

Emergence, Supervenience, and Personal Knowledge

What, if anything, does recent work on emergence in the sciences, in the philosophy of mind, and in theology have to do with the thought of Michael Polanyi? The connection could, of course, be purely external and *ad hoc*: perhaps the coming few minutes will be one of those polite but pointless academic exercises where the speaker nods his head in the direction of the Polanyi Society and then proceeds to trot out some old paper that advances his well-worn ideas on some completely different topic. To be honest, I wasn't sure in advance of being able to avoid such a charade. But the intuitions of Philip Rolnick, who organized this session, were right on the mark. It *is* a new paper, and it *is* about Polanyi.

“About” does not mean uncritical. Polanyi's work is important for contemporary emergence theory in part because he is so right and in part because he is so wrong. Very recent work on emergence has developed metaphysical resources that now avoid what Polanyi thought was a forced choice between accepting reductionism, which leaves no place for the genuinely personal, and accepting what Polanyi called “finalistic” causes. Moreover, some of what Polanyi thought was good science turned out to be bad science, and it served him badly. But I am getting ahead of myself. Let's begin the story at the very beginning.

Polanyi on Emergence

Once upon a time there was a century dominated by the ideal of reductionism. It was a century in which some of the deepest dreams of science were fulfilled. Building on Newton's laws and Maxwell's equations, scientists achieved a structure capable of handling the very small (quantum physics), the very fast (special relativity, for speeds approaching c), and the very heavy (general relativity, or what you might call gravitational dynamics). Chemistry was, for all intents and purposes, completed (or so some are claiming). Crick and Watson discovered the biochemical information system that codes for all biological reproduction and heritable mutations, and not too many months ago the mapping of the human genome was completed and made public. With rather less success, psychology tried to make itself in the same image, spawning schools such as behaviorism, functionalism, and, more recently, sociobiology and evolutionary psychology. (Ironically, the only social science that has come close to matching the mathematical precision of the natural sciences is economics, the science of money. Small wonder.)

Writing in the heyday of this movement, midway between the British Emergentists of the 1920s and the rebirth of the emergence movement in the 1990's, Michael Polanyi functioned as a sort of prophet. Indeed, to the “cultured despisers” of his time, he must have seemed a little bit like Jeremiah, running through the streets of the lost cities and calling the lost to repent. Indeed,

Personal Knowledge was so out of place when it was first published in 1958, and *The Tacit Dimension* so out of place when it was delivered as the Terry Lectures at Yale University in 1962, that the scholars who espoused their teachings were accused of something like a cult following, no unlike the followers of Teilhard de Chardin or even Swedenborg. Indeed, one wonders whether the nine years during which Polanyi worked “almost exclusively”¹ on *Personal Knowledge* were not more like the 40 days in the wilderness than they were like the triumphant entry into Jerusalem.

What was it that Polanyi got right? His argument for the crucial contributions of tacit knowledge and for the irreducibility of the category of personhood are so well known to this group that I will not bore you by repeating them. But I do need to describe his less well known position on emergence, since it will set the stage for the arguments concerning consciousness and supervenience to follow.

In his theory of tacit knowing, Polanyi recognized that thought was motivated by the anticipation of discovery: “all the time we are guided by sensing the presence of a hidden reality toward which our clues are pointing.”² Tacit knowing thus presupposes at least two levels of reality: the particulars, and their “comprehensive meaning” (*TD* 34). Polanyi extended this “levels of reality” insight outward to a variety of fields, beginning with his own field, physical chemistry, and then moving on to the biological sciences and to the problem of consciousness.³ In his view even physical randomness was understood as an emergent phenomenon (*PK* 390f.); all living things, or what he called “living mechanisms,” were classed with machines as systems controlled by their functions, which exercise a downward causation on the biological parts (e.g., *KB* 226f.; *PK* 359ff.); and processes such as the composition of a text serve as clear signs that human goals and intentions are downward causal forces that play a central role in explaining the behavior of *homo sapiens*. Polanyi then wraps each of these argumentative steps together into an overarching philosophy of emergence:

The first emergence, by which life comes into existence, is the prototype of all subsequent stages of evolution, by which rising forms of life, with their higher principles, emerge into existence.... The spectacle of rising stages of emergence confirms this generalization by bringing forth at the highest level of evolutionary emergence those mental powers in which we had first recognized our faculty of tacit knowing (*TD* 49).

¹ *Personal Knowledge*, henceforth *PK* (Chicago: University of Chicago Press, 1958), p. xi.

² *The Tacit Dimension*, henceforth *TD* (Garden City, NY: Doubleday Anchor Books, 1967), p. 24.

³ On the latter see esp. *Knowing and Being: Essays by Michael Polanyi*, henceforth *KB*, ed. Marjorie Grene (London: Routledge and Kegan Paul, 1969), esp. Part 4, “Life and Mind.”

Polanyi's Contributions to Contemporary Emergence Theory

But we are impatient people. We are not interested in historical credits for their own sake, but rather wish to know what in this author's work might be of assistance to us today.

Recently there has been a virtual explosion of interest in the notion of emergence. Scientists from physicists to psychologists are employing the notion to make sense of the data in their fields. Harold Morowitz, for example, has identified 28 distinct levels of emergence in his forthcoming book with Oxford University Press.⁴ Other thinkers have used emergence theory to show that awareness in living things involves not just the monitoring of the external environment (a process too easily confused with perception), but also the monitoring of the organism's own internal states and the modification or potential modification of its behaviors as a result. Most recently, Terrence Deacon has linked emergence to the self-reflexivity of the feedback loop in perception, cognition and language use.⁵ In recent years authors are even beginning to write theologies of emergence.⁶ Is there anything new or helpful that Polanyi can add to this blaze of interest in emergence theory?

Yes. Surveying recent emergence theories, I found several insights in Polanyi which, if they are appropriated, will advance the contemporary discussion. Let me mention just three:

(1) *Active and passive boundary conditions*. Polanyi recognized two types of boundaries: natural processes controlled by boundaries; and machines, which function actively to bring about effects. He characterized his distinction in two different ways: as foreground and background interest, and as active and passive constraint. Regarding the former distinction, he argued, a test tube constrains the chemical reaction taking place within it; but when we observe it, "we are studying the reaction, not the test tube" (*KB* 226). In watching a chess game, by contrast, our interest "lies in the boundaries": we are interested in the chess master's strategy, in *why* he makes the moves and what he hopes to achieve by them, rather than in the rule-governed nature of the moves themselves. (I remember how my kids, when they were young, were completely

⁴ See Harold Morowitz, *Emergences: Twenty-Eight Steps from Matter to Spirit*. Oxford and New York: Oxford University Press, forthcoming

⁵ See, among other works, B. Weber and T. Deacon, "Thermodynamic Cycles, Developmental Systems, and Emergence," *Cybernetics and Human Knowing* 7/1 (2000): 21-43.

⁶ See Arthur Peacocke, *Theology for a Scientific Age: Being and Becoming — Natural, Divine, and Human* (Minneapolis: Fortress Press, 1993); John Haught, *God After Darwin: A Theology of Evolution* (Boulder, CO: Westview Press, 2000); David Ray Griffin, *Unsnarling the World-Knot: Consciousness, Freedom, and the Mind-Body Problem* (Berkeley: University of California Press, 1998); Clayton, *God and Contemporary Science* (Grand Rapids: Eerdmans, 1999); Clayton, *The Emergence of Spirit* (forthcoming). Most radically, see Anthony Freeman, *God in Us: A Case for Christian Humanism* (Thorverton, UK: Imprint Academic, 2000).

fascinated just by the regularities of the chess pieces' movements and by the rules that guided them; fortunately, though, I doubt that my six-year-old was a good example of the advanced appreciation of chess!)

More important than the backgrounding and foregrounding of interest, Polanyi recognized that the “causal role” of the test tube is a passive constraint, whereas intentions *actively* shape the outcome in a top-down manner: “when a sculptor shapes a stone or a painter composes a painting, our interest lies in the boundaries imposed on a material and not in the material itself” (*KB* 226). In *Theology for a Scientific Age* and in his various publications in the Vatican/CTNS series, Arthur Peacocke, a seminal thinker in the theology/science debate, has gone back and forth on whether boundary conditions should really be spoken of as causal forces. If he had appealed to Polanyi’s distinction, the ambiguity would have been dissolved. Messages from the central nervous system cause hormone release in a much more active top-down fashion than does the physical structure of microtubules in the brain. Microtubule structure is still a constraining boundary condition, but it is one of a different type, namely a passive one. Much confusion in the literature would be removed by this distinction.

(2) *The “from-at” transition and “focal” attention.* Already in the Terry Lectures, Polanyi noticed that the comprehension of meaning involved a movement from “the proximal” — that is, the particulars that are presented — to the “distal,” which is their comprehensive meaning (*TD* 34). By 1968 he had developed this notion into the notion of “from-at” conceptions. Understanding meaning involves turning our attention from the words to their meaning; “we are looking *from* them *at* their meaning.”⁷ Mind is a “from-to experience”; the bodily mechanisms of neurobiology are merely “the subsidiaries” of this experience (*KB* 238). Or, more forcibly, “mind is the meaning of certain bodily mechanisms; it is lost from view when we look *at* them focally.”⁸ In my recent work defending panentheism I have used a similar distinction between focal intentions and the body’s autonomous functioning in order to defend the possibility of divine actions without divine interventions that would break natural law.⁹

Note, by the way, that there are parallels to Polanyi’s notion of mind as focal intention in the theory of consciousness advanced by the quantum physicist Henry Stapp, especially in his *Mind*,

⁷ *KB* 235f., my emphasis.

⁸ *Ibid.*; cf. 214. Polanyi writes later, “We lose the meaning of the subsidiaries in their role of pointing to the focal” (*KB* 219). For more on Polanyi’s theory of meaning, see Polanyi and Harry Prosch, *Meaning* (Chicago: University of Chicago Press, 1975).

⁹ See *God and Contemporary Science*, esp. chapter 9.

Matter, and Quantum Mechanics.¹⁰ Both thinkers believe that we can comprehend mind as the function of “exercising discrimination” (PK 403n1). If Polanyi and Stapp are right, it represents good news for the downward causation of ideas, since it means that no energy needs to be added to a system by mental activity, thus preserving the law of the conservation of energy which is basic to all physical calculations.

(3) *The theory of structure and information*. Like many emergence theorists, Polanyi recognized that structure is an emergent phenomenon. But he also preserved a place for downward causation in the theory of structure, arguing that “the structure and functioning of an organism is determined, like that of a machine, by constructional and operational principles that control boundary conditions left open by physics and chemistry” (KB 219). Structure is not simply a matter of complexity. The structure of a crystal represents a complex order without great informational content (KB 228); crystals have a maximum of stability that corresponds to a minimum of potential energy. Contrast crystals with DNA. The structure of a DNA molecule represents a high level of chemical improbability, since the nucleotide sequence is not determined by the underlying chemical structure. While the crystal does not function as a code, the DNA molecule can do so because it is very high in informational content relative to the background probabilities of its formation.

Polanyi’s treatment of structure lies very close to contemporary work in information biology.¹¹ Thus Deacon argues that “it is essential to recognize that biology is not merely a physical science, it is a semiotic science; a science where significance and representation are essential elements.... [Evolutionary biology] stands at the border between physical and semiotic science.”¹² Perhaps other elements in Polanyi’s work could contribute to the development of information biology, which is still in the fledgling phases.

Polanyi Criticized

At the same time that emergence theory has profited from Polanyi, it has also moved beyond his

¹⁰ Henry P. Stapp, *Mind, Matter, and Quantum Mechanics* (Berlin and New York: Springer-Verlag, 1993).

¹¹ Hubert Yockey, *Information Theory and Molecular Biology* (Cambridge: Cambridge University Press, 1992); Werner Loewenstein, *The Touchstone of Life: Molecular Information, Cell Communication, and the Foundations of Life* (New York: Oxford University Press, 1999); Mike Holcombe and Ray Paton, eds. *Information Processing in Cells and Tissues* (New York: Plenum Press, 1998); Susan Oyama, *The Ontogeny of Information: Developmental Systems and Evolution*, 2nd ed. (Durham: Duke University Press, 2000); Roland Baddeley, Peter Hancock, Peter Földiák, eds., *Information Theory and the Brain* (Cambridge: Cambridge University Press, 2000).

¹² Terrence Deacon, “Evolution and the Emergence of Spirit,” SSQ workshops, unpublished paper, p. 6.

work in some respects. Let me briefly indicate three such areas, since they will help to propel us, I hope, toward a more adequate theory of mind and personhood.

(1) *Polanyi was wrong on morphogenesis.* He was very attracted by the work of Hans Driesch, which seemed to support the existence of organismic forces and causes (*TD* 42f., *PK* 390, *KB* 232). Following Driesch, Polanyi held that the morphogenetic field pulls the evolving cell or organism toward itself. He was also ready to argue that the coordination of muscles, as well as the recuperation of the central nervous system after injury, was “unformalizable ... in terms of any fixed anatomical machinery” (*PK* 398). While admitting that the science had not yet been established, he hitched his horse to its future success: “once ... emergence was fully established, it would be clear that it represented the achievement of a new way of life, induced in the germ plasm by a field based on the gradient of phylogenetic achievement” (*PK* 402). He even cites an anticipation of the stem cell research that has been receiving so much attention of late: the early work by Paul Weiss, which showed that embryonic cells will grow “when lumped together into a fragment of the organ from which they were isolated” (*KB* 232). But we now know that it is not necessary to postulate that the growth of the embryo “is controlled by the gradient of potential shapes,” and we don’t need to postulate a “field” that guides this development (*ibid.*). Stem cell research shows that the cell nucleus contains the core information necessary for the cell’s development.

(2) *Polanyi bid, I fear, on the wrong philosopher in his philosophy of biology.* Aristotle is famous for the doctrine of *entelechy*, whereby the future state of an organism (say, in the case of an acorn, the full-grown oak) pulls the developing organism toward itself. In a section on the functions of living beings, Polanyi spoke of the causal role of “intimations of the potential coherence of hitherto unrelated things,” arguing that “their solution establishes a new comprehensive entity, be it a new poem, a new kind of machine, or a new knowledge of nature” (*TD* 44). The causal powers of non-existent (or at least not-yet-existent) objects make for suspicious enough philosophy; they make for even worse science. Worse from the standpoint of biology was Polanyi’s advocacy of Bergson’s *élan vital* (*TD* 46), which led him to declare the affinity of his position with that of Teilhard de Chardin.

(3) *This doctrine of vitalism that Polanyi took over from Driesch meant, in fact, a whole-scale break with the neo-Darwinian synthesis,* on which all actual empirical work in biology today is based. Beyond structural features and mechanical forces, Polanyi wanted to add a broader “field of forces” that would be “the gradient of a potentiality: a gradient arising from the proximity of a possible achievement” (*PK* 398). He wanted something analogous to “the agency of a centre seeking satisfaction in the light of its own standards” (*ibid.*). What we do find in biology is the real-world striving that is caused by the appetites and behavioral dispositions of sufficiently complex organisms. The operation of appetites cannot be fully explained by a Dawkinsian reduction to the “selfish gene,” since their development and expression are often the result of finally tuned interactions with the environment. Combinations of genes can code for appetites, and the environment can select for or against them, without however needing to

introduce mysterious forces into biology.

In the end, Polanyi went too far, opting for “finalistic” causes in biology (*PK* 399). It is one thing to say that the evolutionary process “manifested itself in the novel organism,” but quite another to argue that “the maturation of the germ plasm is *guided* by the potentialities that are open to it through its possible germination into new individuals” (*PK* 400). It is one thing to say that the evolutionary process has given rise to individuals who can exercise rational and responsible choices; but it breaks with all empirical biology to argue that “we should take this active component into account likewise down to the lowest levels” (*PK* 402f.). This move would make all of biology a manifestation of an inner vitalistic drive; and that claim is inconsistent with the practice of empirical biology.

Supervenience and the Person

The prophet has played his role. Polanyi formulated the central principles of emergence during the Diaspora of the 1950's, helping the faithful to resist the slings and arrows of outrageous reductionism until biology could move on sufficiently to rediscover the role of emergence in natural history. To name just one example: re-recognizing the role of environmental influences in triggering gene expression has meant an important break from the radical reductionism of Dawkins and his friends. When Steven J. Gould broke from the reductionist model in the late 1970s, it was due in part to his recognition of the role of the environment not only in selecting for or against structures, but also in causing its development. Gould writes, “Minor adjustment within populations may be sequential and adaptive. ... Evolutionary trends may represent a kind of higher-level selection upon essentially static species themselves, not the slow and steady alteration of a single large population through untold ages.”¹³ The building blocks of the cell do not alone account for the cell's development and functioning; environmental factors and chemical changes at the level of the cell as a whole work to promote the expression of genetic potentials in a sort of “top-down” causation. What's right about the modern synthesis — its ability to account for major changes through a sequence of smaller genetic changes — has been retained; yet it has been *supplemented* by top-down theories that help account for gene behavior. Irreducible complexity has brought with it this expanded notion of emergence, which alone can account for broader, more sudden, and nonsequential change.

The question is: Can a version of emergentism be formulated now that avoids the quasi-Aristotelian metaphysic that peaks out from time to time in Polanyi's writing? Or will we discover that any less “finalistic” metaphysic will be ice too thin for theologians to skate on, such that it will crack under the weight and dump theology again into the frigid waters of physicalism, where it will either freeze to death or at best extricate itself only with the greatest of difficulty? Here the recent supervenience debate has moved us a step further along, as I hope

¹³ Stephen J. Gould, *The Panda's Thumb: More Reflections in Natural History* (New York: W. W. Norton, 1980), p. 15.

now to show. Incidentally, I have heard it said that supervenience is too difficult a notion for use by theologians. I beg to differ. Let me merely summarize in three points what I think is the main contribution of supervenience theory.

(1) In the most general terms, supervenience means that one level of phenomena or type of property (in this case, the mental) is dependent upon another level (in this case, the biological or neurophysiological), while at the same time not being reducible to it. In the last I have used the term *weak supervenience*, adapted from Jaegwon Kim, as a way of expressing this minimal position. *Strong supervenience* positions, by contrast, generally argue for a *determination* of higher-order phenomena by the subvenient level (e.g., for the determination of mental phenomena by the neural substrate), such that the subvenient level provides the “real” explanation for the phenomena in question.

(2) If supervenience is understood to be a *token-token* relationship — an individual instance of a mental property directly supervenes on an individual brain state — then, according to most standard presentations of supervenience, there is no real place for mental causation. For in each case the mental event will be dependent upon its corresponding physical event, making unclear why the explanatory story couldn’t be told in terms of physical events alone. One can *say* that a mental input should be added to the chain of brain states causing other brain states. But it is not clear why the mental cause would not be redundant in this case, turning the resulting position into a de facto epiphenomenalism.

(3) Kim has relied on a version of the multiple realizability argument, which I believe strengthens the case for a *type-type* understanding of the relationship between the mental and the physical. Compared to many authors in the philosophy of mind, Kim shows a deeper appreciation for natural history and the evolutionary origins of mind. I would express a variant of his position in this way: in order to allow adequate place for mental causation, we must reject token-token identity theory. Instead, we understand mental properties to be a type of property, which bear a dependence relation on another type of property, the physical (or neuro-physiological) states of the organism.

If this idea is right, the dependence relation of supervenience now has both a synchronic and a diachronic dimension. Mental properties depend upon the entire natural history that caused this complex brain and central nervous system to evolve, as well as depending on the physical state of the organism at a particular time. (To the best of our knowledge, corpses don’t have qualia.) This mutual dependency is neither logical nor metaphysical — two requirements that have become the hobgoblin of the analytic philosophy of mind. Instead, the assertion of both a diachronic and a synchronic dependence of mental properties is our best reconstruction of the highly contingent natural history that led to organisms like us. Of course, this dependency relation *qua* natural history represents a firm break with all dualist theories of mind, thereby distinguishing the emergence approach as a separate ontological option in the debate.

But the contingent type-type relationship between the mental and the physical also allows one to give a more robust account of the nonreducibility of the mental than the competing accounts provide. Wherein, then, does the nonreducibility of the mental lie? How is it to be characterized? Much turns on this question — I would argue, the entire theory of personhood and personal knowledge. The only way to give an adequate answer, I suggest, is to draw on the resources of emergence theory. Although I will not continue to use the term “supervenience” in the final section of this paper, you could understand the position on human personhood being defended as a version of *emergentist supervenience*.

Toward an Emergentist Theory of Mind

Rather than offering you a survey of the most recent articles and books on emergence — which has become a rather impressive body of literature — let me describe the decision points that one faces and then argue for a particular position on the nature of mind or personhood.

To advocate an emergence approach is already to have made certain decisions. It is to reject reductionist physicalism, the belief that all adequate explanations will finally be given in the terms of contemporary physics. On the other side, it is to reject substance dualism, the view that there are two distinct kinds of substances. (In the substance dualism of Descartes, for example, these were *res cogitans* and *res extensa*, thinking and extended substance).

Of course, there are other options, besides emergence theory, available to those who reject both reductionist physicalism and substance dualism. Among other things, one might be a dual-aspect monist, believing that there is just one kind or level of reality, even though it is apprehended by us sometimes in the mode of mind and sometimes in the mode of body; or one might be a panpsychist, believing that “it’s mind all the way down,” that is, that every level of reality has some sort of mental experience. Until relatively recently it was thought that one might also be a non-reductive physicalist, believing that all things are ultimately physical but not that all explanations (and hence not all causes?) can be reduced to the explanations of physics. I follow Jaegwon Kim, formerly its greatest exponent, in holding that non-reductive physicalism is an inherently unstable position rather than a useful halfway point between other options.¹⁴

So you want to be an emergentist; what are your options? The first decision point you face is between epistemic and ontological versions of the theory. According to epistemic versions, emergence has only to do with limitations in *our* knowledge of the physical order and/or with the particularities of how we come to know; ultimately, ontologically, all that exists are the physical systems whose behaviors are expressed by physical laws. Clearly, however, the more robust — and certainly the more ambitious — version of emergence theory includes ontological

¹⁴ See Jaegwon Kim, *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation* (Cambridge, MA: MIT Press, 2000). This work contains references to Kim’s criticisms of non-reductive physicalism going back some eight years.

emergence. For these theories, emergence entails a genuine ontological difference in the world.

Ontological views, in turn, subdivide into those that only accept emergent properties and those that also accept emergent causal powers.¹⁵ The emergent properties view is consistent with a belief that all that actually exists are physical objects controlled by physical laws. It's just that very complicated physical objects like ourselves give rise to some rather unusual properties, such as thinking of world peace, loving eggnog, and being able to play hopscotch. Such properties don't *do* anything; all the "doing" occurs at the level of the physical processes of which we are constituted. But they exist nonetheless. By contrast, emergent powers add a bit more to the ontology: they must exist in a more robust sense if they are going around *doing things* in the world.

Those of us who accept the existence of emergent powers have to choose between stronger and weaker claims on their behalf. Much of the theory of personhood you finally accept will depend on how strong an account you give of emergent causal powers. Theologically minded persons, I have noticed, leap instinctually toward a strong view of emergent causal powers, or what van Gulick calls "radical kind emergence." The weakest form of emergent causality you can defend (Van Gulick's "specific value emergence") is to insist that wholes and parts must have features of the same kind but may have different subtypes or values of that kind. Thus both the car and its parts have weight, but the car has more of it. A stronger version (Van Gulick's "modest kind emergence") would allow wholes to have features that are different in kind from those of its parts. The most ambitious form of downward causality, however, which I have called *strong emergence*, adds that the holistic features of a complex system need not be necessitated by the sum total of facts about the parts. "Accepting radical-kind emergence," Van Gulick writes, "would be conceding that there are real features of the world that exist at the system or composite level that are not determined by the law-like regularities that govern the interactions of the parts of such systems and their features" (p. 18).

Of course, we want human persons to be fundamental in the furniture of the universe. We want our intentions and goals to matter; we want our thoughts and feelings to be causally efficacious; we want things that we find meaningful (or: things that we *want* to be meaningful) really to *be* meaningful. Hence we tend to think that any views that don't give us these things — eliminativism, the identity theory of mind, non-reductive physicalism — must be false. Of course, we realize that old-fashioned Cartesian dualism conflicts with contemporary neuroscience and faces serious inconsistencies on its own. But some robust form of mental causation, one that allows our wishes and dispositions to be causally efficacious, *must*, we feel, play a role in the universe.

Unfortunately, there are serious difficulties facing mental causation, difficulties that I hinted

¹⁵ See Robert Van Gulick, "Reduction, Emergence and Other Recent Options on the Mind-Body Problem: A Philosophic Overview," *Journal of Consciousness Studies* 8 (2001): 1-34.

at in the critique of Polanyi above. Radical kind emergence, the one that I just described as theologically the most interesting, is particularly vulnerable to such attacks. What are its costs, and should we be willing to pay them? One cost involves the danger of negating scientific study and scientific method; another concerns not being able to specify the evolution of neural states; and a third involves not being able to explain where downward causation takes hold and why it does so when it does. Let me summarize some of these difficulties before coming to my own concluding recommendation.

Recall Van Gulick's observation that for radical-kind emergence, "there are real features of the world that exist at the system or composite level that are not determined by the law-like regularities that govern the interactions of the parts of such systems." This claim — which I am inclined to accept — brings a problem, to put it mildly, for the scientific study of persons and their brains. It's much more amenable to modern science to say that the emergence of new macro-properties in a system is ultimately determined by the sum total of relations between the micro-properties of that system. To know the state of all the registers in your computer *just is* to know the state of the system, even though the *content* of the system may be a new emergent property such as a digitized image of the Mona Lisa.

The problem is a serious one. The neurobiologist William Newsome took me to task recently for suggesting in *God and Contemporary Science* that mental events could give rise to new brain events without there being a full causal story in terms of *prior* brain events. If Newsome's requirement is right, it's hard to understand what it means for mental events to initiate new causal sequences in the brain. Where, for example, would be the point of contact: would "mind" affect the outcome of quantum mechanical indeterminacies in the physics of the brain? Would it change the chemical composition at specific synapses? Or would it exercise its causality only at the level of "the brain as a whole," as Roger Sperry believed?

At first glance, quantum indeterminacies seem to offer the ideal opening for mental causation. Unfortunately, contemporary evidence suggests that quantum effects (say, superimposed quantum states prior to decoherence) would probably be eliminated well before one reaches the level of the neurochemical processes that are basic to brain functioning.¹⁶

The major *non*-quantum accounts of how mental causation might occur appeal to non-linear

¹⁶ See e.g. C. Seife, "Cold Numbers Unmake the Quantum Mind," *Science* 287 (February 4, 2000), and Max Tegmark, "The Quantum Brain," *Physical Review E* (2000): 1-14. Against this view, the quantum physicist Henry Stapp has computed that the scale of microtubules is such that quantum effects associated with sodium ions could affect releases and thus the chemical composition of synapses (personal communication).

dynamics, chaos theory, and the field of complexity studies.¹⁷ One problem is that all three of these approaches are still deterministic. Although they have intriguing features such as the role of “strange attractors” and a sensitive dependence on initial conditions, they do not specify where the input of “mind” or “free choice” could be.

Some scholars are seeking to provide proofs that this entire approach to the problem — trying to provide a linear analysis of mental causation in terms of neuron firings — arbitrarily rules out the possibility of genuine mental causal activity from the outset. Thus E. J. Lowe has argued that mental events must have specific goals: one makes the mental decision to buy a car, get out of bed, or sing the *Krönungsmesse*; they are discrete decisions rather than intermingled states. But neural events, by contrast, are “inextricably entangled.”¹⁸ Physical actions are products of an interconnected web of brain events; there are no discrete groupings that represent the neural antecedents of “deciding to open the door” or “deciding to pick up your books.” Hence no physical account can be given that expresses the steps of the decision in neurological terms. Thus “we think of each decision as giving rise to just its ‘own’ movement and without any contributions from decisions to perform other, independent movements; and to abandon this thought is effectively to abandon mental causation as common sense conceives of that phenomenon” (p. 640). Hence, Lowe argues, we must either give up mental causation altogether, or we must understand it to be something more than some specific set of neural events.

The Dynamical Systems Approach

Perhaps you’ve noticed a certain dynamic in this discussion: it’s like riding a see-saw. If you push off too hard from the mental side, you descend into the morasses of neurophysiological detail, and no mental causes are to be found. If however you push off too hard from the physical side, you end up in the world of purely mental terms, and no connection with the brain can be found.

I believe that more recent work on dynamical systems brings us a little bit closer to an adequate perspective on the human person. Such approaches shift from the neuronal level to broad brain systems as the level of analysis for mind. As Hardcastle, for example, writes:

Hormones and neuropeptides impart data through the extracellular fluid more or less continuously in a process known as “volume transmission.” What is important is that these additional ways of communicating among cells in the central nervous system mean that simple (or even complicated) linear or feedforward models are likely to be

¹⁷ See the excellent presentation in Michael Silberstein, “Converging on Emergence: Consciousness, Causation and Explanation,” *Journal of Consciousness Studies* 8 (2001): 61-98.

¹⁸ See E. J. Lowe, “The Causal Autonomy of the Mental,” *Mind* 102, #408 (1993): 629-44.

inaccurate. ... Discovering the importance of global communication in the brain has led some to conclude that it is better to see our brain as a system that works together as a complex interactive whole for which any sort of reduction to lower levels of description means a loss of telling data.¹⁹

This approach gives at least a hint of the direction one might take in addressing the mind-body problem. Global communication in the brain, if it exists, would reflect the kind of holistic effects that we have seen again and again in studying the ladder of emergence. Admittedly, neuroscientists do not yet understand “large-scale, complex electrophysiological or bioelectrical activity patterns involving millions of neurons and billions of synapses.”²⁰ But these are the *sorts* of processes that we can study and could in principle come to understand.

I therefore conclude that, among many other instances of emergence, the world evidences an emergent phenomenon that we call the mental life, which is associated with a significantly complex brain and central nervous system. From the neurological perspective, we come closer to it when we understand the behaviors of massively distributed systems involving multiple regions of the brain (or possibly, as Sperry thought, the brain as a whole). But only from the first-person or phenomenological perspective can we describe the conscious content of these states. Only when we rise to this level do we get to the lived human experience of emotions, volitions, thoughts — in short, that world of meaning and of meaning relationships in which we live and move and have our being.

Summary and Conclusions

We began by looking at the concept of emergence as developed by Michael Polanyi. Although we found some areas where we had to resist his concept, we also found it useful in setting the agenda for an emergence theory of personhood. We then added in the resources of the concept of supervenience, discovering in it a powerful conceptual tool for expressing emergentist intuitions about the nature of mind. In the third section we looked at even more recent work toward an emergentist theory of mind. The category of *radical kind emergence* allowed us to outline, at least, how one might conceive mental activity in terms of downward causation. At the same time, it reveals the difficulties in giving a full neurological account of consciousness. I argued that, despite the difficulties, the costs are still worth paying.

Some assumptions guided the entire treatment, and I should now lay them out on the table before you. I have assumed, on the one hand, that if a given account of mental influence were incompatible with natural science, that would be a telling argument against it. I do think that Aristotle’s doctrine of entelechies, of future (and thus merely potential) patterns pulling natural

¹⁹ See V. Hardcastle, *The Myth of Pain* (Boston: MIT Press, 1999), quoted in Silberstein, p. 82.

²⁰ Silberstein, p. 82.

processes toward themselves, is incompatible in this way. As a result, I have particularly worked to avoid Polanyi's occasional flirtations with this idea. On the other hand, I assumed that we should select between neuroscientific theories, in part, based on how robust an account of mental causation they can deliver — even if it goes beyond the science we currently know.

This set of assumptions still leaves one with a certain amount of room to maneuver. But it can be refined until it represents a more detailed decision procedure. We need only add that the degree of untestability of a theory of mind correlates with the degree to which it becomes an affront to science; and the degree to which it becomes an affront to science represents at least *prima facie* reason to reject it. This decision procedure amounts, I suppose, to a dual wager. It is a wager that the ultimately victorious account will preserve a place for mental causation (and remember that you have to believe in mental causation even when you plan to argue against it!), and a wager that the victorious account will not invalidate the study of the brain or make such study irrelevant. Accepting this dual wager explains why many of us resist Cartesian dualism: since Cartesian mental substance has nothing whatsoever to do with the physical world (it belongs to another world altogether), brain science would never tell us anything about the nature of thought. On the other hand, explanation through the micro-determinism of neural firings can never explain thought because it has not left a place for ideas to have any causal effect on the brain and central nervous system — and thus on one's actions in the world.

I have not been able to list and develop all the dimensions of human personhood suggested by this account, at least not today. But if I am right, no conceptual roadblocks stand any longer in the way of a full theory of human personhood. It can now draw freely not only on the neurosciences and cognitive psychology, but also on the whole range of the social sciences: psychology, sociology, cultural anthropology, and so on. Indeed, given the universality of religious rites, rituals and beliefs across human cultures, a full understanding of the human person will have to incorporate the spiritual dimension and those disciplines that address it as well. But this is an exercise for another talk.

In closing, as one considers the extremes of the mind-body debate during the last century — identity theorists, eliminative materialists, and reductive physicalists on the one side, and substance dualists of many shapes and colors on the other — one is struck by the power of the emergentist research program. Considering its ability to move further and further from anti-scientific postulations, to enter more deeply into genuine dialogue with the neurosciences, yet without losing the “hard core” of emergence, one must surely rank emergence theories as one of the most progressive and promising research programs in this field over the last 50 years. We must say in one sense that, like the British Emergentists of the 1920s, Polanyi has been superseded by the emergence theories of Roger Sperry, John Searle, certain supervenience theorists, and now by the radical-kind emergence theorists of the most recent years. Yet in another sense it is a testimony to Polanyi that he served as a crucial link in this chain, almost singlehandedly linking together the emergence theorists of the early and the late parts of the century. For this crucial function he continues to deserve our gratitude.