of this chapter.[77] Nonetheless there are a few strands I want to emphasize here that will help clarify the appropriation of "gestalt" by hermeneutical thinkers like Iser, and that will also help us to begin forging connections between hermeneutics and the enactivist theory of perception.

First and foremost is what Iser calls the text's *virtual* dimension, which is the primacy of the relationship between parts, and not the parts themselves. Like a melody, which is not a material thing but an arrangement of musical intervals, one's perception of a text is the product of a relationship, specifically that between the text created by the author (what Iser calls the artistic pole) and the "realization accomplished by the reader" (the aesthetic pole) (Iser 279). The wholeness we perceive in the text emerges through this relationship—it cannot be reduced to any one of these parts. It is an emergent process.

Iser's discussion of narrative gestalts frequently relies on imagistic and representationalist metaphors and analogies. He says, "with a literary text we can only picture things which are not there; the written part of the text gives us the knowledge, but it is the unwritten part that gives us the opportunity to picture things; indeed without the elements of indeterminacy, the gaps in the text, we should not be able to use our imagination" (Iser 288). In this passage we clearly see how the imagination is rendered

[77] This research program continued up until the time of the cognitive turn and Wagemans et al. offer two reasons for the drop-off in gestalt research. The first was pragmatic: after immigrating to the U.S., these early gestalt psychologists stopped training graduate students, and the result was that there were not enough people to carry the torch. The second reason should, by now, sound familiar. They say, "In the 1950s and 1960s, its critics increasingly insisted on causal explanations, by which they meant cognitive operations in the mind that could be modeled as computer algorithms or neural mechanisms that could be attributed to the properties of single cells that were discovered by Hubel and Wiesel in that period" (Wagemans et al. 1179). Gestalt psychology, like Skinner's behaviorism, succumbed to the computationalist pressures of the cognitive turn.

foremost as a pictorial process: when we imagine things we *picture* them. While Iser goes on to say that "the 'picturing' that is done by our imagination is only one of the activities through which we form the 'gestalt' of a literary text" (Iser 288), referring us once again to things like anticipation and retrospection, this does little to substantively alter or nuance the idea of a "gestalt" as a pictorial, image-making, and "illusion-building" process (Iser 289). The language of anticipation and retrospection speaks to an affective dimension to the literary gestalt, but these ideas return us to the realm of formalist aesthetics only in the most general way.

To his credit, Iser avoids the term "meaning," an omission (with one exception) that reflects a commitment to the notion of perception as such, and not to language or representation. Despite this evolution in thinking about the literary text, much of the concrete, embodied nuance that informed the notion of "fore-structure" in Heidegger has been lost. While reader-response theory makes several conscious strides away from meaning as a uniquely linguistic process tied to tradition, its appropriation of "gestalt" from the realm of visual perception is redolent of the "snapshot conception" that Noë argues is so antithetical to enactivism.[78] Gestalt psychology and its associated laws were developed almost exclusively in the realm of visual perception, and reader-response theory and hermeneutics did little to nuance their appropriation of this term or wrest it from these visual underpinnings.

[78] See chapter one, subsection "Emergence of the Enactive Approach…"

In his 2020 book *Stories and the Brain*, Paul Armstrong updates this reader-response and hermeneutic tradition in the context of contemporary neuroscience. He begins by acknowledging many of the hermeneutic practices that evolved from the tradition of phenomenology: from the early work of Martin Heidegger, Roman Ingarden, and Hans-Georg Gadamer, to Wolfgang Iser and Hans Robert Jauss of the Konstanz School. But one of the most important intertexts throughout *Stories and the Brain* is Paul Ricoeur and his three-volume work *Time and Narrative*. A primary point of departure for Armstrong is Ricoeur's revisioning of mimesis as a dynamic process of narrative configuration.

Mimesis means imitation. As Aristotle tells us in *Poetics*, the essence of narratives is their plots, and a plot is the representation (or "imitation") of an action. Poetic mimesis (for all intents and purposes, *literature*) takes the manifold actions of lived experience and emplots them: shapes them into coherent beginnings, middles, and ends; connects them according to the laws of probability and necessity into unified wholes. This is a process that, in Aristotle's view, is meant to render them more intelligible. He says, "poetry is a more philosophical and serious business than history; for poetry speaks more of universals, history of particulars. 'Universal' in this case is what kind of person is likely to do or say certain kinds of things, according to probability or necessity..." (Poetics, 51b5-10). Because poetic mimesis represents universals—that is, general kinds of action—it can better depict the essence of the action. Thus, when properly executed, mimesis can and should add to the percipient's understanding of an action because it goes beyond our everyday perception of the world and its particularities. In stark contrast to our modern conception of art and artists, Aristotle's poet does not *add* to the represented actions in an ontological sense by inventing or imagining characteristics and qualities *ex nihilo*—he ascribes only the essential properties and qualities that are proper to the action. Thus, the poet's *techne* (art, skill) lies primarily in his capacity to arrange and structure what is already before him in the world, and in this way distill the essence of the action from the empirical reality.

Ricoeur's revision involves breaking this "orthodox" view of mimesis into three distinct aspects, which are then analyzed as part of a dynamic system of narrative experience. It assumes the imitative and representational impulses at the heart of this orthodoxy, but it jettisons the essentialist epistemology for a social constructionist position, and it rejects the taxonomizing impulse of narrative "kinds" (e.g., comedy, tragedy, epic), focusing instead on the subjective experience of readers.

Armstrong summarizes mimesis$_1$ as the use of "prefigurative patterns of experience" (Armstrong 29). Here *prefigure* implies that the experiences and actions that constitute plots are ready-to-hand and exist prior to their emplotment. Readers are acculturated by pre-existent patterns of meaning and experience, these are the acquired social and societal conventions that shape one's understanding of action and behavior (e.g., growing up, marriage and domesticity, manners etc.). "Prefigurative patterns" refers to this repository of social and cultural knowledge that is acquired during the process of subject formation. In keeping with the phenomenological approach, we can understand these prefigurative patterns as the projected "wholes" or "horizons of possibility" that constitute the "as-structure" of conscious experience. The *prefiguration* of mimesis$_1$ refers not just to objects, but to patterns of action and behavior, and to what Heidegger would call the "totality of their involvements." These are the patterns that constitute our everyday experiences of being-in-the-world, prior to entering into a textually mediated space. Ricoeur's conception of $mimesis_1$ is grounded in this basic, phenomenological and hermeneutic as-structure of experience.

$Mimesis_2$ is "the configurative work of creating stories by fashioning incidents and events from life into plots" (Armstrong 28). *Configuration* is the act of shaping and ordering these received patterns into other figures with the help of the text—hence *con*-figuration (figuring with). In the configurative processes of $mimesis_2$, mimetic activity is actualized in the consciousness of a reader, as a product of his or her agency and interpretive acts, but it is also mediated by the patterns held ready by the text. In contrast to the experiential (and as Heidegger would say, existential) dimension of the as-structure in $mimesis_1$, Armstrong calls $mimesis_2$ to "as-if" in order to signal that the basic as-structure of $mimesis_1$ is now mediated by the text (Armstrong 141), a mediation that places it at a remove from the everyday experiential forms of interpretation. The "as-if" also signals that texts are speculative places where readers can imagine new possibilities and explore patterns of thought and action normally unavailable to them in their everyday life.

Finally, $mimesis_3$ refers to "the potentially transfigurative experience of comprehending the narrative (Armstrong 29). It is the potential for the narrative to "bring about a reshaping of the audience's emotional, embodied, and culturally mediated sense of its world" (Armstrong 29). The difference between $mimesis_2$ and $mimesis_3$ is slight, but significant. Figuration ($mimesis_1$) is a pervasive feature of cognition. It enables us to structure the self and the world around us. Configuration (mimesis$_2$) is the basic dynamic of mimesis$_1$ but with the unique addition of textual mediation. Transfiguration (mimesis$_3$) is an intensification and elevation of the process of mimesis$_2$, such that this as-if becomes the grounds for new experience as such, one that transcends textual mediation. It refers to the potential for narratives to drastically alter our perception and knowledge of the world. Armstrong will later refer to the effects of mimesis$_3$ as "quasi-experiences" (141), a designation which is meant to signal affinity with, but distinction from, the ordinary, everyday experiences that constitute mimesis$_{1.}$ I believe mimesis$_3$ or transfiguration is tantamount to the Russian Formalist account of defamiliarization, and accordingly, is the place where formalism and aesthetics most clearly announce themselves.

Ultimately, Armstrong wants to move beyond the language of mimesis entirely and adopt instead the language of figuration. He says:

> The language of figuration, configuration, and refiguration is preferable to the terminology of imitation, representation, and copying not only because the latter terms are heavily freighted with referential associations but also because the term 'figure' suggests the activity of constructing gestalts or patterns that is basic to embodied cognition at all levels—from neuronal assemblies to the interactions of brain, body, and world—and that is also integral to the process of telling and following stories. (Armstrong 28)

Here *figuration* functions as a general term to encompass the pre, con, and transfigurative processes of pattern acquisition, pattern making, and novel pattern creation. In this quotation, we can see the basic enactivist critique of representationalism paralleled in the discourse of hermeneutics. Like the enactivist position, the phenomenology of figuration emphasizes the dynamic processes and actions that constitute (in this case, narrative) experience, over and above reference and representation. A figure is a basic pattern or gestalt that we encounter in the world, and the means by which we actively create our own narrative patterns (configure), and which we can return to time and again to make ever more artful and defamiliarized patterns (re or transfigure). As Armstrong contends, not only do figurative approaches like Ricoeur's help move us beyond traditional notions of mimesis, but they also move beyond the static, taxonomizing impulses of structuralist narratologies that were either unable to, or uninterested in, analyzing these structures in a dynamic—i.e., temporally unfolding, recursive way (Armstrong 15, 46).

Furthermore, this passage returns us to the notion of the literary gestalt established by Iser and the Konstanz school. I have explicated Armstrong at some length because his analysis demonstrates how the hermeneutical tradition inaugurated by Heidegger and advanced by later theorists such as Gadamer, Iser and Ricoeur can be easily grafted onto a more explicit, dynamical system framework—that the phenomenological processes of configurative reading and the "as-if" of narrative engagement does not proceed by static, representational means but by dynamic enaction.

What I want to do now is compare this approach to that outlined by Noë's enactivism (dynamic sensorimotor approach). In the following section, I will return to Alva Noë's *Action in Perception* and show how this dynamic process of textual interpretation and gestalt formation can be filled out along embodied (sensorimotor) lines.

Sensorimotor Knowledge:

Noë's thesis in *Action in Perception* is that perception is an inherently active process that acquires specific content via skillful activity and embodied know-how: "We bring content to experience, by action. We *enact* content" (Noë 100). We accomplish this by building up a repository of sensorimotor knowledge based on patterns of movement and change, and how these changes structure our intuitive, kinesthetic, and proprioceptive engagement with the world. Noë calls these patterns of movement and change *sensorimotor dependencies,* and at other times *sensorimotor contingencies,* in order to emphasize how the percept is in fact dependent on these dynamic feedback loops that connect real world objects with our sensorium. Perception depends on these loops because it is movement and variation around and among objects that enables one to vary the appearance of the environmental stimulus and in this way build up these impressions into coherent perceptual frameworks over time.

Consider the problem of dimensionality reduction and our perception of volume as a case in point. How is it that we can perceive three-dimensionality (voluminousness) when the optics our eye can only record the visual scene in two-dimensions? Noë says:

> Your perceptual experience of the tomato as voluminous depends on your tacit understanding of the ways its appearance (how it looks) depends on movement. You visually experience parts of the tomato that, strictly speaking, you do not see, because you understand, implicitly, that your sensory relation to those parts is mediated by familiar patterns of sensorimotor dependence. (Noë 77)

First of all, we need to acknowledge that Noë is presenting us with a phenomenological analysis. He is not interested in pinpointing this or that neural correlate to the experience. He is attempting to explain why the percipient has a particular kind of perceptual experience (i.e., an experience *like this* as opposed to some other way). As part of this experience, we rely on certain empirical neurological facts (a two-dimensional retinal image), but our experience of the tomato cannot be exhaustively analyzed from this fact. Notice too that this sketch of voluminosity depends on what Noë calls a “tacit understanding” understanding based on “familiar patterns.” What ultimately fills out our two-dimensional profile of the tomato is the percipient’s ability to draw on a repository of “sensorimotor knowledge” (Noë 65).

Noë says, “The need for this further knowledge is clear: How can you experience a strictly unseen bit of an occluded surface as perceptually present? Your sense of its presence cannot be explained simply by reference to the fact that you receive stimulation from it. Because when it is occluded, you do not…” (Noë 65). The only thing that can explain this perceptual presence that we experience (i.e., the wholeness of the tomato) is this repository of sensorimotor knowledge. Because the percipient has built up a repertoire of embodied know-how, they implicitly understand that, were they to move this way or that in relation to the tomato that the tomato would change its profile in a predictable way. The argument is that these predictable patterns of change and variation are established piecemeal from basic sensorimotor dependencies.

Thus, sensorimotor knowledge (i.e., practical mastery of sensorimotor dependencies) is not knowledge in the propositional sense. Noë says, “It would be unsatisfying to explain this mastery by appeal to the perceiver's tacit grasp of propositions describing the sensory effects of movement" (Noë 119). Unsatisfactory for two reasons. For one, there is no mechanism for explaining the connection between the proposition and experience. In other words, propositional knowledge has no way of addressing the phenomenological experience; and two, "this seems to put the cart before the horse" (120) because "knowledge of the propositions (tacit or otherwise)…is *consequent* on the experience" (Noë 120). Thus, to state Noë's position unequivocally: "There is no sense, then, in which the enactive approach is committed to the idea that perceivers have cognitive access to the content of experience prior to their grasp of sensorimotor knowledge. Sensorimotor knowledge is basic" (Noë 120). This is the insight we have seen recapitulated at every scale of analysis: content does not preexist cognition, it is enacted through ongoing, body-dictated processes of self-constitution.

For one, we can see how this notion of sensorimotor dependencies—and by extension, sensorimotor knowledge—recapitulates well-established frameworks from cognitive neuroscience and developmental psychology, such as the "temporal contexts" outlined by Albright, which depend on "associative learning" and "memory retrieval"; as well as Jean Piaget's theory of cognitive development, which situates the "sensorimotor stage"—i.e., the stage in which infants use their incipient motor skills such as looking, sucking, grasping etc.—as the period from birth to two years of age.[79] Noë's

[79] While Piaget's well-defined progression of different cognitive stages has since been criticized as too rigid and unable to adequately account for individual variation, the basic insight that human cognitive development occurs in well-defined milestones that serve as the foundation for more advanced cognitive functions has remained a key part of cognitive science (Lones et.al. 23).

phenomenological approach is substantiated by this empirical overlap, despite differing idioms and aims.

So how can enactivism fill out or complicate some of the static, imagistic, and snapshot-esque assumptions that are baked into the conception—and appropriation—of a literary gestalt? The most important consequence of Noë's analysis of sensorimotor dependencies for the purposes of the literary gestalt is his assertion that one's perception of the tomato is *virtual*, by which he means that it is "present to perception as accessible" (63). He elaborates:

> According to the enactive approach, the far side of the tomato, the occluded portions of the cat, and the unseen environmental detail are present to perception virtually in the sense that we experience their presence because of our skill-based access to them. Phenomenological reflection on the character of perceptual presence suggests that the features are present as available, rather than as represented. (Noë 67)

He likens this virtuality to the way that information is stored and accessed on computer's desktop. Virtual access means that one does not need to have all the environmental detail consciously represented all at once, you just need effective real-time access. He says, "Just as you don't need to download, say, the entire *New York Times* to be able to read it on your desktop, so you don't need to construct a representation of all the detail of the scene in front of you to have a sense of its detailed presence" (Noë 50).[80]

This is one of the central claims of 4E more generally: why would the brain go through the trouble of creating detailed mental representations when cognitive processes can be distributed and offloaded onto the body (embodied), other objects (embedded) and throughout the environment (extended). The *embodied* approach describes our thought processes as they are shaped by the physiological hardware we possess and by the sensorimotor loops that constitute our distinct perceptual modalities. The *extended* approach expands these loops beyond the body and examines how cognitive processes become deposited in objects and tools in the environment.[81] The *embedded* approach,

---

[80] There are many experiments that validate Noë's assertion that the perceptible world is accessible, not represented. Andy Clark, who works under the banner of "extended" cognition) draws our attention to one such example. The "block copying task" is an experiment performed on visual processing in which subjects are given a model pattern of colored blocks that they are then asked to copy by "moving similar blocks from a reserve area to a new workspace" (Clark 11). Using these spare blocks, participants must recreate the model by moving one block at a time using a computer mouse to click and drag the blocks into place and was designed to pinpoint the particular problem-solving strategies employed by the subjects. The cognitivist-representationalist paradigm would predict that participants "look at the target, decide on the color and position of the next block to be added, and then execute the plan by moving a block from the reserve area" (12) because such a procedure suggests that participants are building up a detailed inner model of the scene, and then using this information to guide their actions. What the authors found was that "repeated rapid saccades (spontaneous scanning eye movements) to the model were used in the performance of the task, and many more than you might expect" (Clarke 12). I cite this example because it demonstrates (a) that "movement" in Noë's theory does not just refer to movement of the whole organism but can be extended to its most granular cases (e.g., eye saccades), and (b) that mental representations, to the extent that they are operative in such a task, cannot be divorced from the ongoing processes that link brain, body, and world. These types of experiments lend credence to Noë's theory of sensorimotor dependencies because they unsettle the *primacy* of representational models. We can see how Noë's claim that the visual scene is virtually present because it is accessible, has analogues in empirical neuroscientific research. One focuses on phenomenological experience, the other on measurable problem-solving tasks, but both effectively demonstrate that the mental processes involved extend beyond the brain and redistribute the act of cognition throughout the body and the environment.

[81] A clear expression of extended cognition can be found in "the parity principle" developed by Clark and Chalmers in their 1998 article "The Extended Mind." It reads: "If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process"

"facilitates thinking by causally exploiting an object in the environment" (Carney 77). It foregrounds the specific, material causes present in an environment and how these constrain one's cognitive processes. For example, I sit down with my friend to play a game of chess only to discover that we are missing three pawns. We can substitute any number of objects in this context provided they are manipulable, small, and identical, e.g., we could place pennies where the pawns should be, but the embedded, material aspects of the process (the rules, the board, the movements required etc.) mean that we could not use droplets of water. In *Incomplete Nature*, Deacon often invokes the idea of "boundary conditions," and this strikes me as an appropriate analogue to embedded cognition because it describes the functional limits of a given process or procedure.

Notice that all four of these approaches are tightly intertwined. The language of *enaction* highlights the temporal dimension—that the world is not passively awaiting representation but instead emerges via dynamic unfolding of the system and its environment. Yet this enaction takes place foremost through embodied means. The dynamic sensorimotor approach that constitutes Noë's enactivism highlights this synthesis of enaction and embodiment. In like manner, as soon as we grant the enacted and embodied underpinnings of cognition, environmental extension and context-dependent embedding of those processes read like straightforward logical consequences. Again, I think this helps explain my particular investment in enactivism because it strikes me as the most foundational of the 4E approaches.

(qtd. in. Clark 77). Simply put, extended approaches try to understand the consequences of this principle, of how we offload cognitive processes onto objects in the world.

The idea that perceptual content is virtual is one of the most unique claims advanced by Noë, and it speaks to the broader enactivist view of perception as a non-representational process of bringing forth self and world. It also allows us to complicate the virtuality that underpins the notion of a literary gestalt. We have seen how the literary gestalt is a virtual (relational) product of text and reader, but Noë's analysis give us a way of complicating what Iser conceives of merely on relational grounds. Rather than simply denoting a relation, the sensorimotor approach renders this virtuality as a fundamentally embodied process. It is not a space of received images, types, categories etc. because, as Noë argues, such images-types-categories are foremost the product of sensorimotor dependencies that have been built up over time via the reader's embodied agency. To be sure, cultural constructions will inevitably color and colonize this repository of perceptual knowledge to greater or lesser degrees, but we can see how, at bottom, these are processes that obtain for all similarly situated observers, regardless of their acculturation. There is nothing ideological at work in one's perception of depth, spatiality, contiguity, etc. In other words, we are presented with a cognitive space that exists on embodied, not ideological grounds, and that I would call transcendental in the Kantian sense of the word. It describes certain, universal conditions of possibility. [82]

---

[82] Capra and Luisi also emphasize these transcendental/universal aspects to cognition. They say, "The mind's embodiment can easily be illustrated by our use of spatial relations, which are among our most basic concepts. As Lakoff and Johnson (1999, pp. 34-5) explain, when we perceive a cat 'in front of' a tree, this spatial relationship does not exist objectively in the world, but is a projection from our bodily experience. We have bodies with inherent fronts and backs, and we project this distinction onto other objects. Thus, 'our bodies define a set of fundamental spatial relations that we use not only in orienting ourselves, but in perceiving the relationship of one object to another'" (Capra and Luisi 272).

Furthermore, the notion of "accessibility" at the heart of Noë's use of the term "virtual" also reflects the "as-if" that defined Armstrong's exposition of narrative configuration. When we read, we create a constructed and speculative world uncoupled from the real. One that allows us to imagine and explore—not just ideas, characters, plots etc.—but the perceptual process itself.

Interestingly, Noë makes only one connection to art throughout *Action in Perception*, and to visual art at that. He says:

> From the standpoint developed in this book, it is clear that the task of phenomenology and of experiential art, ought to be not so much to depict or represent or describe experience, but rather to catch experience in the act of making the world available. Experience is a kind of activity, an activity that acquires content, as we have seen, thanks to the perceiver's application of a kind of sensorimotor knowledge. (Noë 177)

At the end of this section Noë even advances a tentative approach to aesthetics. He says, "Most writings on the importance of art for perception focus on pictures and paintings as *objects of perception*" and on all the ways in which perception depends on pictorial representations (Noë 178). He suggests that rather than think about pictures as objects of perception, that we think about the act of making pictures as a way to illuminate all the ways we enact experience. He says, "Picture making takes up the phenomenological stance on the world. For this reason, it is to the activity of picture making that phenomenology can turn for instruction about how to do phenomenology" (Noë 179). I think Noë's commitments to visual perception throughout the book explain the rather outsized importance he places on visual art in this section. If the goal is to disabuse us of the "snapshot" view of perception, then examining how artists actively produce visual art (undoubtedly in an analogous manner to the block-copying discussed by Clark above) is yet another powerful way to demonstrate the embodied, embedded, enacted, and extended processes of perception. But the idea that we should reverse the paradigm with respect to art, i.e., approach the art object foremost as a process of creation rather than some kind received product is the very essence of reader-response theory and the tradition of hermeneutics I have been outlining.

Stated simply, if perception is fundamentally a virtual process, then I would argue that this virtuality is recapitulated in the novelistic space more than any other type of art. I think Armstrong's analysis of figuration and the as-if structure of interpretation helps us synthesize these embodied, active, and virtual threads that are at work in the making of a literary gestalt. There is nothing passive about reading literature. Even when we think of the story as conforming to genre conventions and delivering us a narrative experience that is "right down main street," so to speak, we can still see how this is active construction—not passive consumption. We first internalized the conventions of genre fiction (e.g., the detective novel) by engaging with it as an as-if space—one that we actively explored, constructed, and accessed in a piecemeal fashion, and not all at once. The sensorimotor approach would have us conceive of genre patterns like other perceptual patterns—as repositories of actions, behaviors, and familiar types which we have built up over time and as a function of embodied perception, not static representations or taxonomies.

So, while I understand the reasons behind Noë's recourse to visual art, literature strikes me as a much more rich and complex space through which to explore the phenomenology of action in perception. By exploring novels (and formally complex, experimental novels to boot), rather than an explicitly multisensory medium like film or music, we can underscore the widespread applicability and utility of embodied cognitive claims, claims that obtain in all manner of cognitive processes, even those that, on their surface, seem wholly abstract and skull-bound.

Neuroscience and Narrative: Habituation, Immersion, Defamiliarization

In a 1991 article titled, "The Narrative Construction of Reality," psychologist Jerome Bruner elaborates ten features of narratives that he believes constitute "an instrument of mind" and not just objective features of texts (Bruner 6). But I think the most interesting moments in this essay occur in his initial framing. He critiques cognitive science and its focus on "…the child's growth as 'little scientist,' 'little logician,' 'little mathematician" (Bruner 4), and asserts, "It is curious how little effort has gone into discovering how humans come to construct the social world and the things that transpire therein" (Bruner 4). He says:

> For once the 'cognitive revolution' in the human sciences brought to the fore the issue of how 'reality' is represented in the act of knowing, it became apparent that it did not suffice to equate representations with images, with propositions, with lexical networks, or even with more temporally extended vehicles such as sentences. It was perhaps a decade ago that psychologists became alive to the possibility of narrative as a form not only of representing but of constituting reality…" (Bruner 5)

Overlooking his recourse to the language of representation, the claim Bruner makes here is an important one: reality—the totality of cognitive activities—reads like a narrative. He acknowledges that this is not a particularly novel claim for narratology or literature (nor, we might add at this point, in phenomenology), but a relatively new one in the sciences (Bruner 5).

Bruner's insights impress upon us the need to distinguish between two different modes (or scales) of narrative structuration. The first is immanently cognitive and perceptual, a kind of endogenous cognitive "scaffolding" that enables all manner of meaningful behavior and intersubjective engagement. It is an ongoing suite of cognitive processes that help constitute a narrative construction of reality, rather than some kind of acquired, discrete, socially-mediated object (e.g., a book, or oral storytelling).

One of the most fundamental expressions of this can be found in the brain's Default Mode Network (DMN)—a complex of cortical areas that activates when humans are not engaged in any specific task but are merely in a state of restful awareness such as mind-wandering and daydreaming. Mind-wandering and daydreaming are mental processes that synthesize and intertwine imagination and memory: future-oriented hopes and goals, past occurrences and events, and ongoing speculation about the world and its inhabitants. Because these are all activities that we would recognize as essential to telling and following stories, it demonstrates how narrativizing is in fact a default mode of cognition (Armstrong 85). We also broached a fundamental aspect of narrative cognition at the end of chapter one with respect to neurophenomenology and time consciousness. The brain is a systems-based organ that distributes cognition in discrete regions (visual, motor, auditory cortexes etc.) and scales (neurons, neuronal assemblies, and larger functional regions). It takes time for the action potentials of neurons to fire, for assemblies of neurons to form, and for whole brain regions to synchronize and self-organize via parallel processing. The argument we encountered in that section was that these temporal requirements of the brain are directly responsible for the neurophenomenology of time. Recall also that the longest of the three temporal scales elaborated this neurophenomenology of time consciousness is known as the "narrative" scale. Regardless of whether you want to commit to that label, this analysis lays the foundation for thinking about the cognitive role of narrative objects—as tools for structuring and organizing time.[83]

The second mode of narrative structuration is constructed on these baseline cognitive processes, but it externalizes them, and in this way lays the foundation for mediating between different cognitive agents. This is narrative understood as a received, socially constructed objects such as films, novels, epic poetry etc. On this scale, narratives are the product of social and linguistic conventions (things like genre); and while narrative objects are actualized in the minds of individual percipients as a kind of inner experience, it is an experience that is also made possible by the "intentionality held ready by the narrative" (Armstrong 36). The self and the world that are enacted in narrative experience of this kind results in what Armstrong calls a "doubling of consciousness," by which he means that the intentionality the reader brings to the text remains distinct from, but

---

[83] One need look no further than the urtext of literary theory, Aristotle's *Poetics*, and his thesis that the essence of narrative is plot!

intersubjectively coupled with, the intentionality held ready by the text.[84] In other words, narrative objects are repositories for the intentionality of the author-cum-narrator, and so the enaction of narrative experience that proceeds from a literary object is not reducible solely to the consciousness of the reader. It is a joint product—a doubling of consciousness. Crucially for our purposes, Armstrong takes this phenomenological, social constructionist narratology and updates it in the light of contemporary neuroscience. He says, "Figuration, or seeing-as, is the key concept joining cognitive science and narrative theory, and it is central to the neurophenomenological model of narrative that I construct…" (Armstrong 26).

In the context of 4E cognition, narratives need to be viewed foremost as *patterns* that help structure perception, guide cognition, and shape emotion. Not only can they

[84] Armstrong says, "The experience of following a story does not collapse the differences between selves, however, but entails a doubling of my consciousness with the intentionality held ready by the narrative, thereby enacting the paradoxical combination of community and solipsism that characterizes intersubjectivity" (Armstrong 36). Putting aside for the moment this somewhat off-the-cuff description of intersubjectivity as a "combination of community and solipsism," I do think this quotation captures something true about narrative experience, what Armstrong refers to here as the "doubling of consciousness." When we engage with a narrative, we maintain our autonomy and agency—our own sense of the world as separate from the narratively constructed world we encounter in tandem with the text. This is because narratives are animated in the consciousness of readers and on the basis of their autonomy and environment. And yet, this process is also the product of another subjectivity, that which is "held ready by the narrative," i.e., the consciousness responsible for the words we encounter on the page. In this way, the self that the reader enacts in narrative experience—in concert with the intentionality of the author—is not entirely his or her own. This is what Armstrong means by the doubling of consciousness. However, I would not go so far as to call this process *paradoxical* as this implies an inherent contradiction in terms, one which does not square with sociological account of language acquisition (and I would argue, poststructuralist accounts of subject formation). According to these approaches, intersubjectivity is a constitutive feature of subjectivity and so it seems unhelpful to label something so fundamental "paradoxical." Recall too that Thompson sees intersubjectivity as an ineluctable aspect of our cognitive lives. Where enactivism differs from these accounts is in the role of bodily autonomy. One of the major upshots to the autonomy perspective of enactivism and embodied cognition is that there are aspects of bodily experience that escape ideological determination (i.e., determination by others).

represent these processes thematically, as explicit objects of thought—but they rely on this perceptual and embodied know-how in order to produce the experiences we often associate with literary works—things like escapism (the pleasure of losing oneself in a narrative space and identifying with its world), cultivating empathy and ethical sensibilities (identifying with characters both like and unlike oneself), speculation (imagining alternative realities), and altering the reader's perception of the world by defamiliarizing it, making it feel strange and new. By focusing on these affective and perceptual facets of literary engagement over and above its conceptual, representational, or referential aspects, a 4E exploration of narratives is fundamentally formalist.

But this embodied dimension to narrative creation and engagement represents a new kind of formalism, one that rejects the so-called "autonomy" of the work of art—i.e., the idea that works of literature possess an inherent unity and coherence, separate from their realization by readers—in favor of the embodied cognitive dynamics that instantiate that form. Certainly formal patterns exist in the literary object. But enactivist theory challenges the objectivist pretensions of these older formalisms by asking, "how we can we understand the forms and patterns that underly narratives as a function of embodied processes, sensorimotor knowledge, and ecological embeddedness?" In this way, I believe an enactivist approach to literature can offer ways of understanding how literary narratives move between representations of embodied experience and embodied experience itself.

The cognitive scaffolding of narrative structuration is the foundation for many of the aesthetic, formalist concerns that predominate in the study of art and literature. As

Armstrong tells us, within this dynamic-systems context of the mind, stories must be seen as:

> biocultural hybrids—recurrent configurations that develop because certain repeated characteristics of our species' shared experiences of birth and death, collaboration and competition, propagation and violence interact with biologically based cognitive proclivities to produce statistically discoverable regularities in cultural institutions, including the stories we circulate in our communities. (Armstrong 25)

This is an (overly) dense way of saying that stories are both socially constructed objects, and essential cognitive architecture. The language of "recurrent configurations" "shared experiences" and "biological proclivities" also gestures towards the universal features of mind and human sociality that underpin the cognitive role of narrative—a claim that will be anathema to many people working from within the disciplinary inheritances of the hermeneutics of suspicion, but which I think a cognitive science and systems theoretical perspective necessarily entails. Literary theory and criticism have spent the better part of the last century unpacking the social construction of literature. With our present-day renewed interest in consciousness, as well as improved brain scanning, mapping, and computing techniques, we are in a better position to understand and expound on the biological and behavioral mechanisms that undergird its socially constructed aspects.

In one of the most overarching observations about the cognitive role of narrative offered in the book, Armstrong says, "Stories help the brain negotiate the never-ending conflict between its need for pattern, synthesis, and constancy and its need for flexibility, adaptability, and openness to change" (Armstrong 11).[85] We are given two basic behavioral poles: constancy and change. The brain must affirm patterns that will enable ongoing fitness and evolutionary success, but it must be flexible enough to reorganize those patterns in the face of novelty. Armstrong calls the synthesizing, agreement-seeking dimension to narrative experience "concordance," while the ways in which narratives disrupt and defamiliarize perception he labels "discordance." In reference to work by Frank Kermode (*The Sense of an Ending* 1967) and later Ricoeur (*Time and Narrative* 1983), Armstrong argues that emplotment is a process of "concordant discordance" (Armstong 12), which is his attempt to further situate the configurative aspects of narrative engagement in cognitive scientific and neurophenomenological discourse.

Armstrong says, "If congruent processes of figuration in the telling and the told can facilitate immersion or instruction, disjunctions between discourse and story can defamiliarize and foreground cognitive processes that may ordinarily go unnoticed in reading or in life because we are absorbed by our habitual ways of knowing" (Armstrong 34). In this passage we see the behavioral poles recapitulated in the language of structuralist narratology, demonstrating how phenomenological (i.e., figurative) approaches to narrative can still exploit basic structuralist concepts—in this case, the perennial distinction in narratology between the "telling" and the "told" (or discourse/story), in which the told refers to the actions, characters, events being represented, while the telling refers to the way that raw material is sequenced and shaped

[85] In this passage, we see the two basic poles outlined by Thompson and Varela at the neuronal scale (i.e., the integration and dissolution of various neural assemblies)—recapitulated in terms of the larger behavioral dynamics that these cortical processes facilitate.

into a plot by the narrator. In other words, we can still use the language of static forms and structures in a phenomenological analysis, provided that "naming a technique or a structure is only the first step in analyzing the configurative process…" (Armstrong 37).

As Armstrong tells us, pattern-affirming concordances in narrative can lead to *immersion* in the story word, which is arguably the most fundamental aesthetic pleasure that people take from stories. Such immersion into the story world is vital in facilitating the reader's instruction in its ideas, object lessons, and in the cultivation of empathy and affect so often associated with literature. Conversely, discordances can bring to light those habits of thought and perception that have faded into the transparency of one's experience and thereby make the world new.

One of the most obvious and enduring ways in which narratives manifest discordance and make the world feel new is in their ability to scramble time and create systematic discontinuities that work against the various concordant processes of time-sequencing. For example, one of the primary ways that modernist and postmodernist literature was able to break with the traditions of the past was by experimenting with formal methods for emploting time: from the stream-of-consciousness style of Joyce and Woolf, to the so called "delayed-decoding" of Conrad; the austere minimalism of the Beckett trilogy, to the episodic and highly fragmented structure of William Borroughs' *Naked Lunch*. This tendency towards "discordance" is also reflected in many 20th century works on literary aesthetics, for example Victor Shklovsky's concept of *defamiliarization*, or Darko Suvin's concept of *cognitive estrangement*. What these formal experiments in literature and the aesthetics of Shklovsky and Suvin have in common is the role of art in disrupting perception and combatting habituation.[86]

Defamiliarization acts in the realm of perception, *not* understanding. It argues that the purpose of art is to impede perception—to slow it down—and thus make the world strange, unfamiliar, and new. Shklovsky says, “The purpose of art is to impart the sensation of things *as they are perceived and not as they are known*. The technique of art is to make objects ‘unfamiliar,’ to make forms difficult, to increase the difficulty and length of perception because the process of perception is an aesthetic end in itself and must be prolonged” (Shklovsky 12; emphasis added). Here Shklovsky speaks in terms of art in general, but the decisive point for literature is the distinction between practical and poetic language. Practical language is aimed at economy and intelligibility. It tries to reduce the cognitive load, improve understanding and intelligibility, and more fluidly integrate us with our environment. Shklovsky calls this process *automatization*. Poetic language (like art more generally) is not aimed at the intelligible but at the perceptible. Ultimately for the Russian Formalists, defamiliarization was a concept that helped to distinguish literary from non-literary objects because it defined the essence of literature (literariness) and thus conformed to their scientific pretensions.[87] It also helped to define literature’s main purpose. This was an aesthetic distinction also made use of by Iser, e.g., the “gaps” that emerge in the reading process). Not only am I convinced by their position, but it speaks to

[86] Cognitive estrangement, while differing slightly from the Russian Formalist’s conception of defamiliarization in that the estrangement aimed at is “cognitive” (i.e., conceptually explicit cognition), still cleaves to this basic idea of defamiliarizing the known and thereby making it more salient for perception.

[87] We encountered this distinction in Iser’s work, e.g., the “gaps” that emerge in the reading process.

a well-defined and well-studied set of phenomena in cognitive science and psychology known generally as *habituation*.

Habituation is a fundamental aspect of learning and behavior. Rankin et al. in their retrospective analysis of the concept titled "Habituation revisited…" state the general scholarly consensus that:

> Researchers who work on this form of learning believe that because habituation allows animals to filter out irrelevant stimuli and focus selectively on important stimuli, it is a prerequisite for other forms of learning. Therefore, to fully understand the mechanisms of more complex forms of learning and cognition it is important to understand the basic building blocks of habituation. (136)

In other words, all roads in the realm of behavior and learning lead back to the phenomenon of habituation in some way. It is the prerequisite and foundation for all other forms of learning. They point out how standard measures of habituation have traditionally been external bodily markers of the nervous system such as pupil response, sweating, muscle contraction and even hormone release (136). Most of the core ideas around habituation were the result of research done in the 1960s. Technological advancements in the intervening years have meant that researchers can now measure "cellular or molecular responses or neuronal activity, including population activity, such as measured with EEG or functional imaging. These responses at the molecular, cellular or population levels may be monitored in an effort to identify underlying mechanisms or they may be used as indices of habituation" (136). Thus, not only is habituation posited as the keystone of learning and behavior in general, but it remains an ongoing area of research that is amenable to the tools and techniques of cognitive neuroscience. While Shklovsky and his contemporaries were undoubtedly ignorant of the underlying mechanisms of the nervous system involved in habituation, they nonetheless created a cogent aesthetic theory around the intuition that perception operates by automatization (i.e., habituation). The Russian Formalists, by explicating this intuition in the realm of literature, provide another lucid framework for integrating cognitive science with literature and aesthetics.

Emergent Dynamics and Defamiliarization:

In the above analysis, we have broached many useful concepts for situating narratives in cognitive terms. Distinctions between mimesis/configuration, consonance/dissonance, habituation/defamiliarization, telling/told, practical/poetic language provide some essential coordinates for mapping enactivism onto the realm of literary studies.

I want to end this chapter by asking where does Deacon's theory of emergent dynamics fit into this new, embodied formalism? I think the best way to start answering this is to return to his insights regarding the nature of *work*. According to Deacon's theory, work results from the coupling of different orthodynamic tendencies (e.g., the piston and the gas). Because this coupling constrains each system's spontaneous tendencies, work is a reflection of non-spontaneous change (327), or stated differently, the production of contragrade change (337).

We tend to pigeonhole and/or reduce the concept of work to observable changes in bulk thermodynamic properties, particularly in mechanical and engineering contexts. One reason for this rather myopic conception of work is that once a system undergoes an emergent transition—e.g., from homeo to morphodynamics—it becomes very easy to lose sight of the underlying thermodynamic processes that are still operative within it. But by adhering to emergent dynamics' more abstract change-oriented conception of work we can start to trace its manifestation in more complex systems and their interactions.

For example, we know that morphodynamic systems are extremely susceptible to disturbance and dissolution because they depend on a precise balance of constraint elimination and production. The sensitivity of morphodynamic processes means that "describing the way that the interaction between morphodynamic processes can transform their orthograde dynamics into contragrade change in the morphodynamic domain (i.e., the description of a morphodynamic engine, if you will) is far more difficult than for thermodynamic systems" (350). This means that examples of morphodynamic work in nature are fewer and far between compared to thermodynamic work (347). Nevertheless, we have already encountered and dealt with one exemple in detail: the autogen, which Deacon calls a "special case, because of its precise recursive synergistic organization" (350).

At this point we have also dealt at length with teleodynamic work: *evolution by natural selection*. Organisms are teleodynamic systems that have an orthograde tendency towards self-maintenance, repair, and reproduction. When organisms interact and compete for finite resources, over time we see the differential preservation and inheritance of certain advantageous traits. This means that natural selection logic can also be expressed as the myriad interactions of teleodynamic systems, each with their own unique "equilibrium" points and assemblages of constitutive constraints, all of which results in a kind of "engine" that can drive phylogenesis.

The most important examples of teleodynamic work for our purposes are the cognitive processes of organisms with nervous systems. The orthograde dynamics of nervous systems involve all kinds of subsystems, such as the internally-oriented feedback loops that regulate bodily functions like breathing, digestion, and heart rate, or higher forms of interoception; to the dynamical fluctuations of neural assemblies that occur in the brain's default mode network and that constitute a kind of basal state of cognition; or the externally-oriented sensorimotor feedback loops that constitute perception. In teleodynamic work, these myriad, orthograde cognitive processes become coupled to, and thus constrained by, other teleodynamic systems—including their repositories, such as the written word. Deacon says:

> Teleodynamic work is what we must engage in when trying to make sense of an unclear explanation, or trying to produce an explanation that is unambiguous. It is what must be produced to solve a puzzle, to persuade resistant listeners, or to conduct scientific investigations. It is also the sort of work that goes on in board meetings and in domestic arguments, and which leads to the design of machines and governments. And it characterizes what is difficult about creative thought processes. (359)

This passage demonstrates how all higher-order cognitive processes can be understood as forms of teleodynamic work, and it implies that when we think alongside and through linguistic objects, such as literary narratives, we are also engaging in teleodynamic work. This is where the rubber meets the "emergent dynamic" road for literary studies.

The idea that narratives are tools for shaping cognitive processes has long been at the heart of formalist analysis. We have just seen how the Russian Formalist language of "defamiliarization" and "automatization" can speak to contemporary neuroscientific findings regarding habituation. We saw how habituation and its inverse sensitization (or more broadly, neural plasticity), were initially used to describe the behavioral dimensions of learning but have now been extended into the underlying structure of neurons and neuronal assemblies. We also saw this habituation/plasticity dynamic at work in Paul Armstrong's analysis of "concordant discordance." At this point I think we can safely assert that *habituation and defamiliarization are also different ways of naming the brain's orthograde and contragrade processes.*

A formalist approach is the most salient for mediating between literary studies and cognitive science because formalism tends to avoid or diminish a narrative's objects of reference and representation (what novels are *about*) and looks instead to the patterns that constitute them. This attention to form over content can make it easier to abstract patterns and isolate their material (i.e., cognitively measurable) effects. Narrative patterns are interesting objects of study in terms of emergent dynamics because of their capacity to alter or amplify uninterrogated or habitual emotional and intellectual patterns. Enculturation itself is a process of pattern recognition and assimilation, and the preservation of certain narrative patterns in culture (e.g., religious archetypes, kinship roles, etc.) can even be construed as a kind of cultural phylogenetic inheritance. This is because inevitably certain patterns will be more or less successful in a given cultural or intellectual milieu. As these patterns get established, they entrain and constrain our capacity to generate and interpret meaning and thus contribute to our social and psychological fitness. Narratives exploit and undermine these cultural patterns to great effect, producing defamiliarizing or habituating aesthetic effects.

Literary scholarship itself has often been directed into well-defined patterns of thinking in this respect. For example, Marxist critiques of labor and culture try to understand how art functions as an ideological and historical tool; psychoanalytic analyses examine how linguistic structures and desire become coded in the stories we tell ourselves and the cultural archetypes we give credence to; while liberal humanist approaches appeal to universal norms and the edifying nature of canonical works. These approaches all operate on the axiom and intuition that literature has the capacity *to do things* in the world.

Why is all of this relevant to the field of literary studies? Don't the above examples attest to the fact that this insight (*that stories do actual work in the world*, be it cultural, intellectual, moral-psychological etc.) has been at least intuitively operative in literary analysis for the better part of the twentieth century, if not all the way back to Aristotle's *Poetics* and his notion of catharsis?

My contention here is that while we intuitively understand that stories have the capacity to do work, we do not always have a clear or consistent grasp on where that work starts or how it ramifies into other disciplines. By defining literary and interpretive work in the abstract and protean terms outlined by emergent dynamics, I believe there is the potential to synthesize all manner of psychological, social, political, and cultural critique. I believe that this is particularly true for literary scholars working in the field of ecocriticism. By digging into the thermodynamic fundament, and by coming to terms with more general notions of change and causation, we can provide more clear and cogent links to other disciplines, particularly to many of the hard science disciplines on which ecocriticism depends. Not only is this an interesting exercise in its own right, but it strikes me as the ultimate eco-critical move. Deacon supplies us with the means of understanding these patterns of concordant discordance as some of the most complex examples of systemic constraint.

Conclusion:

In this chapter we have emerged from the embodied depths and traced enactivism's sensorimotor claims firmly into the world of literary narratives and their interpretation. Not only can we reconceive of literary engagement and its gestalts as an embodied perceptual process, but we can also see how formalist approaches that have long urged us to focus on form over content—perception over representation—have been legitimated by contemporary neuroscientific work on learning and behavior. Finally, we have seen how this new embodied (or sensorimotor) formalism can also be translated into the language of Deacon's emergent dynamics, and in so doing, give ecocritics a way to ground their labors in thermodynamic work—in true environmental and ecological bedrock.

Dissertation Conclusion:

I am the eye with which the Universe
Beholds itself, and knows itself divine;

—Percy Shelley, "Hymn of Apollo"

This is the opening line from the final stanza of Percy Shelley's "Hymn of Apollo," which is written from the first-person perspective of Apollo, Greek god of poetry, music, light, healing and prophecy. Throughout the poem the speaker celebrates the transformative power of light on the myriad processes of nature. Even though most of the poem is a celebration of the speaker's *own* deific powers to participate in, and transform, these processes, in this line the speaker casts doubt on the source of this agency and power. The speaker asserts himself as a subject—a locus of action and causation—in the very act of speaking; but at the same time, this agency and subjectivity emerges as an extension of a larger cosmic and sublime whole. The autonomy of the "I" is instrumentalized into an "eye," subordinated to the metaphysical will of the universe and its innate desire to *know* itself. The universe enacts its self-knowledge through the eye of the speaker, and the speaker enacts his "I" in the context of this cosmic and sublime whole.

I think this passage nicely stages what Capra and Luisi name *systems thinking*, "…the understanding of a phenomenon within the context of a larger whole. This is, in fact, the root meaning of the word 'system,' which derives from the Greek *syn* + *histanai* ('to place together'). To understand things systemically means literally to put them into a context, to establish the nature of their relationships" (Capra and Luisi 64). This mode of thinking involves a shift away from material objects and structures and towards nonmaterial processes and patterns. It involves viewing nature in terms of "multileveled structures of systems within systems" (Capra and Luisi 64), in which each system forms a whole with respect to its parts, but which can, in turn, be understood as constitutive of still larger wholes. Systems thinking is an approach for defining and circumscribing the world in terms of mutable relationships and shifting patterns of parts and wholes. Throughout this dissertation we have looked to the ways in which systems thinking makes the relationships and connections of nature salient. It is an extremely protean and generative heuristic for connecting all manner of disciplines. The ability to clearly define the boundaries of one's objects, to abstract patterns of interaction both within and outside of these boundaries, and to unequivocally describe the role of the observer and his or her effects on the system being described are universal features of knowledge production.

The above passage by Shelley also highlights a basic tension inherent to all living organization—from the semi-divine powers of Apollo (which, in the context of the poem, is a thin stand-in for the semi-divine powers of humanity) to the bacteria that have served as our cognitive urtexts. As Capra and Luisi tell us, "The double role of living systems as parts and wholes requires the interplay of two opposite tendencies: an integrative tendency to function as part of a larger whole, and a self-assertive, or self-organizing tendency to preserve individual autonomy" (Capra and Luisi 65). Shelley stages this tension in the Deistic metaphysics characteristic of the Romantic movement: the universe is transcendent and divine. It remains, in some fundamental sense, *outside* the totality, and this transcendence enables it to suffuse all of nature's workings. But what Shelley renders as transcendent, systems thinking allows us to render in terms that are fully immanent and ecological. Ecological thinking is synonymous with systems thinking, and I believe that much of the conceptual progress that has occurred in ecocriticism and the environmental humanities is the result of theorists availing themselves of the tools of systems theory, in one form or another.

Throughout this project, we have seen how dynamical systems theory is currently the only theory that stands a promise of bridging the gap between humans and their beginnings in non-human modes of existence. We have also seen how autopoietic enactivism can provide us with the scientific and phenomenological tools that can contextualize human cognition—and all its pretenses to superiority—within the larger self-organizing tendencies of nature. In this way, autopoietic enactivism has an important role to play in the environmental humanities and what is arguably its primary philosophical aim: to overcome the anthropocentric bias of modern thought. The face of nature speaks to us at all scales of existence, from the bacterial to the planetary. In order to ethically respond we must first overcome this bias.

Glossary

absential (Deacon)**.** The property of not being materially present, of constitutive lack. See also **ententional**.

autogen (Deacon). A thought experiment designed to describe in precise, stepwise detail the emergence of purposive (or teleological) organization—organization that defies the second law of thermodynamics (entropy). It is the most simple systemic unity that results from the process of autogenesis. When paired with our understanding of autopoiesis, the autogen can clarify the energetic and thermodynamic underpinnings which are left unexamined in that process.

autogenesis (Deacon). Deacon's alternative to autopoiesis that encompasses non-living systems. A process of self-organization, self-maintenance, and self-repair that results from teleodynamic organization.

autonomy. The opposite of **heteronomy** (q.v.). The capacity for/property of self-determination. For Maturana/Varela, the principal characteristic of life.

autopoiesis (Maturana/Varela). Literally self-making (auto-poiesis). The process that characterizes life, as opposed to inorganic systems. An autopoietic unity is a unity that creates and maintains itself by continuously producing its components and incorporating them into a stable pattern of organization. A system which keeps its organizational pattern as the variable that remains constant despite the constant exchange of matter and energy.

autopoietic enactivism (Thompson). Differs from Varela et al.'s approach to enactivism by incorporating the insights of autopoiesis, and thus moving beyond enactivism's initial neuroscientific context.

cognition. Has many different meanings, depending on context. For Maturana/Varela, all the behaviors that enable and actualize an organism and its autopoiesis—the process of life itself. Existing at all levels of life and lower levels of cognitive ability, cognition is thus broader than consciousness.

cognitivism. First phase of cognitive science in which the mind is understood as a computer-like symbol-processing machine working on a linear input/output model.

connectionism. Second phase of cognitive science in which the mind is understood as a neural network composed of myriad nodes and links, and which utilizes feedback and control mechanisms (circularity) to surpass cognitivism's linear input/output model.

contragrade. The opposite of **orthograde** (q.v.) Forced changes or tendencies in a system, i.e., the result of external forces. Changes or tendencies that "go against the flow," i.e., that disrupt or alter an established pattern of organization.

constraint (Deacon). A system state *not* realized. A restriction in dynamical possibilities. In Deacon's emergent dynamics, constraints arise from the interaction of orthograde tendencies (thus, constraint is a function of contragrade (forced) change). Constraints are central to Deacon's theory of emergent dynamics because it provides an alternative to the concept of order/pattern. An ordered arrangement or pattern implies a repetition of similar features and thus implicitly depends on the perspective of some outside observer. *Constraint redefines this in negative terms—as fewer total differences*. The more constraints, the less total differences (i.e., the more similarities). Thus, an increase in constraints can also be understood as a decrease in entropy. In this way, we arrive at a notion of "order" that does not appeal to the perspective of some situated observer.

cybernetics. The science that studies control and feedback mechanisms in both living and mechanical systems. These control mechanisms are understood in terms of their circular organization—i.e., positive and negative feedback loops.

developmental systems theory (DST). Originated in the work of Susan Oyama. An approach to systems theory in which an organism's development and genetic "information" is no longer confined to genetics, and instead emerges in dynamical interrelation with myriad aspects of an organism's genes and their environments. DST is enactivism's holistic answer to the reductionist tendencies at work in neo-Darwinism and beyond.

dynamical systems theory. A mathematical approach to modelling complex, dynamical (i.e., constantly changing) systems using topological and geometrical techniques. It is considered a "qualitative" approach, meaning that it cannot predict an exact system state at a particular point in time, but rather a system's general behaviors and patterns of organization as they develop in geometrical space (i.e., attractors).

embodied dynamicism. The third phase of cognitive science in which the mind is understood as a connectionist network, but now this network is expanded to include the constitutive roles played by the body and its immediate environment.

emergence (in dynamical systems theory). The appearance of novel properties at higher-order levels of complexity (e.g., whirlpools, life) that do not exist at lower levels. A (seemingly) radical change in causal processes that is endogenous, caused by the interaction of a system's parts. In the context of autogenesis (Deacon), the creation of a new orthodynamic tendency.

emergent dynamics. Deacon's approach to dynamical systems theory, which focuses on general patterns of change (orthograde/contragrade) and the role of system states not realized (constraint), to describe the emergence of complex systems.

enaction. The co-emergence of self and world realized through the processes of cognition. It holds that self and world do not preexist thought; they are brought into being by cognition (i.e., thought-as-action). An organism enacts its own meaningful milieu (Umwelt) and thereby also gives rise to itself. This meaning is not **representation** (q.v.) because it does not exist prior to its enaction.

enactivism (Varela, Thompson, Rosch). An approach to cognitive science that critiques **representation** (q.v.) and which avers that **enaction (q.v.)** is the mind's quintessential activity.

ententional/ententionality (Deacon). Quality of existing in relation to something non-intrinsic. A class of phenomena that are intrinsically incomplete in that they exist in relationship to something materially absent (a purpose, a meaning, a possibility). (The human heart exists for the sake of keeping the organism alive. Books exist to be read.)

ethics. Responsibility (i.e., response-ability) towards other living things and the ecological collectives they inhabit. Active recognition of another organism's autonomy and interiority, as well as their capacity to suffer.

face (Levinas). The way in which another being—possessing interiority and autonomy—confronts and questions one's own sense of autonomy and one's own sense of self. Any *expression* of this other beings' interiority and autonomy. In the phenomenology of Emmanuel Levinas, the face is an encounter with the infinite (the more within the less).

heteronomy. Determination by an outside force. Determination of a system boundary or pattern of organization by an outside observer. The quintessential perspective of materialist science.

homeodynamics (Deacon). Deacon's recasting of thermodynamics. Homeodynamic organization is analogous to basic thermodynamic systems and their bulk (e.g., temperature, pressure) or statistical (e.g., microscopic motions of atoms at their interactions) properties, but reconceived in terms of orthograde and contragrade change, and constraint. The second law of thermodynamics represents the paradigm case, but for Deacon, its characteristics (the elimination of constraints such as dissipation to the point of thermodynamic equilibrium—maximum entropy) can also apply in many other contexts (my stream of consciousness, for example).

negative feedback. Mechanisms that damped or constrain a process and keep a system within a desired range.

negentropy. The opposite of entropy. An increase in order.

operational closure (in autopoietic theory). The condition of autonomy of living systems ensuring that external force can only disturb them but never determine their behavior. When a system's "operations" or pattern of organization is maintained via *internal* processes that compensate for change and *not* as the effect of some external cause or input.

orthograde (Deacon). Spontaneous (unforced) changes or tendencies in a system, i.e., that arise without external influence. Changes or tendencies that "go with the flow," i.e., that continue an established pattern of organization.

morphodynamics (Deacon). Deacon's recasting of self-organization. The class of system defined by the amplification or propagation of constraints. Morphodynamic organization is analogous to self-organization, but reconceived in terms of orthograde and contragrade change, and constraint.

phenomenology. Study of the way things appear to mind. The study of interiority. Also, a meditative or self-reflective practice for dispensing with preconceptions and/or habitual ways of knowing and perceiving.

perception. Depends on sensation (mechanistic and passive). An active process that interprets sensation.

positive feedback. mechanisms that amplify a process and thus move a system beyond equilibrium.

representation. Any sign, signal, or configuration of parts that stands for some objective/preexisting reality or that mediates between a preexisting outside world (object) and an inner world of thought (subject). An isomorphic relationship between a mental state and some outside domain. Works like a map.

second-order cybernetics. The logic of cybernetics (i.e. circular organization/causality) applied to itself. The science of recursion (formal self-reference).

sentience (Thompson). The most fundamental feeling of being alive. An interiority and inner experience characterized by significance and valence.

Shannon information. Has nothing to do with semantic (content) information. A mathematical theory for determining the carrying capacity of a communication medium based on the medium's entropy. It calculates the entropy expected in the signal medium and uses this to measure the degree of order/repetition in received in the signal (i.e., the information therein). The lawful (predictable and quantifiable) relationship between signal and signal medium described Shannon information theory has been extended throughout scientific discourse to describe how any two things can vary lawfully in relationship to one another (e.g., the rings of a tree carry "information" about the age of the tree; the byte capacity of USB drives).

structural coupling (in autopoietic theory). Structural congruence; a history of mutual determination and interdependence between two or more systems (including the system and its environment).

teleodynamics (Deacon). Class of system based on the creation, propagation, and most importantly preservation of constraints. The highest level of an emergent dynamic description.

Works Cited:

Albright, Thomas D. “Perceiving.” *Daedalus*, vol. 144, no. 1, Jan. 2015, pp. 22–41. *JSTOR*, https://doi.org/10.1162/daed_a_00315.

Aristotle. *Poetics*. Translated by Gerald Frank Else, University of Michigan Press, 2012.

Armstrong, Paul B. *Stories and the Brain: The Neuroscience of Narrative*. Johns Hopkins University Press, 2020.

Atkins, Peter William. *The Laws of Thermodynamics*. Oxford University Press, 2010.

Badin, Antoine-Scott, et al. “The Features and Functions of Neuronal Assemblies: Possible Dependency on Mechanisms beyond Synaptic Transmission.” *Frontiers in Neural Circuits*, vol. 10, 2017, pp. 114–114, https://doi.org/10.3389/fncir.2016.00114.

Bakhtin, Mikhail. “Discourse in the Novel.” *The Dialogic Imagination: Four Essays*, translated by Michael Holquist, University of Texas Press, Austin, TX, 1982.

Baring, Edward. “Levinas and Derrida.” *The Oxford Handbook of Levinas*, edited by Michael L Morgan. Oxford University Press, 2018, *Oxford Handbooks Online*, Accessed 2 Nov. 2024.

Belsey, Catherine. *Critical Practice*. 2nd ed., Routledge, 2002.

Ben-Menahem, Yemima. “Philosophy of Mind of Hilary Putnam.” *Encyclopædia Britannica*, Encyclopædia Britannica, inc., www.britannica.com/biography/Hilary-Putnam/Philosophy-of-mind. Accessed 9 Nov. 2024.

Best, Stephen, and Sharon Marcus. “Surface Reading: An Introduction.” *Representations*, vol. 108, no. 1, 2009, pp. 1–21. *JSTOR*, https://doi.org/10.1525/rep.2009.108.1.1. Accessed 24 Mar. 2025.

Bray, Dennis. *Wetware: A Computer in Every Living Cell*. Yale University Press, 2011.

Bruner, Jerome. “The Narrative Construction of Reality.” *Critical Inquiry*, vol. 18, no. 1, 1991, pp. 1–21, https://doi.org/10.1086/448619.

Cain, Mark J. *The Philosophy of Cognitive Science*. E-book, Polity Press, 2015. *ProQuest Ebook Central*, https://ebookcentral.proquest.com/lib/washington/detail.action?docID=4187285.

Capra, Fritjof, and Pier Luigi Luisi. *The Systems View of Life: a Unifying Vision*. Cambridge University Press, 2019.

Caracciolo, Marco, and Karin Kukkonen. *With Bodies Narrative Theory and Embodied Cognition*. The Ohio State University Press, 2021.

Carney, James. "Thinking avant la lettre: A Review of 4E Cognition." *Evolutionary Studies in Imaginative Culture*, vol. 4, no. 1, spring 2020, pp. 77+. *Gale Academic OneFile*, link.gale.com/apps/doc/A623924810/AONE?u=wash_main&sid=bookmark-AONE&xid=f9eb8c1b. Accessed 30 Jan. 2023.

Carson, Rachel. *Silent Spring*. Mariner Books, 2022.

Chandler, Daniel, and Rod Munday. "apparent motion." *A Dictionary of Media & Communication*. 3rd ed., Oxford University Press, 2020. Oxford Reference. Date Accessed 12 Apr. 2024

Clark, Andy. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press, 2011.

Clarke, Bruce. *Gaian Systems: Lynn Margulis, Neocybernetics, and the End of the Anthropocene*. University of Minnesota Press, 2020.

---. "Heinz Von Foerster's Demons." Clarke, Bruce, and Mark B.N. Hansen, editors. *Emergence and Embodiment: New Essays on Second-Order Systems Theory*. Duke University Press, 2009.

Clarke, Bruce, and Mark B.N. Hansen, editors. *Emergence and Embodiment: New Essays on Second-Order Systems Theory*. Duke University Press, 2009.

Commoner, Barry. *The Closing Circle: Nature, Man, and Technology*. Dover Publications, Inc, 2020.

Costandi, Moheb. *Neuroplasticity*, MIT Press, 2016. *ProQuest Ebook Central*, https://ebookcentral.proquest.com/lib/washington/detail.action?docID=4648373.

Damasio, Antonio R. *Feeling & Knowing: Making Minds Conscious*. Pantheon Books, 2021.

Deacon, Terrence William. "How molecules became signs." *Biosemiotics*, vol. 14, no. 3, 25 Sept. 2021, pp. 537–559, https://doi.org/10.1007/s12304-021-09453-9.

---. *Incomplete Nature: How Mind Emerged from Matter*. W. W. Norton, 2013.

Dennett, Daniel C. "Aching Voids and Making Voids." *The Quarterly Review of Biology*, vol. 88, no. 4, The University of Chicago Press, 2013, pp. 321–24, https://doi.org/10.1086/673760.

---. "Shall We Tango? No, but Thanks for Asking." *Journal of Consciousness Studies*, vol. 18, 2011, pp. 23–34, https://dl.tufts.edu/concern/pdfs/44558r75m.

Derrida, Jacques. *Writing and Difference*. Translated by Alan Bass, Univ. of Chicago Press, 2002.

Descartes, René. *Discourse on Method and Meditations on First Philosophy*. Translated by Donald A. Cress, 4th ed., Hackett Publishing Company, 1998.

"distribution." *The Concise Oxford Dictionary of Mathematics*: Oxford University Press, , 2021. *Oxford Reference*. Date Accessed 21 May. 2024 <https://www-oxfordreference-

com.offcampus.lib.washington.edu/view/10.1093/acref/9780198845355.001.0001/acref-9780198845355-e-886>.

Emmett, Robert S., and David E. Nye. *The Environmental Humanities: A Critical Introduction*. 1st ed., The MIT Press, 2017, https://doi.org/10.7551/mitpress/10629.001.0001.

Foer, Jonathan Safran. *Eating Animals*. Little, Brown and Company, 2009.

Foucault, Michel. *Discipline and Punish: The Birth of the Prison*. Translated by Alan Sheridan, Vintage, 1995.

Gadamer, Hans-Georg. *Truth and Method*. Translated by Donald G. Marshall and Joel Weinsheimer, Second, Revised Edition ed., Continuum, 2004.

Gallagher, Shaun, and Dan Zahavi. *The Phenomenological Mind*. 3rd ed., Routledge, 2021.

Pace Giannotta, Andrea. "Autopoietic Enactivism, Phenomenology, and the Problem of Naturalism: A Neutral Monist Proposal." *Husserl Studies*, vol. 37, no. 3, 2021, pp. 209–28, https://doi.org/10.1007/s10743-021-09286-x.

Glotfelty, Cheryll, and Harold Fromm. *The Ecocriticism Reader: Landmarks in Literary Ecology*. University of Georgia Press, 2009.

Grider, Michael H., et al. "Physiology, Action Potential." *StatPearls [Internet].*, U.S. National Library of Medicine, 8 May 2023, www.ncbi.nlm.nih.gov/books/NBK538143/.

Grondin, Jean. "The Hermeneutical Circle." *The Blackwell Companion to Hermeneutics*, edited by Niall Keane and Chris Lawn, Wiley-Blackwell, 2016, pp. 399–407.

Haddad, Wassim M. *A Dynamical Systems Theory of Thermodynamics*. Princeton University Press, 2019.

Harman, Graham. *Tool-Being: Heidegger and the Metaphysics of Objects*. Open Court, 2002.

Hayles, N Katherine. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature and Informatics*. Univ. of Chicago Press, 1999.

Heidegger, Martin. *Being and Time*. Translated by John Macquarrie and Edward Robinson, Harper Perennial, 2008.

Hoopes, James. "Introduction." *Peirce on Signs: Writings on Semiotic by Charles Sanders Peirce*. Edited by James Hoopes, e-book, The University of North Carolina Press, 2014, *Project Muse*.

Houdé, Olivier, et al. "Perception." *Dictionary of Cognitive Science*, Routledge, 2004, pp. 348–54, https://doi.org/10.4324/9780203486030-82.

Hutto, Daniel D. “Philosophy of Mind’s New Lease on Life: Autopoietic Enactivism meets Teleosemiotics.” *Journal of Consciousness Studies*, vol. 18, no. 5-6, 2011, pp. 44–64.

Hutto, Daniel D., and Erik Myin. *Radicalizing Enactivism: Basic Minds Without Content*, MIT Press, 2012. *ProQuest Ebook Central*, https://ebookcentral.proquest.com/lib/washington/detail.action?docID=3339551.

Irigaray, Luce. *This Sex Which Is Not One*. Translated by Catherine Porter, Cornell University Press, 1985.

Iser, Wolfgang. “The Reading Process: A Phenomenological Approach.” *New Literary History*, vol. 3, no. 2, 1972, pp. 279–99. *JSTOR*, https://doi.org/10.2307/468316. Accessed 21 Apr. 2025.

“Isomorphic, Adj.” *Oxford English Dictionary*, Oxford UP, December 2024, https://doi.org/10.1093/OED/7501387173.

Johnston, Adrian. “Jacques Lacan.” *Stanford Encyclopedia of Philosophy*, Stanford University, 24 Dec. 2022.

Kallestrup, Jesper. *Semantic Externalism*, Taylor & Francis Group, 2011. *ProQuest Ebook Central*, https://ebookcentral.proquest.com/lib/washington/detail.action?docID=838177.

Kant, Immanuel. *Critique of Pure Reason*. Translated by Werner S. Pluhar, Hackett Publishing, 1996.

Keysers, Christian, and Valeria Gazzola. “Hebbian learning and predictive mirror neurons for actions, sensations and emotions.” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1644, 5 June 2014, https://doi.org/10.1098/rstb.2013.0175.

Kohn, Eduardo. *How Forests Think: Toward an Anthropology beyond the Human*. University of California Press, 2013.

Kornblum, Sylvan., and David E. Meyer. “Part III Attention: Introduction” *Attention and Performance XIV : Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience*. MIT Press, 1993.

Lacan, Jacques. *Ecrits: The First Complete Edition in English*. Translated by Bruce Fink, WW Norton & Co, 2007.

Laozi. *Tao Te Ching*. Translated by Stephen Mitchell, HarperPerennial, 1992.

Lazari, Alberto, et al. “Hebbian Activity-Dependent Plasticity in White Matter.” *Cell Reports (Cambridge)*, vol. 39, no. 11, 2022, pp. 110951–110951, https://doi.org/10.1016/j.celrep.2022.110951.

Legg, Catherine. "Discursive Habits: A Representationalist Re-Reading of Teleosemiotics." *Synthese (Dordrecht)*, vol. 199, no. 5–6, 2021, pp. 14751–68, https://doi.org/10.1007/s11229-021-03442-8.

Legg, Catherine, and Joshua Black. "What Is Intelligence For? A Peircean Pragmatist Response to the Knowing-How, Knowing-That Debate." *Erkenntnis*, vol. 87, no. 5, 13 Aug. 2020, pp. 2265–2284., https://doi.org/10.1007/s10670-020-00301-9.

Levinas, Emmanuel. *Nine Talmudic Readings*. Translated by Annette Aronowicz, Indiana University Press, 2019. *ProQuest Ebook Central*, http://ebookcentral.proquest.com/lib/washington/detail.action?docID=30657059. Accessed 26 Oct. 2024.

---. *Otherwise Than Being, Or, Beyond Essence*. Trans. Alphonso Lingis. Pittsburgh, PA: Duquesne UP, 1998. Print.

---. "Philosophy and the Idea of Inifnity." *Collected Philosophical Papers*. Translated by Alphonso Lingis, Duquesne University Press, 1998.

---. *Totality and Infinity: An Essay on Exteriority*. Trans. Alphonso Lingis. Pittsburgh, PA: Duquesne UP, 1994. Print.

Liddell, Henry George, and Robert Scott. *Liddell & Scott Greek-English Lexicon, Abridged*. Follett Pub. Co. 1949.

Lones, John, et al. "From sensorimotor experiences to cognitive development: Investigating the influence of experiential diversity on the development of an epigenetic robot." *Frontiers in Robotics and AI*, vol. 3, 26 Aug. 2016, https://doi.org/10.3389/frobt.2016.00044.

Luijpen, William A. *Phenomenology and Humanism*. Duquesne University Press, 1966.

Maturana, Humberto R., "Biology of Cognition." *Autopoiesis and Cognition: The Realization of the Living* by Humberto R. Maturana and Francisco J. Varela. D. Reidel, 1980, pp. 5-58.

Maturana, Humberto R., and Francisco J. Varela. "Autopoiesis: The Organization of the Living." *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel, 1980, pp. 73-134.

---. *The Tree of Knowledge: The Biological Roots of Human Understanding*. Shambhala, 1992.

Merleau-Ponty, Maurice. *Phenomenology of Perception*. Routledge, 2014.

"Mikhail M. Bakhtin: 1895-1975." *The Norton Anthology of Theory and Criticism*, edited by Vincent B. Leitch et al, Norton, New York, NY, 2010, pp. 1072–1075.

Moati, Raoul. *Derrida/Searle : Deconstruction and Ordinary Language*. Translated by Timothy Attanucci and Maureen Chun, Columbia University Press, 2014, https://doi.org/10.7312/moat16670.

Moran, Dermot. *Introduction to Phenomenology*. Routledge, 2007.

Morton, Timothy. *The Ecological Thought*. Harvard University Press, 2012.

Newen, Albert, et al., editors. "4E Cognition: Historical Roots, Key Concepts, Central Issues." *The Oxford Handbook of 4E Cognition*. Oxford University Press, 2018. https://doi.org/10.1093/oxfordhb/9780198735410.013.1

Nixon, Rob. *Slow Violence and the Environmentalism of the Poor*. Harvard University Press, 2013.

Noë, Alva. *Action in Perception*. The MIT Press, 2006.

Palmisano, Annalisa, et al. "Face pareidolia is enhanced by 40 Hz transcranial alternating current stimulation (tacs) of the face Perception Network." *Scientific Reports*, vol. 13, no. 1, 4 Feb. 2023, pp. 1–15, https://doi.org/10.1038/s41598-023-29124-8.

Papineau, David. "Teleosemantics." *How Biology Shapes Philosophy: New Foundations for Naturalism*, edited by David Livingstone Smith, Cambridge University Press, 2017, 95-120.

Passingham, Richard. *Cognitive Neuroscience: A Very Short Introduction*. Oxford University Press, 2016.

Peirce, Charles Sanders. "How to Make Our Ideas Clear." *Peirce on Signs: Writings on Semiotic by Charles Sanders Peirce*. Edited by James Hoopes, e-book, The University of North Carolina Press, 2014, *Project Muse*. pp. 160-179.

Pinker, Steven. *The Language Instinct: How the Mind Creates Language*. Harper Perennial, 2007.

Putnam, Hilary. "Meaning and Reference." *The Journal of Philosophy*, vol. 70, no. 19, 1973, pp. 699–711. *JSTOR*, https://doi.org/10.2307/2025079. Accessed 19 Sept. 2024.

Rankin, Catharine H., et al. "Habituation Revisited: An Updated and Revised Description of the Behavioral Characteristics of Habituation." *Neurobiology of Learning and Memory*, vol. 92, no. 2, 2009, pp. 135–138., https://doi.org/10.1016/j.nlm.2008.09.012.

Rey, Georges. "Remembering Jerry Fodor and His Work." *Mind & Language*, vol. 33, no. 4, 2018, pp. 321–341., https://doi.org/10.1111/mila.12206.

Robert, Jason Scott, et al. "Bridging the Gap between Developmental Systems Theory and Evolutionary Developmental Biology." *BioEssays*, vol. 23, no. 10, 2001, pp. 954–62, https://doi.org/10.1002/bies.1136.

Sagan, Dorian. "Introduction." *A Foray into the Worlds of Animals and Humans: With a Theory of Meaning* by Jakob Von Uexküll. Translated by Joseph D. O'Neill, University of Minnesota Press, 2010.

Saussure, Ferdinand de. *Course in General Linguistics*. Translated by Wade Baskin, Columbia University Press, 2011.

Segundo-Ortin, Miguel, and Daniel D. Hutto. "Similarity-Based Cognition: Radical Enactivism Meets Cognitive Neuroscience." *Synthese (Dordrecht)*, vol. 198, no. Suppl 1, 2021, pp. 5–23, https://doi.org/10.1007/s11229-019-02505-1.

Shea, Nicholas. "Developmental Systems Theory Formulated as a Claim About Inherited Representations." *Philosophy of Science*, vol. 78, no. 1, Jan. 2011, pp. 60–82. *JSTOR*, https://doi.org/10.1086/658110.

Shklovsky, Viktor. "Art as Technique." *Russian Formalist Criticism: Four Essays*, translated by Lee T. Lemon and Marion J. Reis, University of Nebraska Press, 2012, pp. 3–24.

Spector, Nicole. "Why Our Sense of Time Speeds up as We Age - and How to Slow It Down." *NBCNews.Com*, NBCUniversal News Group, 26 Nov. 2018, www.nbcnews.com/better/health/why-our-sense-time-speeds-we-age-how-slow-it-ncna936351.

Surette, Leon. "Deconstruction: A Misprision of Saussure and Charles Sanders Peirce." *Philosophy and Literature*, vol. 43, no. 2, 2019, pp. 411–40, https://doi.org/10.1353/phl.2019.0021.

Sutherland, Thomas. "The Contortions and Convolutions of the 'Speculative Turn.'" *Diacritics*, vol. 49, no. 1, 2021, pp. 108–26, https://doi.org/10.1353/dia.2021.0005.

Suvin, Darko. *Metamorphoses of Science Fiction: Studies in the Poetics and History of a Literary Genre*. Yale University Press, 1979.

Thompson, Evan. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press, 2007.

---. "Life and Mind: From Autopoiesis to Neurophenomenology." *Emergence and Embodiment: New Essays on Second-Order Systems Theory*. Edited by Bruce Clarke and Mark B.N. Hansen, Duke University Press, 2009.

"Topology, N." *Oxford English Dictionary*, Oxford UP, July 2023, https://doi.org/10.1093/OED/8435562293.

Uexküll, Jakob Von. *A Foray into the Worlds of Animals and Humans: With a Theory of Meaning*. Translated by Joseph D. O'Neill, University of Minnesota Press, 2010.

Varela, Francisco J., et al. *The Embodied Mind: Cognitive Science and Human Experience*. Rev. ed., The MIT Press, 2016.

Vološinov, Valentin N. *Marxism and the Philosophy of Language: V.N. Vološinov*. Translated by Ladislav Matejka and I.R. Titunik, Harvard Univ. Press, 1986.

Wagemans, Johan, et al. "A Century of Gestalt Psychology in Visual Perception: I Perceptual Grouping and Figure–ground Organization." *Psychological Bulletin*, vol. 138, no. 6, Nov. 2012, pp. 1172–217. *EBSCOhost*, https://doi-org.offcampus.lib.washington.edu/10.1037/a0029333.

Westling, Louise Hutchings. *The Logos of the Living World: Merleau-Ponty, Animals, and Language*. Fordham University Press, 2014.

Wittgenstein, Ludwig. *Philosophical Investigations*. Translated by G.E.M. Anscombe et al., 4th ed., Wiley-Blackwell, 2009.

Zavadskii, K. M., et al. "Engels and Darwinism." *Soviet Studies in Philosophy*, vol. 10, no. 1, 1971, pp. 63–80, https://doi.org/10.2753/RSP1061-1967100163.