The Emergence of Mind: *Personal Knowledge* and Connectionism

Jean Bocharova

Key words: Connectionism, Michael Polanyi, emergence, mind, neurocomputation, PDP

ABSTRACT

At the end of Personal Knowledge*, Polanyi discusses human development, arguing for a view of the human person as emerging out of but not constituted by its material substrate. As part of this view, he argues that the human person can never be likened to a computer, an inference machine, or a neural model because all are based in formalized processes of automation, processes that cannot account for the contribution of unformalizable, tacit knowing. This paper revisits Polanyi's discussion of the emergence of consciousness and his rejection of neural models in light of recent developments in connectionism. Connectionist neural modeling proposes an emergentist account of brain structure and, in many ways, is compatible with Polanyi's philosophy, even if it ultimately neglects questions of meaning.*

In his discussion of evolution in "The Rise of Man" at the end of *Personal Knowledge*, Polanyi touches on the emergent properties of human development. He argues in this section that the movement from embryo to fully developed human person cannot be explained either as mere preprogrammed maturation or as the result of an "external creative agency" (395). Rather, human development involves something he calls the "intensification of individuality" (395). According to this view, stages of development—new achievements of a developing human person—arise in a manner similar to the emergence of new scientific discoveries: both processes require the crossing of a "gap," a heuristic gap in the case of the scientist or an ontological gap in the case of the human person. Just as the scientist strives toward a truth that can only be intimated, so too does the infant passionately strive toward an achievement yet to be realized but intimated as possible. The result of such striving is the emergence of personhood, achieved most fully when a child enters into the "traditional noosphere," his or her culture's "lasting articulate framework of thought" (388). For Polanyi, this intensification of individuality at the level of human development is consistent with his view that higher-order structures and characteristics of the human mind are not predetermined in the material substrate of biology but emerge indeterminately as a result of an individual's personal commitment (395-397). In this way, his view of the emergence of human consciousness is part of his larger refutation of a Laplacean conception of the universe as reducible to the laws of physics and chemistry.

Related to Polanyi's discussion at the end of *PK*, the concept of emergence has recently begun to gain prominence among cognitive neuroscientists who model brain function using connectionism. Connectionist models of brain architecture assume that higher-order cognitive functions can only be understood globally in terms of patterns of activity distributed over multiple connections in the brain. In this sense, and for readers familiar with Juarrero's work, it could almost be called a dynamical systems approach to human cognition (e.g., McClelland et al. 2010). Connectionism stands in opposition to "grandmother cell" theories that try to locate thoughts in specific neurons or groups of neurons; representational nativists (e.g., Pinker and Chomsky) who argue that humans are born with significant domain-specific knowledge located in specialized, predetermined modules in the brain; probabilistic models of cognition, which advocate a top-down modeling approach to study cognitive processes; and other computational theories of mind (e.g., Fodor and Pylyshyn) which argue that the brain operates like a digital computer. In contrast to these other theories and approaches, connectionists argue that the complex brain architecture of an adult is emergent from simpler neural structures (Rumelhart 1987; McClelland et al. 2010, 348; McClelland 2011, 134). These structures, rather than being pre-programmed to mature a certain way or to specialize for pre-determined functions, acquire their abilities by encountering inputs in their environment (McClelland 2010, 753). As a revision of the brain-as-computer metaphor, connectionism informs many of the most prominent contemporary discussions about the mind-body relation. It provides a foundation for neurophilosophy, eliminative materialism, embodied cognition, dynamic core theory, and work in artificial intelligence.[1] Because connectionism is so fruitful in the cognitive sciences, it is worth considering how it might agree with or depart from Polanyi's understanding of emergence, especially as it relates to his larger arguments about human development and the relationship between mind and body. Below I explain how many of Polanyi's objections to the neural model in *PK* do not apply to current neural models based on connectionist assumptions. Connectionism, which favors pattern recognition rather than logic as a descriptor of cognitive processing, agrees with many of Polanyi's points about the nature of tacit knowing. Despite such agreement, however, there remains a divergence regarding the status of the human person as an active center.

Connectionism: An Overview

Connectionism traces its origins to the 1940's with the McCulloch and Pitts neural model, but it is generally understood to have begun to take its current form in the 1980's as a result of studies in artificial intelligence (McCleod et al. 1998, 314). It was at this time that David Rumelhart, James McClelland, Geoff Hinton and others developed computer models of brain function that operated through parallel distributed processing (PDP). PDP involves many small "neuron-like" units operating simultaneously over a multi-layered network.[2] In these networks, information is not carried in whole chunks (such as binary units of 1 or 0). Instead, it is conveyed as a pattern of activity among many units. For example, whereas a localist or symbolic representational system might assign a whole concept, such as "dog," to a single neuron, a small group of neurons, or a single unit in a computer network, distributed systems do not represent or store such a concept in any single place. Rather, a representation of "dog" would arise from a pattern of activity among many different units, units which are also used to represent other concepts, like cats or coyotes (Elman et al. 1999, 90-91). This pattern of activity is generated and stored as a potential in the weights between connections in the network. These weights reflect the probability that a unit will activate given various levels of input (McClelland 2000, 583).

Thus, connectionism treats the human brain primarily as an information processor. But unlike other brain-as-computer theories, connectionism rejects the notion that the brain operates through symbolic processing, with preprogrammed and sequential steps, local storage of memory, and discrete packets of information. Instead, they propose that it is more likely that the brain operates through weighted connections that store and generate information over a distributed network, with units operating in parallel and with larger systems emerging from simpler architectures (Elman et al. 1999, 50-56).[3]

Since the 1980s, parallel distributed processing models of cognitive function have shown that significant cognitive tasks, such as learning the meaning of words and identifying similarities and differences between objects, can be performed by multiple simple units working in parallel in layered networks (Rumelhart and Todd 1993, 14-15; Elman 1990, 200). Much current work in connectionism focuses on modeling human learning and development, and researchers in the field have built computer models that mimic how humans acquire and perform higher-order cognitive tasks such as learning how to pronounce words (Plaut et al. 1996), learning advanced rules of syntax (Elman 1993), recognizing faces from multiple angles (McLeod et al.. 1998, 294-300), predicting the effects of weight and distance on a balance beam (McClelland 1989), and reconstructing a whole image from a partial association (Hertz et al. 1991). To demonstrate the biological plausibility of the models, the performance of these networks is often compared with studies in human cognition, including studies of children's acquisition of similar tasks as well as studies of patients with brain damage. More recently, researchers at The Neurosciences Institute have developed brain-based devices (BBDs) with artificial nervous systems that allow these devices to learn about their environment and to navigate in response to it. As these devices learn about their environment, each one develops in its artificial brain a unique activity pattern between connections (Edelman 2006, 131-141).

In short, connectionists argue that many higher-order cognitive functions—as well as physiological structures (e.g., specialized brain regions) and developmental processes (e.g., critical periods)—are emergent (McClelland 2010, 751). In these theories, the term "emergence" is assumed to be the appearance of unpredictable structural novelty, especially those with causal potency. Complex abilities, structures, and processes involve a large number of simple elements that are shaped into a new structure by dynamic, contextual forces (McClelland 2010, 754). The novel form cannot be reduced to any of its constituent elements; it is dependent on but more than the sum of its parts. A complementary assumption is that new structures by their very nature are not designed. Although connectionists allow for initial constraints to be placed on individual components of a complex system (e.g., a learning rule or simple starting architecture), they argue that these initial constraints cannot determine in themselves what the final emergent form will be. The neurons that will be involved in processing language, for example, are not specialized for this task until they find themselves in an environment saturated with linguistic input (Elman et. al. 1999, 265-267). The complexity of a network's final configuration is not pre-programmed.[4]

Polanyi's Critique of the Neural Model

These dynamic features of connectionist networks distance contemporary neurobiology from Polanyi's critique of the "neural model" of mind. In *PK*, Polanyi objected to a neural model of the human mind for two reasons. The first reason is that a neural model, by focusing on the subsidiary contribution of the body to the comprehensive entity which is the mind, destroys the thing it is trying to understand. Just as reducing life to the laws of physics and chemistry is meaningless, he writes, "it is likewise meaningless to represent mind in terms of a machine or of a neural model" (*PK*, 382). I discuss this first critique toward the end of the paper. The second reason, which is more prominently discussed in *PK*, is that neural models are characterized as automated, formalized, specified systems of explicit inference, systems incapable of the tacit inferences that characterize human thought. A closer look at Polanyi's understanding of a "neural model," however, demonstrates that it differs from connectionist assumptions about the nature of neurocomputation.

When Polanyi refers to neurology or the "neural model" in *PK*, he consistently likens it to a symbolic processing machine. He objects to the neurological model precisely because he objects to those who would see the mind as a kind of computer. But in *PK*, "computer" means *digital* computer, which is significantly different than a computer working on a parallel distributed processing model. The chief problem that Polanyi identifies with digital computers, inference machines, and automation in general is that they try to formalize and specify the unformalizable and the unspecifiable; in so doing they eliminate the tacit coefficient that constitutes the active center of a human person (*PK*, 257-8). Automation, if it is cleverly designed, may seem intelligent, but this intelligence is illusory: it can never be more than the manipulation of symbols through the use of prescribed rules specified and formalized by its program. Because rules are fed to the machine in advance, there is no room, he says, for unformalizable, unspecifiable (i.e. tacit) components to enter into its operation: "A routine game of chess can be played automatically to the extent to which the rules of art can be specified. While such a specification may include random elements, like choices made by spinning a coin, no unspecifiable skill or connoisseurship can be fed into a machine" (*PK*, 261). By extension, neural models, for Polanyi, also operate by "fixed symbolic operations" (*PK*, 337), assuming pre-determined rules that foreclose the possibility of unspecified, tacit contributions to human life. He calls this the "automatic neurological model" (*PK*, 262), a model that, should it admit of human consciousness, could only allow for a consciousness that was superfluous to the automatic functioning of a nervous system over which it could exert no influence (*PK*, 336). In short, Polanyi rejects neurology because it is "based on the assumption that the nervous system—*functioning automatically* according to the known laws of physics and chemistry—*determines* the workings which we normally attribute to the mind of an individual" (*PK*, 262; emphasis mine).

Like Polanyi, connectionists also reject the metaphor of mind as a digital computer. Unlike digital computers—which run input symbols through a series of rules in a formal program in order to reach an output—connectionist networks do not work by symbols or deterministic programs. Instead, patterns of activation trigger other patterns of activation that lead to an output. Over time, as the network encounters common patterns in the environment, some pathways between connections are strengthened while others are weakened. Whereas digital computers are characterized by "fixed symbolic operations," the connectionist network that has discovered structure in its environment is not the same as the one that was originally built by a human technician. Similarly, unlike most who use the metaphor of the mind as digital computer, connectionists also reject the claim that genes "code" for higher-order cognition or complex neural structures. Instead of likening genes to a "computer program" or a "blueprint," they use the metaphor of genes as "catalysts" that enable—but do not direct—the emergence of new structures and processes (Elman et al. 1999, 350-351). In this sense, connectionism agrees with Polanyi's view that DNA "*evokes* the ontogenesis of higher levels, rather than *determining* them" (*KB*, 235; emphasis original).

Emergence

Given this brief sketch of connectionist perspectives, the question now arises: do connectionist networks achieve "higher levels" of being in the sense that Polanyi means it? True emergence, for Polanyi, is more than mere holism: the irreducibility of emergent entities, he writes, "must not be identified with the mere fact that the joining of parts may produce features which are not observed in the separate parts" (*KB*, 230). Instead, emergence requires the addition of higher supervening principles—boundary conditions that harness lower levels of being. For a connectionist model to be truly emergent, it would have to be under dual control, with a higher operational principle harnessing the boundary conditions left open by lower levels. In this section, I want to suggest that the set of weights found in a trained network may constitute such an operational principle. Just as personal acts of knowing comprise the tacit integration of subsidiaries into focal wholes, so too do networks integrate subsidiary inputs into global patterns of activity. Higher order operational principles emerge when networks cross their own conceptual "gap," a point in the training where the network achieves the ability to better discriminate between inputs. The set of constraints found in a trained network enable it to complete tasks which other networks with random weight assignments cannot complete.

The possibility of a network's emergence due to dual control can be seen in connectionist explanations of the mechanics underlying sudden insights and new developmental achievements. It is worth here recalling that Polanyi sees the process of scientific discovery—a process marked by sudden insight as one crosses a heuristic gap—as paradigmatic for tacit knowing (*TD*, 24-25) and for the emergence of mind (*PK*, 395). The new interpretative framework that arises when we cross a heuristic gap is "of the same kind" as the emergence of individual personhood, a process undergone in stages and culminating in the emergence of human culture (*PK*, 395). Personal commitment leads to a passionate striving toward a hidden reality and to the eventual achievement of discovery (*TD*, 25; *PK*, 311-312). Because such striving and achievement involves personal commitment, the process of scientific discovery and other forms of emergence related to the human person cannot be formalized or prespecified (*TD*, 25; *PK*, 311-312). Instead, we integrate subsidiaries informally, achieving focal awareness and understanding in an unspecified way. Most striking is how connectionism suggests that this process of tacit integration can be observed in the global patterned activity of networks processing inputs across a large number of small (subsidiary), weighted connections.

Tacit integration of this kind can be seen in connectionist explanation of sudden learning. Like humans moving from one phase of knowledge to another, where they may at first not understand or only partially understand a concept before gaining mastery, artificial connectionist networks sometimes demonstrate sudden, unexpected advancements in their ability to process information. Such complex outcomes, however, are the result not of new enabling interventions or dramatically new learning mechanisms but rather of simple learning mechanisms making small changes over time. One example is a network developed to identify the sum of a sequence as even or odd. After 17,999 training cycles, the network was limited in its ability to identify whether the sequence was odd or even and had not been able to extract a general principle. On the very next training cycle, cycle 18,000, the network was suddenly able to give the right answer in all instances (Elman et al. 1999, 230-236). Looking from the outside it would seem that the network passed through a dramatic stage in its development—or that it crossed a heuristic gap. This jump in ability, however, was the result of a single learning algorithm making a large number of small adjustments to connection strengths. These small changes amounted not to a linear change in output—with each adjustment corresponding to an observable change in the system's behavior. Rather, the small adjustments to weigh strengths resulted in non-linear change at output. The network's behavior seemed not to be affected until a critical threshold of change had been passed, after which point outputs changed significantly. Another example of sudden, seemingly dramatic learning can be found in a network developed by Plunkett and Marchman to learn vocabulary. Like children, who exhibit sudden vocabulary spurts around 22 months, the network also demonstrated distinct jumps in its ability both to produce and comprehend vocabulary (Plunkett et al. 1993). As with the first network, the large change in output did not occur as a result of a large change or addition to the network's basic learning mechanisms. In both cases, and as is common in dynamical systems (Juarerro 1999, 123-125), small incremental changes led to complex, seemingly sudden and dramatic changes in the systems' behavior.

For Polanyi, the child's ability to speak represents an emergent achievement. Through the intensification of his or her own individuality, higher-level operational principles take control, allowing the knower to form meaningful coherences out of subsidiaries—in this example, to distinguish and produce meaningful utterances from a collection of various sounds (*KB*, 233, 235). It is the child's own operation of tacit knowing that allows him or her to move from unbounded subsidiaries (mere sound) to meaningful coherences (words) controlled by new operational principles. What exactly happens then when an artificial network gains the ability to distinguish words from sounds? Like other machines, such networks seem to be working according to a higher operational principle irreducible to physics and chemistry; like living beings, they seem to have achieved such higher operational principles through an emergent process. Given the absence of passionate striving and personal commitment, however, we cannot say that such a network has achieved something in the same way that living beings do. Yet clearly something has happened when a network that has not been programmed to do so acquires an ability to make conceptual distinctions and to formulate conceptual prototypes. Like Polanyi's description of the developing child at the end of *PK*, whatever has happened to the network is not the result of a program running through its operations or of an external agent intervening at every step. Polanyi describes the first view as the belief that a genetic program already contains all of the instructions for the mature adult and that maturation is simply the carrying out of these prespecified directions. This view is not shared by connectionist theories of human cognition or by connectionist modelling. Although connectionist networks begin with basic starting architecture (cf. basic wiring of neurons in the brain) and are equipped with a learning algorithm (cf. properties of neurons that allow them to strengthen or weaken connections with other neurons), it would not be accurate to say that a network's successful performance is the result of its simply running its program. Nor can it be said that the programmer intervenes and adjusts the network to meet the task at hand. The weights of an "immature" connectionist network respond to stimuli it encounters in the environment. Although programmers "intervene" by providing the network with a training environment, it is significant that they do not actually rewire connections or rewrite programs to help the network perform. The same algorithm used in one task could be used by another network performing a very different task. The formative variable is the training environment.

Are these networks then "determined by" their training environment? Perhaps, and I will revisit this question below**.** For now, suffice it to say that the new weight space that develops in a connectionist network could be said to function as a new operational principle bringing the system under dual control. That is, the new weight configurations of a trained network, a set of constraints that processes information predictably and using a pattern it had not previously used, harness the principles left open by the boundary conditions of the network's lower levels of being, the *random activation* that precedes the trained network's *patterned activation*. Thus, the trained network would seem to exist, in Polanyi's terms, as an emergent from the untrained network.

It should not be surprising then to see connectionists themselves noticing those characteristics of thought and intelligence that Polanyi considers throughout his writings. In an article about the topic of emergence in cognitive science, James McClelland, one of the leading original members of the PDP research group that first developed connectionist networks, discusses the possibility of the emergence of consciousness in a way that evokes Polanyi's concepts of achievement, success, convivial societies, heuristic striving, and tacit knowing. The similarities are so striking that McClelland (2010, 752-753; emphases mine) deserves to be quoted at length:

> But, in fact, these simple regularities [emergent forms of dynamic systems] are not the essence of intelligence or the supreme *achievements* of nature. When it comes to intelligence, the real stuff consists of *human success* in everyday acts of perception, comprehension, inductive inference, and real-time behavior—areas where machines still fall short after nearly 60 years of effort in artificial intelligence—as well as the brilliant creative intellectual products of scientists and artists such as Newton, Darwin, Einstein, Shakespeare, Michaelangelo [sic], and Beethoven. According to an emergentist perspective, all of these products of the mind are essentially emergents. I do not think anyone who emphasizes the importance of emergent processes would deny that planful, explicitly goal-directed thought plays a role in the greatest human intellectual achievements. However, such modes of thought themselves might be viewed as *emergent consequences of a lifetime of thought-structuring practice supported by culture and education* (Cole & Scribner, 1974). Furthermore, proponents of the essence of human thought as an emergent phenomenon might join with Hofstadter (1979) and others in suggesting that key flashes of insight and intuition may not have arisen from planful, explicit goal-directed thought alone, but instead might reflect a massive subsymbolic constraint-satisfaction process taking place *outside of awareness*. In the case of Darwin, for instance, biographers (e.g., Quammen, 2006) have written about the origins of his work on his theory of evolution. It appears that Darwin set his mind to this investigation

> *knowing intuitively that there was something interesting to discover, while not knowing exactly what it was*. This intuition, arguably the key factor in his discovery, might have arisen as an emergent consequence of a subconscious constraint-satisfaction process, which then led him to engage in more intentional (yet still perhaps intuition-guided) exploration. This sequence in discovery may be the rule even in formal domains such as mathematics and physics, where the *intuition may come first, followed only later by formal specification and rigorous proof* (Barwise & Etchemendy, 1991).

For McClelland, explicit inferences, conscious thought, and scientific discovery may arise from the tacit workings of an individual who has lived and worked in a convivial society: They are "emergent consequences of a lifetime of thought-structuring practice supported by culture and education." The bulk of this paragraph sounds as if it were written by someone familiar with Polanyi's writings, with the exception of the phrase "sub-symbolic constraint satisfaction process"—and here lies the crucial departure from Polanyi. A central—perhaps *the* central feature of Polanyi's theory of personal knowledge and tacit knowing is the existence of an "active center," a human person who strives towards that which is only intimated. In McClelland's quote above, we see the hallmarks of this active center: achievements, successes, heuristic striving, intuitive intimation of future discovery. But in the place of the active center out of which these things proceed, we get a sub-symbolic constraint satisfaction process. What emerges in these models is a set of probabilistic constraints governing the interaction of physical neurons and groups of neurons.

Gradients of Probability and "Enlarged Laws of Nature"

How might such a vision of emergence correspond to Polanyi's understanding of the emergence of an active center? The idea that probabilistic constraints may play a role in the emergence of mind is not in itself antithetical to Polanyi's view and, in fact, plays a major role in his writing. The distinction between the emergence of machine-like principles in nature and the emergence of mind involves the relationship of the emergent to its gradient of probability. Returning to his discussion of DNA, for example, we see this gradient of probability working as a causal force. Here he argues that if DNA is to be seen as a blueprint for a complex machine, the machine's growth "seems to require a system of causes not specifiable in terms of physics and chemistry, such causes being additional both to the boundary conditions of DNA and to the morphological structure brought about by DNA" (*KB*, 231-2). That system of causes responsible for the development of the embryo, he describes variously as an "integrative power," "field-like powers," and "a gradient of potential shapes."

It is significant here that Polanyi refers to Waddington's epigenetic landscape, which depicts a marble in a hilly landscape moving toward a wall (Waddington 1957, 67). He says that these images "show graphically that the growth of the embryo is controlled by the gradient of potential shapes, much as the motion of a heavy body *is controlled by* the gradient of potential energy" (*KB*, 232; emphasis added). In other words, as a heavy body at the top of a hill will, with minimal force, be pulled to the bottom, so will an embryo be pulled towards the most probable shapes along the course of its development.[5]

This process of being pulled along a gradient of greatest probability is described in connectionist terminology as descending along an error gradient or settling into a weight space. They too refer to Waddington, noting the similarity between his landscapes and their graphic representation of a network error surface. A network moving toward a solution to a problem in a weight space is said to be seeking "low ground" in its error surface—the point in the landscape where errors will be minimal (Elman et al. 1999, 17-18). The descent of a network along an error gradient as it arrives closer to achieving a solution to a problem set resembles Polanyi's description of how a knower makes contact with reality in a heuristic field:

> We assume that the gradient of a discovery, measured by the nearness of discovery

> prompts the mind towards it…. The assumption of a heuristic field explains now how it is possible that we acquire knowledge and believe that we can hold it, though we can do this only on evidence which cannot justify these acts by any acceptable strict rules. It suggests that we may do so because an innate affinity for making contact with reality moves our thoughts—under the guidance of useful clues and plausible rules—to increase ever further our hold on reality (*PK,* 403).

For connectionists, this "innate affinity" is the computational properties of neurons. Under both models—connectionist error surface and Polanyi's heuristic field—making contact with objective reality is described as descending along a gradient. For connectionists, it is the potential of the initial weight conditions, the network's learning algorithm, the quality of inputs, and the changing internal constraints of the developing set of weights that pull a network toward low points. For Polanyi, however, there comes a point in the development of the embryo where a higher operational principle—the individual's active center—harnesses this movement in the field of potentialities so that it no longer moves passively along lines of force but actively seeks its own direction. Gradients of probability, he says, control the development of the *embryo*, but not the active center of an individual. For the latter, such gradients only go so far as to open up possible opportunities for behavior, without determining or controlling that behavior (*PK*, 403).

It is unclear how connectionist theories would respond to Polanyi's distinction between entities passively following a line of force along a probability gradient and those—like the human person—which actively traverse their environment. Even taking into account the role of modulatory and attentional systems, it would be difficult to reconcile with connectionism this jump from passive movement to active movement. Active movement in the sense that Polanyi means it would simply be a much more complex network following lines of probability, no more active than simpler networks. The emergence of a complex brain which "speaks mainly to itself" (to use Gerald Edelman's words [2006, 20]) would still seem, from the perspective of connectionism, to be largely the product of a unique dynamic system responding to inputs in the environment, inputs which exist along too many dimensions for an error landscape to fully represent. Despite the seemingly deterministic role that the training environment still seems to play in shaping these networks, however, connectionist "emergence" does seem to correspond to Polanyi's intimation of "enlarged laws of nature," laws that would enable the rise of consciousness in biological material and that would allow consciousness to shape that material for its own ends: "Since action and reaction usually arise together in nature, it would seem reasonable, on the contrary, that the new laws of nature, which would allow for the rise of consciousness in material processes, should also allow for the *reverse* action, that is, of conscious processes acting on their material substrate" (*PK*, 397). In viewing the brain as a complex, dynamical system, connectionist thought suggests that as higher order cognitive functions emerge they direct and shape the very material on which they depend—that is, the physical synapses that carry and store information—by choosing among alternative behaviors, beliefs, and interpretive frameworks (Edelman 2006, 95). In this way, the nonlinear dynamics of connectionist networks do seem to use "enlarged laws of nature" that make possible a third alternative to the unwelcome views of human development as either genetically pre-determined or as requiring external intervention at every stage (*PK*, 395).

Despite many points of confluence, prominent connectionists who speak about the existence of mind tend to make statements antithetical to Polanyi's philosophy. At one end of the spectrum, for example, the Churchland's eliminative materialism posits that concepts used in "folk psychology" (e.g., desire, belief, fear, intention) can and should be eliminated in favor of material, neurobiological explanations (Churchland 1992, 6). At the other end of the spectrum, Gerald Edelman (2006), an anti-reductionist, describes consciousness as the activity of the brain's dynamic core. For him, our thoughts and feelings are entailments of brain states: just as the spectrum of hemoglobin is not separated from but entailed in the molecular structure of hemoglobin, qualia are non-causal but faithfully informative of our brain states (91-92). Despite the remarkable strides made by connectionism and the fruitfulness of it as an explanatory theory, the central problem that Polanyi identifies still remains: so long as we focus on mechanics, including the dynamics of a complex system, the human person disappears. Neurobiological explanations account for failures of cognition and provide a detailed explanation of the conditions that it requires. But they cannot define success. The meanings we find in the world—a skillful dance performance, a word, the expression on a familiar face—are real, though artificial. They are created by a knower who constructs a coherent mental representation out of subsidiaries. Just as these realities disappear when we focus on their subsidiary components, so too does the mind when we focus on neurobiology.

Thus, perhaps the biggest point of divergence between Polanyi and various strands of connectionist thought is located in Polanyi's articulation of the relationship between knowing and being. Scientists who see themselves as endeavoring to strip away illusions or as trying not to be fooled by appearances have already subscribed to a generative metaphor—one that places them in an imagined position of mastery, or potential mastery, over both the untrustworthy shows of nature and the masses who are still enchanted. This knower-as-master contrasts with the knower who gropes around seeking to touch or make contact with reality. Polanyi's assertion that we know by dwelling in a comprehensible object or process thus points to the inadequacy of descriptive explanation as a full map of reality. If, as connectionists say, our physical bodies change in response to what we encounter in our environment, connections become stronger or weaker and neuronal structures are built and dismantled, we might call these changes a form of indwelling, as information takes on physical (non-linguistic, sub-symbolic) form in the brain's microcircuitry. We may then know through indwelling as these subsidiary connections register high-dimensional discriminations. But, as Polanyi notes, what we know, what interests us, is something other than mere prototype vectors. If connectionists are right that our conscious thoughts and feelings correspond in some way to brain networks that register discriminations among a very large number of dimensions, then the descriptive language used in the sciences is inadequate to the task of fully recording and replicating such fine-grained, quickly changing activity. Attempts to articulate our knowledge, including attempts by scientists, will always leave out many dimensions of experience, including the experience we are trying to describe—we will always know more than we can say. As creatures who know through indwelling, we must articulate those truths we have come to know by drawing on multiple forms of expression, including figurative language, music, image, gesture, and even silence. Thus, it may not be as helpful to talk about emergent *things*, as if an emergent must be fully described and describable. Instead, we ought to look (as Polanyi does) at emergence as a process of knowing, whereby the many dimensions of being unfold, shift, open themselves up to us in ways that exceed our ability to capture fully in language. Regarding the mind-body question, Polanyi's post-critical philosophy continues to challenge us both to embrace new discoveries of neurobiology and to reject the idea that they are sufficient to explain what we are as humans and how we ought to live.

ENDNOTES

[1]See, for example, on neurophilosophy Churchland (1989); on eliminative materialism, Churchland (1992); on embodied cognition, Lakoff and Johnson (1999) and Feldman (2006); on dynamic core theory, Edelman (1987); on artificial intelligence, Smolensky (1987).

[2]Rumelhardt (1987) describes his frustration with forms of AI based on symbolic processing as a motive for his looking to the brain for inspiration. If we took seriously the notion that the brain is a computer, he asked, what kind of computer would it be and how would it process information? The fruitfulness of connectionist models is found in their radically different method of processing information, not in a claim

for strict biological plausibility. Artificial connectionist networks are not intended to correspond exactly to neurons or neuronal groups, which exist on a scale much greater than any model that can be currently built and which involve the interaction with various modulatory hormonal systems. Rather, these networks are supposed to process information in ways that more closely resemble the operations of the brain than other computational theories of cognition based on a Turing-Machine, von Neumann, digital-computer model.

[3]Connectionism does not foreclose the possibility of emergent higher-order structures working in ways that appear to be rule governed. As such, they are not opposed to computational models that try to capture these phenomena (McClelland et al. 2010, 349-350; Feldman 2006). The disagreement arises when models of larger brain structures or processes are used as the bases of explanatory theories (McClelland et al. 2010, 350). The connectionist position on modularity is, perhaps, best summed up by Elman et al. (1999) in the following: "domain-specific representations can emerge from domain-general architectures and learning algorithms and . . . can ultimately result in a process of modularization as the end product of development rather than its starting point" (115).

[4]For an account of how human brains might have evolved to take on initial architectural constraints rather than genetically determined brain regions see Elman et al. 1999, 238-317; McLeod et al. 1998, 303-313**;** Calabretta and Parisi 2005.

[5]See David Agler (2014) for an extended discussion of Polanyi's references to experimental embryology and its influence on his writings on emergence.

REFERENCES

Agler, David. 2014. "Emergence from Within and Without: Juarrero on Polanyi's Account of the External Origin of Emergence." *Tradition and Discovery: The Polanyi Society Periodical.* 40 (3): 23-35.

Barwise, Jon and John Etchemendy. 1991. "Visual Information and Valid Reasoning." In *Visualization in Teaching and Learning Mathematics*, edited by Walter Zimmerman and Steve Cunningham, 9-24. Washington, DC: Mathematical Association of America.

Calabretta, Raffaele and Domenico Parisi. 2005. "Evolutionary Connectionism and Mind/Brain Modularity." In *Modularity: Understanding the Development and Evolution of Complex Natural Systems*, edited by Werner Callebaut and Diego Rasskin-Gutman, 309-330. Vienna Series in Theoretical Biology. Cambridge: The MIT Press.

Churchland, Patricia. 1989. *Neurophilosophy: Toward a Unified Science of the Mind-brain.* Cambridge, MA: The MIT Press.

Churchland, Paul. 1992. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge, MA: The MIT Press.

Cole, Michael and Sylvia Scribner. 1974. *Culture and Thought: A Psychological Introduction*. New York: Wiley.

Edelman, Gerald. 1987. *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books.
_____. 2006. *Second Nature: Brain Science and Human Knowledge*. New Haven: Yale University Press.

Elman, Jeffrey L. 1990. "Finding Structure in Time." *Cognitive Science* 14: 179-211.

_____. 1993. "Learning and Development in Neural Networks: The Importance of Starting Small." *Cognition* 48: 71-99.

Elman, Jeffrey L., Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett. 1999. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge. MA: The MIT Press.

Feldman, Jerome. 2006. *From Molecule to Metaphor: A Neural Theory of Language*. Cambridge, MA: The MIT Press.

Hertz, John, Anders Krogh, and Richard Palmer. 1991. *Introduction to the Theory of Neural Computation*. Addison-Wesley, Redwood, CA.

Hofstadter, Douglas. 1979. *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books.

Juarrero, Alicia. 1999. *Dynamics in Action: Intentional Behavior as a Complex System*. Cambridge, MA: The MIT Press.

Lakoff, Geroge and Mark Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

McClelland, James L. 1989. "Parallel Distributed Processing: Implications for Cognition and Development." In *Parallel Distributed Processing: Implication for Psychology and Neurobiology*, edited by Richard Morris, 8-45. New York: Oxford University Press.

_____. 2000. "Connectionist Models of Memory." In *The Oxford Handbook of Memory*, edited by Endel Tulving and Fergus I.M. Craik, 583-596. New York: Oxford University Press.

_____. 2010. "Emergence in Cognitive Science." *Topics in Cognitive Science* 2: 751–770.

_____. 2011. "Memory as a Constructive Process: The Parallel-Distributed Processing Approach." In *The Memory Process: Neuroscientific and Humanistic Perspectives,* edited by Suzanne Nalbantian, Paul Matthews, and James L. McClelland, 129-151. Cambridge, MA: The MIT Press.

McClelland, James, Matthew Botvinick, David Noelle, David Plaut, Timothy Rogers, Mark Seidenberg, and Linda Smith. 2010. "Letting Structure Emerge: Connectionist and Dynamical Systems Approaches to Cognition." *Trends in Cognitive Sciences* 14: 348–356.

McLeod, Peter, Kim Plunkett, and Edmund T. Rolls. 1998. *Introduction to Connectionist Modelling of Cognitive Processes*. New York: Oxford University Press.

Plaut, David, James McClelland, Mark Seidenberg, and Karalyn Patterson. 1996. "Understanding Normal and Impaired Reading: Computational Principles in Quasi-Regular Domains." *Psychological Review* 103: 56-115.

Plunkett, Kim and Virginia Marchman. 1993. "From Rote Learning to System Building: Acquiring Verb

Morphology in Children and Connectionist Nets." *Cognition* 48: 21-69.

Polanyi, Michael. 1962. *Personal Knowledge: Towards a Post-Critical Philosophy*. Chicago: The University of Chicago Press.

_____. 1966 [2009 re-issue]. *The Tacit Dimension*. Chicago: The University of Chicago Press.

_____. 1969. *Knowing and Being Essays by Michael Polanyi*, edited by Marjorie Grene. Chicago: The University of Chicago Press.

Quammen, David. 2006. *The Reluctant Mr. Darwin*. New York: Norton.

Rumelhart, David E. and Peter Todd. 1993. "Learning and Connectionist Representations." In *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience*, edited by David E. Meyer and Sylvan Kornblum, 3–30. Cambridge, MA: The MIT Press.

Rumelhart, David. 1987. Embedded Video. "Brainstyle Computation and Learning." *The Science Network*. http://thesciencenetwork.org/programs/cogsci-2010/david-rumelhart. 1/13/14.

Smolensky, Paul. 1987. "Connectionist AI, Symbolic AI, and the Brain." *Artificial Intelligence Review* 1: 95-109.

Waddington, Conrad. 1957. *The Strategy of the Genes*. London: Allen & Unwin.