Quintessence of Dust: The Science of Matter and the Philosophy of Mind

Value Inquiry Book Series

VOLUME 345

Cognitive Science

Edited by

Francesc Forn i Argimon, *University of Catalonia*

The titles published in this series are listed at *brill.com/vibs* and *brill.com/cosc*

Quintessence of Dust: The Science of Matter and the Philosophy of Mind

By

Harry Redner

BRILL
RODOPI

LEIDEN | BOSTON

Cover illustration: Eugène Delacroix: Hamlet and Horatio in the Graveyard, 1839. Collection Musée du Louvre, RF 1942; Reproduction compiled by the Yorck Project and licensed under the GNU Free Documentation License.

The Library of Congress Cataloging-in-Publication Data is available online at http://catalog.loc.gov

Typeface for the Latin, Greek, and Cyrillic scripts: "Brill". See and download: brill.com/brill-typeface.

ISSN 0929-8436
ISBN 978-90-04-42602-3 (hardback)
ISBN 978-90-04-42686-3 (e-book)
DOI 10.1163/9789004426863

This book is printed on acid-free paper and produced in a sustainable manner.

Contents

Prelude

"There are more things in heaven and earth, dear Horatio, than are dreamt of in your philosophy", Hamlet retorts when Horatio expresses astonishment at the appearance of the ghost of Hamlet the king. We are not told explicitly what is Horatio's philosophy, but from his later remarks we might glean that he is some kind of Stoic. If he were alive today he would be a positivistic Materialist. We know even less about Hamlet's philosophy, or even that he is a philosopher, though as a student at Wittenberg University he would certainly have been taught Aristotelian philosophy. In a recent book on the play, the literary critic Rhodri Lewis devotes a whole chapter to Hamlet as philosopher, even though he is highly critical of him in that capacity.[1] But perhaps there is less justification for that than Lewis supposes.

We know that Hamlet is referring to ghosts or the supernatural element in his philosophy, which Lady Macbeth in the later play calls "metaphysical".[2] In our philosophy both the supernatural and the metaphysical have been banished for what there is in heaven and earth is determined by science. Science forbids us to believe in ghosts, for if we allowed for the ghost in the cellarage we would also be bound to admit the ghost in the machine and that would cause immense difficulties for our view of mind, as well as for our science. In that respect we must side with Horatio's philosophy.

But Hamlet is right in that there is more to our minds than is dreamt of in our philosophy. Hamlet stands for all that which our present philosophy and science neglect and repress, especially in dealing with the mind. He stands for *Homo Ludens*, that side of our mind that concerns art and mimesis, imagination and creation, or, in other words, the whole play aspect of our being. The term is the title of the well-known book by Johan Huizinga and it can stand as the watchword of our approach to the science and philosophy of mind.[3]

According to Huizinga, Homo Ludens together with Homo Faber and Homo Sapiens are the essential characteristics of human mind and culture. As he puts it, "in culture we find play as a given magnitude existing before culture itself existed, accompanying and pervading it from the earliest beginnings right up to the present phase of civilization we are now living in".[4] As we shall see,

1 Rhodri Lewis, *Hamlet and the Vision of Darkness* (Princeton: Princeton University Press, 2017), Chapter 5.
2 *Macbeth,* Act I, scene v.
3 Johan Huizinga, *Homo Ludens: A Study of the Play Element in Culture,* trans. anonymous (London: Routledge and Kegan Paul, 1949).
4 Ibid, 4.

the contemporary palaeoanthropologist Merlin Donald has taken up Huizinga's thesis and developed it in a scientific evolutionary way to account for the origins of the human mind. According to Donald, it is mimetic play that is the starting point of humanity even prior to language. Language alone does not make us human, what is also called for is artistic creativity as evidenced in material symbols and pictures. The age-old Western view of Homo Sapiens as the speaking animal must be complemented with Homo Ludens as the playing and creating animal.

What has this to do with Hamlet? Once more Huizinga provides the answer: "It is more than a rhetorical comparison to view culture *sub specie ludi*. The thought is not at all new. There was a time when it was generally accepted, though in a limited sense quite different from the one intended here: in the 17th century, the age of the world theatre… It was the fashion to liken the world to a stage."[5] Hamlet, the connoisseur of Elizabethan theatre, plays his tragic part on the world stage. He has much to teach us about our own *theatrum mundi* on which we play our parts. Certainly, our world is very different historically from Hamlet's world: we play on a global stage with instruments that command far greater powers than he could have conceived. Our capacity to do ourselves harm is also beyond his imagination. And yet, the human drama has not changed in its basic essentials, and the philosophical issues at play are not all that different despite our boastful claims of progress.

This is the reason that Hamlet figures so markedly in our text and that words drawn from his play are quoted repeatedly throughout; or putting it in modish literary parlance, *Hamlet* acts as the subtext to our main text. It is also meant to act as a reminder that philosophy and science do not just proceed in a world of abstract ideas divorced from realities, but in the real world of pain and suffering and death – the tragedy of human life. Hamlet provides the tragic human touch to our abstruse theorizing, the sense that we live in a historical world where the time is out of joint and that we must try to set it right or suffer tragic consequences such as could not have been even dreamt of in Hamlet's time.

The reasons why our time is out of joint are, of course, very different from those in Hamlet's time. Then it was the surreptitious murder of a king that wrought something rotten in the state of Denmark and brought corruption into the body politic. Now there are different kinds of corruptions in the present state of the world, for we are still traumatized by the mass murders that occurred within the living memory of the oldest, and the youngest live in dread of the even far greater possible catastrophes that hang over our heads like the sword of Damocles.

5 Ibid, 5.

Our time is also out of joint because our eyes are out of focus, for we no longer have a clear vision of ourselves as human beings. The sciences and technologies of our time have distorted our view of ourselves and seem to mandate a conception of our minds as machines. We are the machines from which the ghost has been exorcized through the ministrations of our philosophies and sciences. The instrument these have employed to carry out this exorcism is the computer, an invention that arose during the dark days of the last Great War when so many of our current technologies were first developed. These technologies now enmesh our lives, none more so that computer networks. So that even if we have still not fully accepted the idea of being machines, we are in thrall to the machines that control what we do and determine how we live.

But let us leave the tragic side of things and turn more to the comic which is also there in Hamlet. Hamlet reminds us that "the play is the thing to catch the conscience of the king". Play, as we shall show, catches more than just conscience, it grasps hold of consciousness itself. It does so more decisively even than thought. According to Descartes, who came shortly on the heels of Hamlet, it is thought that defines us primarily as conscious beings. Hence he pronounced his famous dictum "Cogito, ergo sum". But perhaps it is even more appropriate to declare "Ludo, ego sum", for playing in the general sense of miming, imitating and representing is more characteristic of being human than is thinking. "I play, therefore I am" might be the more profound truth since, like Hamlet, we are all players before we are thinkers; and too much thinking can make cowards of us all, as Hamlet rues. However, "nothing is either good or bad but thinking makes it so", that is, without thought there could be no judgement.

After Hamlet, Descartes is the next most crucial figure in our story. It is to Descartes that we owe the initial exposition of the scientific method of mathematical quantification and of experimentation that is at the heart of our science and philosophy. Descartes together with Galileo and Kepler initiated the Scientific Revolution out of which our world of science and technology developed. Consequently, Descartes is the hero of the standard histories of the modern world; but in our account he serves more as the anti-hero, for out of his philosophy have sprung the contemporary forms of Cartesianism with which we now have to battle and attempt to overcome. In declaring the body to be a machine, Descartes prepared the ground for the later progression to the view that the mind, too, is a machine.

Hamlet had as yet no inkling of any such Cartesian mechanistic philosophies; even though he did sign his letter to Ophelia with the words "thine, evermore, dear lady whilst this machine is to him, Hamlet." But this is obviously only a fortuitous reference to the machine, rather than an anticipation of the much later *homme machine* doctrine. And the same is true for his apparent premonition of Cartesian doubt. In the same letter he writes “doubt that the stars are fire, doubt that the sun doth move, doubt truth to be a liar”, which is obviously mere poetic hyperbole and not real scepticism. Lewis is most probably right in arguing that Hamlet could not have even entertained the Copernican hypothesis. And yet, the neuroscientist Antonio Damasio contends that “Hamlet is aware of the science of the day – physics and biology, such as they are; he goes to the university of Wittenberg after all – and he knows about the intellectual dislocation brought about by Martin Luther and Jean Calvin.”[6] Damasio is certainly right about Luther, for not only was Wittenberg also Luther’s university, but Hamlet points directly at Luther when he puns on “the convocation of worms”.

The real hero in Damasio’s account is neither Hamlet nor Descartes, but Spinoza. He notes, however, that all three are linked when he remarks that “Spinoza was born into an age of questioning, an era that might well be known as Hamlet’s age”. But there is a difference in their questioning. For Hamlet “to be or not to be, that is the question”; for Spinoza being and non-being are no longer the main question, rather it is that of the being with contradictory attributes, human being. On this key issue of reconciling thought and extension or mind and body Damasio sides with Spinoza as against Descartes. This also means that he is against the contemporary forms of mechanistic Cartesianism based on the computer model of mind. Since what he has to say about that fits so well the overall thesis of this work, we will allow ourselves to quote him at some length right at the very start:

> …we can be certain Spinoza was changing the perspective he inherited from Descartes when he said, in *The Ethics*, Part 1, that thought and extension, while distinguishable, are nonetheless attributes of the same substance, God or Nature. The reference to a single substance serves the purpose of claiming mind as inseparable from body, both created, somehow, from the same cloth. The reference to the two attributes, mind and body, acknowledged the distinction of two kinds of phenomena, a formulation that preserved an entirely sensible “aspect” dualism, but rejected substance dualism. By placing thought and extension on equal footing, and by tying both to a single substance, Spinoza wished to overcome a problem that Descartes faced and failed to solve: the presence of two substances and the need to integrate them. On the face of it, Spinoza’s solution no

6 Antonio Damasio, *Looking for Spinoza: Joy, Sorrow and the Feeling Brain* (Orlando/Florida: Harcourt, 2003), 225.

> longer required mind and body to integrate or interact; mind and body would spring in parallel from the same substance, fully and mutually mimicking each other in their different manifestations. In a strict sense, the mind did not cause the body and body did not cause the mind.[7]

The last sentence, as we shall show, is particularly noteworthy.

Spinoza – as seen through the eyes of Damasio and interpreted with Damasio's knowledge of science – is as good an introduction to the mind-body problem as that to be found in any philosopher. But it is only an introduction and much needs to be added to it and amended in it before it will begin to provide answers to our main questions. Damasio's own sciences, neurology and psychology, need to be consulted and the contribution that both of them have to make to the issue spelled out. How they relate to each other forms the nub of our account of mind. The difficulty is one of determining how these two kinds of science come together. Neurology is at the "hard" sciences end of the interface where mind meets body; psychology is at the "soft" sciences end – it is where mind reaches out outside its own body to the external social environment of language and culture. Tying up these loose ends is the knot intrinsicate of our problem.

This is not a problem with one solution that might be formulated once and for all as the true answer. It is much more like a historical work in progress, an ongoing scientific and philosophical exploration that humanity engages in age by age. But there is real progress in this work for ever newer discoveries are being made and better theories are advanced; and we can judge this by how far we have come since Hamlet. The next few centuries will undoubtedly bring as much if not more progress again. But that does not mean it will bring us closer to an ultimate truth or final solution. It could turn out to be an endless endeavour, humanity's ceaseless preoccupation with understanding itself. Or it might turn out that there will be some way of bringing this quest to an end or ending it due to the sheer impossibility of going on any further.

All such speculations about what the future might or might not bring are nothing more than that – speculations. They are unsubstantiated and unverifiable since there is no way of peering over the horizon of historical time to see what it will reveal on the other side, as it were. However, that does not prevent us surveying all that which is within range of our temporal vision at present and even a little beyond into the future. We can study the sciences as they now exist and as they might develop in the near future, and so ascertain what they might disclose about the mind. We can interrogate them philosophically and

7 Ibid, 209.

make them yield answers to the questions we can pose to them. The following are some of the things we might ask.

What do the "hard" sciences and the soft "sciences" have to contribute to our knowledge of ourselves as minds and bodies? Why are their contributions so very different? What is it that makes neurology a "hard" science and psychology a "soft" science, what distinguishes them in their methods and procedures? Why do we even need two such different types of sciences? How do these two sciences fit in with all the others: how does neurology relate to biology and the sciences of evolution and how does psychology relate to anthropology and sociology? How can the humanities, such as literary criticism of plays like *Hamlet*, contribute or have any bearing on the sciences at all? At present the vast majority of scientists and philosophers would hold that the humanities have no relevance to any such scientific issues. Most would argue that the "soft" sciences are not all that relevant to the mind-body problem either, and would look solely to the "hard" sciences for answers. This is one of the reasons why the humanities and "soft" sciences have fallen on such hard times at present, so much so that many scientists and philosophers believe that they do not really count at all. Here again, our reference to *Hamlet* serves a purpose in reminding us that literature does matter and that it cannot be simply dispensed with for it has so much to teach us about ourselves.

The trend to relying on the "hard" sciences alone in unlocking the mystery of the mind has become ever more pronounced since the end of the Second World War. The spectacular advances in technology that then began and have continued ever since have given rise to developments on two broad fronts: on the one hand, we now have instruments and recording devices that enable us to peer into and see what is going on in the brain in ever greater detail; on the other hand, we can build machines that seem to have some of the same capacities as the brain, the computer being the foremost among these. The coming together of these two developments has led to a great variety of new scientific approaches and whole new sciences. Neuroscience, cybernetics, cognitive science, artificial intelligence, connectionism, robotics and many more have been new departures in the "hard" sciences. What has ensued has been a veritable mechanization of the mind. The history of this whole trend in the sciences of mind has been masterfully summarized in a two volume book by Margaret Boden.[8] This work represents both a triumphalist vindication and validation of the evident success of this movement; but also a testament to its ultimate

8 Margaret Boden, *Mind as Machine: A History of Cognitive Science* (Oxford: Oxford University Press, 2006).

failure as its subtitle, *A History of Cognitive Science*, implies, for when the history of a science is retrospectively being written this usually indicates that it is past and its time is up.

It is this whole post-war technologically conditioned trend that we are setting ourselves against. But instead of tackling it head on right from the very start, we shall adopt the indirect approach and by indirection find direction out. Only at the very end will it be possible to openly confront the "mind as machine" advocates and other "hard" sciences exponents of mind as brain. For prior to any such confrontation we must first arrive at an alternative view of the nature and origin of the mind. This acts as the positive foil of the negative critique; the one cannot exist without the other, like the two sides of a coin; which comes first is a matter of tactics, and the presentation we have adopted could just as easily have been reversed. But to have done so and begun with a direct frontal attack would have forfeited the advantages of the indirect approach.

Right at the very start we begin our indirections by asking a few childishly simple questions: what are the origins of matter, life and mind, the three most fundamental modes of being? But we soon discover that as with all such naïve questions about origins they are all but unanswerable, not because we have no answers, rather because we have too many, a veritable *embarrass de richesse*. Unfortunately, we have no way of decisively choosing between them. Determining how matter, life and mind arise is all but undecidable since we are unable to prove or disprove empirically the various explanations on offer. All the answers are plausible and there are no experimental tests or other decision procedures for falsifying or ruling out any of them. This means that we have no commonly accepted theoretical explanations of the fundamental issues in physics, biology or psychology. Our initial task will be to explain how such an anomalous situation has arisen in the sciences and what it tells us about the course of science as a whole. Does it mean, as some claim, that science is coming to an end?

Our preoccupation in this work is, of course, with the third of these issues, that of the origin or emergence of mind. The other two, the origins of matter and of life, are only there as prolegomena to the main text. Matter – which concerns physical reality – is a highly technical issue that spans the whole gamut from elementary particle physics to solid state physics, including chemistry. Life, which concerns biological reality, is also the province of numerous sciences from biochemistry and molecular genetics through to animal ethology. Mind, which concerns animal and human cognitive capacities, has even more scientific specialities both in what we have called the "hard" and the "soft" sciences. And besides, there are numerous philosophies of mind, some of which we will deal with but also many which we cannot even consider purely for lack of space, if for no better reasons.

If understanding the mind means understanding ourselves as human beings, then it seems as if we have both more knowledge of ourselves than of anything else, and yet we are further from knowing ourselves than of knowing anything else, short of the universe as a whole. We know that our brain is the most complex single object in the universe; but we do not know whether we will ever be able to fully plumb its workings. There are, indeed, philosophers who argue that the human brain does not have the capacity to know itself, that only a superior brain can do that; for as we can understand animals, so only superior beings are able to understand us. On the face of it, such an argument seems specious for the problem is not that of an individual brain knowing itself and so being defeated by its own self-limitation, but that of the joint activities of a multitude of brains engaged in a collective undertaking to analyse and decipher piece by piece the operations of this single system. There is nothing to prohibit that in principle, though it might take a long time in practice; how long we do not know.

However, in practice there might be limitations in how far we might go in unravelling the inner tangles of our brains if, for example, its workings prove to be too complex or chaotic. So it is possible that at some point in the distant future there will be diminishing returns on the scientific effort expended on understanding the brain. It is possible that in respect of mind as well, truly novel and significant knowledge will eventually peter out. These are conceivable possibilities, but we have no reason at present to believe that we have as yet or will soon encounter any such barriers. Productive returns are still being achieved in both the "hard" and "soft" sciences of brain and mind, and this might continue for some time still.

However, there are thinkers who disagree and argue that not only the brain and mind sciences but the sciences in general have now reached the stage of ever diminishing returns. This argument was first launched by the molecular geneticist and neuroscientists Gunther Stent and was later restated by the science journalist John Horgan. Stent tries to show that all the crucial fundamental discoveries in all sciences have already been made and that there is nothing left to do except fill in the details to the last decimal point. He relies on Thomas Kuhn's theory of scientific revolutions according to which the crucial discovery is the initial paradigm and all the rest is merely working out and branching out or what Kuhn calls normal science. It seems, according to Stent and Horgan, that all the paradigms have already been discovered, only their application to the infinite variety of concrete particulars remains to be laboriously elaborated. Stent takes geography as an analogy for all other sciences:

> Geography, for instance, is bounded because its goal of describing the features of the Earth is clearly limited. Even if the totality of the vast number of extant topographic and demographic details can *never* be described, it seems evident nevertheless that only a limited number of significant relations can ultimately be abstracted from these details. And, as I tried to show in *The Coming of the Golden Age*, genetics is not only bounded, but its goal of understanding the mechanism of transmission of hereditary information *has*, in fact been all but reached. Indeed, and here I will probably part company with some who might have granted me the preceding example, even such more broadly concerned scientific taxa as chemistry and biology are also bounded.[9]

Stent believes that all the major problems in biology have fundamentally been solved or about to be so; and he explicitly refers to three key issues not solved which he confidently believes will be soon overcome: the origin of life, the mechanism of cellular differentiation and the functional basis of the higher nervous system. He places his confidence in "the central dogma of molecular genetics" as elaborated by Crick and Watson, and on "the host of biologists now standing ready to do battle and vast armoury of experimental hardware at their disposal". Hence he expects that "origin of life, differentiation, and the nervous system cannot help but soon suffer the fate that was accorded to heredity in these last twenty years."[10] It is now over forty years since that sanguine prediction was made and though undeniably there have been advances, these problems are still very far from solved.

Stent's thesis has been taken up by Horgan, first in a book, *The End of Science*, and subsequently in other publications.[11] Horgan reviews the state of the sciences in a much more sceptical spirit even that Stent. He holds that in many fields science has reached a dead end where no further progress can be expected. This he holds to be the case in relation to the three fundamental issues we previously raised, the origins of matter, life and mind. He sees many of the theories on offer in these fields to be as much philosophical as scientific, consequently not subject to empirical testing, and therefore outside the ambit of science as he understands it.

Both Horgan and Stent before him make many insightful comments on where the sciences stand at present, but the conclusions they draw from them must be rejected. Science is not coming to an end; this only appears to be so

9 Gunter, S. Stent, *The Paradoxes of Progress* (San Francisco: Freeman, 1978) 49.

10 Ibid, 49.

11 John Horgan, *The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age* (New York: Basic Books, 1996).

because a certain kind of reductive science has now mainly accomplished what it set out to do. The work of the kind of science that in an earlier book, *The Ends of Science,* we called Classical science is now nearly at an end because the reductive task it set itself is nearly over.[12] Once we have attained the fabled *Theory of Everything* then it will be over; but nobody can now tell when, if at all, this will happen.

This is the reductive quest in science that began with Descartes and was first articulated in his *Discourse on Method.* Subsequently it was carried much further by Newton who declared "hypothesis non fingo", namely, he was not concerned with metaphysical issues of the fundamental nature and ground of things, merely to explain them empirically by reference to the laws governing them. All other subsequent scientists continued this quest to reductively explain the whole of nature. Physics came close to completion with Relativity Theory and Quantum Theory and subsequently with the Standard Model of elementary particle physics; and the other physical sciences followed suit. Chemistry was successfully reduced to physics starting with Bohr's work. Biology proved a harder nut to crack but a good start was made in genetics by Crick and Watson's reduction of the principles of Mendelian heredity to the atomic structure of DNA, and it was believed, as Stent said, that the rest would follow. Even psychology seemed finally to be susceptible to reductive treatment through the Cognitivist Revolution based on information processing.

Reductive science aims to explain entities on a higher and more complex level by reference to theories couched in terms of entities on a lower and more basic level. It is pre-eminently an analytic and atomistic mode of explanation, breaking things down to their lowest constituents and then reconstituting them again. The first analytic move is relatively easy, but the second synthetic one rarely succeeds. It is not possible to derive chemistry from physics or biology from chemistry. The successful reduction of biology leaves unanswered most key questions, such as the ones Stent confidently expected to be soon settled. The reason for this is that reduction in principle tells us nothing about what is actually possible in practice. When Dirac discovered the positron, he believed he had solved the fundamental problems of physics, and he famously declared "all the rest is chemistry", namely, a matter of laboriously solving the equations. But the equations could not be solved.

Einstein expressed himself in a similar vein in a statement which we will quote in full because it is so indicative of how a great reductivist master thinks about science:

---

12 Harry Redner, *The Ends of Science: An Essay in Scientific Authority*, (Boulder, Co.: Westview Press, 1987).

> In regard to his subject matter … the physicist has to limit himself very severely: he must content himself with describing the most simple events which can be brought within the domain of our experience; all events of a more complex order are beyond the power of the human intellect to reconstruct with the subtle accuracy and the logical perfection which the theoretical physicist demands. Supreme purity, clarity, and certainty at the cost of completeness. But what can be the attraction of getting to know such a tiny section of nature thoroughly, while one leaves everything subtler and more complex shyly and timidly alone? Does the product of such a modest effort deserve to be called by the proud name of a theory of the universe? In my belief the name is justified; for the general laws on which the structure of theoretical physics is based claim to be valid for any natural phenomenon whatsoever. With them, it ought to be possible to arrive at the description, that is to say, the theory, of every natural process, including life, by means of pure deduction, if that process of deduction were not far beyond the capacity of the human intellect. The physicist's renunciation of completeness for his cosmos is therefore not a matter of fundamental principle.[13]

The telling phase in that declaration is "if that process of deduction were not beyond the capacity of the human intellect", for almost always deducing a natural phenomenon from already known laws and initial conditions is beyond the scope of the human intellect. Even if we have the relevant equations at hand, they are usually too difficult to solve; in such cases they stare back at us mutely and uncomprehendingly. Hence, being able to reductively explain something in principle tells us almost nothing about what is actually the case in practice. Einstein was certainly wrong to hold that life can be arrived at by "pure deduction".

We can reductively explain life in terms of chemistry and molecular genetics; Crick and Watson demonstrated that conclusively. But we cannot start from the laws of chemistry and molecular structure and deduce from these the origin of life. Despite the fact that we know all about the molecular and chemical constitution of the living organism, we still do not know how life originated though there are many plausible, unfortunately mutually contradictory, accounts about it. In brief, as we showed in *The Ends of Science*, reduction cannot be reversed and transformed into deduction; from the fact that X can be reduced to Y, it rarely follows that X can be deduced from Y.

---

13 Albert Einstein, *Ideas and Opinions* (New York: Crown, 1954), 225.

In seeking to reverse reduction, as it were going backwards from Y to X, we come upon a concept that is playing an increasingly more important role in contemporary science, that of emergence. We can say that X emerges from Y, and we can also proceed to explain how and why this takes place. To scientists brought up in the reductivist philosophy of Classical science this seems to be either an illegitimate move, an illusionist sleight-of-hand trick of getting something out of nothing, or something utterly mysterious and incomprehensible. The re-emergence of emergence in science and philosophy has consequently been an uphill struggle and bitter battles have ensued in many fields. But there is no way of dealing with the issues of origin, above all that of the mind, which we previously raised except through invoking the notion of emergence.

Emergence is not a new idea. It had first arisen in the work of J.S. Mill in the mid-nineteenth century and a little later it flourished in connection with Darwin's theory of evolution. It was particularly prominent in British philosophy and science till just before the onset of the Second World War. But after the war it lapsed due to the reductivist triumphs that then ensued. So much so that it was made unmentionable for a long time. It went into a kind of conceptual hibernation which lasted till almost the end of the twentieth century. Only a few daring thinkers had raised it during this period, such as the neurologist Roger Sperry and the philosopher Donald Campbell in the 1970s. But around the turn of the millennium it aroused considerable attention from many scientists and philosophers. By now it is back with a vengeance as if to make up for lost time.

We shall attempt to make sense of emergence and show that it is neither a mystery nor a trick but the logical consequence of reversing reduction. It has a clear logic of its own which is that of indissoluble non-identity, such that to assert that X emerges from Y entails both that X cannot be detached from Y and that X cannot be identified with Y. We shall also show that it is in some broad sense an empirical matter whether or not emergence does or does not apply in a given case, namely, whether the concept of emergence can be invoked as part of an explanation in a given scientific context. It is undoubtedly the case that there are very many phenomena and systems where the concept of emergence has been invoked with considerable justification and explanatory plausibility. This has occurred in all scientific fields, which raises the further question of whether an overarching theory of emergence can be formulated to cover all of them. There have been many attempts to do so, frequently presenting themselves as unified theories of systems or complexity or adaptive matter and many others. We shall study a number of these and seek to ascertain whether they have succeeded in presenting an all-embracing theory of the many diverse phenomena of emergence. To forestall any doubts about this matter and anticipate our eventual conclusion, we might reveal even now that generally such attempts have failed. Whether or not they will succeed in the future is another matter.

What emergence represents for the history of science depends on how all these issues will work themselves out in the future. At present we cannot tell whether it stands for another kind of science, perhaps even another era in science after the Classical age; or, on the contrary, whether it is a trivial pastime of speculative minded scientists and philosophers at the end of science, as Horgan maintains. Horgan is particularly critical of theories of complexity and chaos and other such modish fashions in science which promised a great deal but delivered very little. But one must not take such disappointments – frequently the result of excessive expectations raised by journalistic hype – to prejudge the issue of where science stands at present. It is most likely that we are in an in-between time waiting for something decisive to happen; like Hamlet's ghost, we are wandering between two worlds, one dead, the other powerless to be born.

In a historical period when the time is out of joint, there are even more serious problems besetting philosophy. The end of philosophy is also a frequently mooted topos which we tried to set right in a book that complements the one on science, entitled *The Ends of Philosophy*.[14] Once again, the point of that title is to stress *ends* and not *the end*, *telos* and not *finis*. The book sought to propose other ends for philosophy than those that had been dominant during the long era of Positivism.

During the period of Positivism two variants of this extensive movement became prominent: Logical Positivism and Linguistic Positivism. Wittgenstein played a central role in both. Logical Positivism, a name which it acquired much later, began prior to Wittgenstein with the technical triumphs in formalizing logic and mathematics first initiated by Gottlob Frege and subsequently completed by Bertrand Russell and A.N. Whitehead in their masterwork *Principia Mathematica*. This formal "language" became the basis for Wittgenstein's initial quasi-metaphysical sketch the *Tractatus Logico-Philosophicus* that combined Frege with Schopenhauer, the two opposed poles, conceptual Realist and Idealist, of German philosophy.[15]

However, as is well known, Wittgenstein changed his mind later in his career when he turned away from formalized "language" to common language and

14 Harry Redner, *The Ends of Philosophy: an Essay in the Sociology of Philosophy and Rationality* (London: Croom-Helm, 1986).

15 See Harry Redner, *Malign Masters: Philosophy and Politics in the Twentieth Century*, (Basingstoke: Macmillan, 1997).

from logical critique to a kind of therapeutic linguistic analysis modelled on psychoanalysis.[16] He was undoubtedly also much influenced by the native British philosophy of common sense as expounded at the time by his friend G.E. Moore. The closeness of Wittgenstein's linguistic analysis to their own traditions made it very appealing to British philosophers. Thus under this dual influence a school of Linguistic philosophy arose at Oxford headed by Gilbert Ryle.

Both these types of Positivism were brought over to America where they were amalgamated as two sides of Analytic philosophy which has dominated philosophy departments since the Second World War. Logical Positivism arrived in America through the agency of leading exponents of the Vienna Circle, Rudolf Carnap and Herbert Feigl, who came as refugees. Linguistic philosophy was introduced by Wittgenstein's pupils, such as Norman Malcolm, and through lecture courses given by Oxford philosophers, such as J.L. Austin. Both gained many distinguished American adherents, such as Willard Quine on the one side and Richard Rorty and Stanley Cavell on the other.

It is our contention that the ends of philosophy practiced and preached by both the "turning to logic" and the "turning to language" movements, though perhaps salutary in their time, have by now clearly turned into dead ends. Analytic philosophy in general has more than outlived its time and is now moribund if not yet altogether dead. The time has surely come to seek for new ends of philosophy, new ways in which philosophizing can still be carried on, for otherwise it will indeed come to an end. This is by no means an idiosyncratic opinion but one shared by a number of outstanding philosophers in America whose views will be discussed in this book. The two who will be extensively dealt with are Evan Thompson and Mark Johnson and there are others who might be considered as well.

A new start in philosophy must be made because not only has it sunk into a sterile Positivism of the logical or linguistic varieties, but also because it made common cause and modelled itself on the misguided scientific-technological project to mechanize the mind. This is the "mind as machine" movement which flourished in American science in the period after the war, first as Cybernetics and subsequently as the Cognitivist Revolution to which we previously adverted. American philosophers almost *en masse* fell into step with this scientistic development and propounded theories of mind accordingly. None was more receptive in this respect that Daniel Dennett who started his endeavour at

16 See Harry Redner, *The Tragedy of European Civilization: Towards an Intellectual History of the Twentieth Century*, (New Brunswick, NJ: Transaction Publishers, 2015).

Oxford under Ryle. This demonstrates how easily Analytic philosophy could accommodate itself to scientistic Cognitivism.

Due to war-time exigencies which we shall later spell out, Cognitivism began at the Massachusetts's Institute of Technology (MIT) at the hands of its great luminaries, Marvin Minsky and Noam Chomsky. Both exercised a profound influence on philosophers all over America and eventually in Britain and its academic spheres, as exemplified pre-eminently by Margaret Boden. A precursor to Boden's book on the history of the Cognitive Revolution, Howard Gardner's *The Mind's New Science*, points to the attraction that Cognitivism held for philosophers and explains why they were so taken with it.[17] The mind's new science quickly spawned the mind's new philosophy in a number of variants of this approach. One very influential version was produced by a pupil of Chomsky, Jerry Fodor, and went under the designation of Functionalism. Another was Dennett's philosophy of mind based on Minsky's society of mind theory. And there were many other such developments. Philosophy once more returned to the subservient position to which Locke had once sought to consign it, that of the underlabourer to science.

Philosophy as the handmaiden of science linked its fortunes to the science which it served. It happened to be Cognitivist science which was in the ascendency. Since then the Cognitivist Revolution has shown itself to be far less revolutionary that it was originally thought to be and, like so many others, it has ended up as a failed revolution. The philosophies of mind that based themselves on it are also registering the same symptoms of failure. The whole "mind as machine" movement both in science and philosophy is by now showing all the signs of an exhausted undertaking. It was a mistaken enterprise from the start and the errors inherent in it are by now glaringly obvious, as we shall show in the very last chapter of this book, when finally our indirections find direction out.

A new start must be made in philosophy, one that will once more align it with science but in a different way to that which prevailed during the Positivist era with its reductivist preoccupations. Philosophy must reengage with those scientific issues where there is a need for philosophical mediation and where science alone cannot settle matters. These are the kinds of issues of emergence that we previously touched on and will go on to consider in greater depth, such as the questions regarding the origins of matter, life and mind. It is the last of these, the emergence of mind, which will preoccupy us throughout most of this book. We will address this problem at once on the scientific and philosophical

---

17 Howard Gardner, *The Mind's New Science: A History of the Cognitive Revolution,* (New York: Harper Collins, 1985).

levels and show how the two can be integrated to provide answers to the common problems that arise in both domains. However, for heuristic reasons of clarity of exposition we shall concentrate in Part 2 mostly on the science of mind and in Part 3 predominantly on the philosophy of mind. But the two cannot really be separated for at every point science and philosophy come together on the very same problems and solutions.

Thus the book might serve as a demonstration of a new way of bringing science and philosophy together in dealing with fundamental issues that concern both. This kind of partnership might be entered into on many other major topics apart from mind. Part 1 of the book that deals with the science of matter offers some suggestions as to how philosophy might re-enter the discussion of key problems in science from which philosophers have been largely excluded or, better put, they have excluded themselves by adopting the previously outlined blinkered Positivistic approaches. We leave it to others better qualified in the relevant scientific fields to once more take up the interpretative tasks that call for and invite philosophical participation.

On the matter of mind, which Hamlet calls the quintessence of dust, we might once more begin by turning to his play. For it is Hamlet who coins this expression at the conclusion of his peroration on the nature of Man. So we turn to that first, and let Hamlet speak to us in his own words.

PART 1

The Science of Matter

CHAPTER 1

The Origin and Nature of Things

1 Matter, Life and Mind

> What a piece of work is a man, how noble in reason, how infinite in faculties, in form and moving, how express and admirable in action, how like an angel in apprehension, how like a god! The beauty of the world; the paragon of animals; yet to me what is this quintessence of dust?[1]

The whole point of this book is to demonstrate the truth of Hamlet's assertion: Man is the quintessence of dust, but to do so with somewhat more up-to-date scientific and philosophical means. By "dust" we understand matter in general. So putting it in modern terms, we can say that Man or the human mind emerges from matter; which today is relatively uncontroversial and might even have been said in Hamlet's time, though in vastly different ways in either Aristotelian or Lucretian terms. But to speak of the "quintessence of dust", which in Hamlet's time had a simple philosophical and alchemical meaning, requires today a greater degree of interpretative license. We shall take it that it can be translated as referring to a fifth level of emergence from an original state of matter. This version can be given a sound scientific sense and we shall proceed to do so.

Emergence is a concept that has come to play a large role over the last thirty years or so both in science and philosophy, so that the idea of the emergence of Man from dust, as it would traditionally have been known, or in modern terms as the emergence of mind from basic matter, is not a strange one to us now. Indeed, scientists and philosophers see the whole of the universe and everything solid that it contains, including this too solid flesh, as emerging from intergalactic dust, literally stardust. Or as Hamlet puts it, "this most excellent canopy, the air, this majestical roof fretted with golden fire, why, it appeareth nothing but a foul and pestilent congregation of vapours". This can be also interpreted in modern terms as the universe of stars emerging from an original golden fire called the Big Bang. This is how Harold Morowitz puts it in his comprehensive book about emergence entitled *The Emergence of Everything*, and many other scientists have written similar accounts for popular consumption.[2]

1 *Hamlet*, Act II, scene ii.
2 Harold Morowitz, *The Emergence of Everything* (New York: Oxford University Press, 2004).

So the question inevitably arises of how many essential stages of emergence ensued between the Big Bang and the human mind? We shall attempt to show that there were at least five such stages and that, therefore, Hamlet got it exactly right - Man *is* the quintessence of dust. To do so will require a very brief detour through these essential stages each of which raises fundamental questions and problems that science is at present unable to solve, that is, to answer unambiguously with commonly agreed on theories and definitions, and establish with empirical certainty. In other words, science is unable to deal with the difficulties raised by the essential stages of emergence relying on purely scientific methods. Inevitably, science has to bring in the aid of philosophy to the discussion. And, as is well known, philosophers always differ in their answers to a difficult question. Hence, as we shall see, there are always different answers to the major questions posed by the sciences of emergence. This is, above all, true of the emergence of mind from matter, which is the major, almost the sole, problem dealt with in this work. The other issues of emergence raised in this introductory chapter are only supplementary to the main topic and are not intended to figure in their own right, though each of them raises interesting and difficult scientific and philosophical conundrums of its own.

Authors such as Morowitz and, indeed, the great majority of books on science make it seem as if science has proven answers to just about all problems raised in science. Thus in his book Morowitz makes it seem as if science can equally well explain the whole scope of creation from alpha to omega, from the initial Big Bang at the origin of the universe till human mind, society and culture at the end of terrestrial evolution. There seems to be an established and confirmed science at every point in this vast panorama of emergence; we seem to know everything we need to know about how anything arises from its preceding conditions. However, a closer look at Morowitz's book or that of many others quickly dispels this illusion.

All one need do is ask one simple question: how did matter, life and mind originate? Indeed, Morowitz himself prompts this very question when he states that "the three principal beginnings are origin of the universe, origin of life, and origin of mind."[3] As we shall soon see and as Morowitz notes, to ask for the origin of the universe and to ask for the origin of matter is in fact the very same question, for at this level the almost infinitely large and the infinitesimally small come together in the one scientific quest as cosmology and elementary particle physics merge. Hence, we might rephrase our initial question thus: does science have indisputable or, at least, empirically testable theories of the origin of matter, life and mind?

3 Ibid, 26.

But one can soon surmise, implicitly from Morowitz's own account of the emergence of everything, that science at present has no such theories. When it comes to the emergence of matter, life and mind there are numerous theories available which might or might not even agree on fundamentals. Furthermore, and even worse, no experimental tests can be devised at present or in some cases can even be thought of as to how such differences might be resolved and how such theories might even in principle be falsified. The reasons for this anomalous situation in science are very different in the three cases, as we shall show. But whatever the reasons, they all share this unprecedented condition in science of not being subject to experimental proof or disproof; each can only be argued for or against on general grounds in which experimental findings or mathematical and computer calculations only provide supportive evidence. Available theories concerning these primordial origins do not behave like theories in all other areas of science. Elsewhere proposed theories are either upheld or rejected, but not so where the emergence of these fundamental entities is concerned. Why is this so?

The problem of the origin of matter, life and mind is completely bound up with that of the nature of matter, life and mind. For if we do not know what something is, that is, how it is to be defined, we cannot know what it is, or look for its origins since we do not know what it is we are looking for. But also inversely, if we do not know the origins of something, that is, how it is constituted, we cannot know what it is or what its nature consists of. Thus posing the problem of the origin of something as basic as matter, life and mind is also *ipso facto* posing that of the nature of its being. And answering the one question is simultaneously also answering the other for the two are bound up with each other. Hence, as we shall see, any proposed solution to the problem of the origin of matter, life and mind will also contain, explicitly or implicitly, a specification of what is matter, life and mind. And that is where disagreements enter.

For it follows from this that as theories of the origin of matter, life and mind differ so do the definitional specifications of what they really are; and vice versa, as their definitions vary so do the theories of their origins. This is what makes scientific disputes about fundamental origins so difficult to resolve because the contending opposed theories are not explaining the same things, but conceptually considered quite different things. This is the reason that such disputes are so intractable and cannot be settled by reference to experimental evidence or observation alone, even when these are available or can be envisaged, which is not always the case. New experimental results or observations matter, but by themselves they cannot be decisive in settling fundamental disputes. Hence, such contentions continue, even though they continually assume new and updated forms.

All such points about the intractability of fundamental issues in science become very evident if we ask the first question about the nature and origin of things: "what is matter?" or, better put, "what is the fundamental nature of matter?" The philosophical differences that separate different schools of scientists in debating their various answers to this question are often hidden or disguised because of the forbiddingly technical, inevitably mathematical "language" in which they are couched. Only someone cognisant with the mathematical and experimental complexities has any understanding of what is involved, and such people, of whom there are not many in the world, are not prone to admitting that they are engaged in philosophical debates as much as scientific ones.

An exception is the renowned astrophysicist and mathematician George Ellis who has stated unequivocally:

> You cannot do physics or cosmology without an assumed philosophical basis. You can choose not to think about that basis: it will still be there as an unexamined foundation of what you do. The fact that you are unwilling to examine the philosophical foundations of what you do does not mean those foundations are not there; it just means they are unexamined.[4]

He especially sets himself against equally renowned physicists, such as Lawrence Krauss, Stephen Hawking and Neil de Grasse Tyson who have expressed strong opposition and animadversion against the idea of philosophy having any part to play in physics. As against them, Ellis refers himself to historical precedents where philosophy made a crucial contribution to physics, as in "Einstein's musings on Mach's principle [which] played a key role in developing general relativity [and] Einstein's debate with Bohr and the EPR paper have led to a great deal of good physics testing the foundations of quantum physics"; he also refers to his own recourse to "the Copernican principle in cosmology [which] has led to exploration of some great observational tests of spatial homogeneity that have turned an untested philosophical assumption into a testable – and indeed tested – scientific hypothesis".[5] And, of course, the very issue of whether all scientific theories need to be testable or falsifiable, which is such a crucial issue in current theories, such as multiverse theories and string theory, is fundamentally a philosophical issue.

4 George F.R. Ellis in an interview with John Horgan on July 22, 2014, reprinted in *Scientific American* (July, 2014).

5 Ibid.

Ellis has expounded and substantiated such more or less off the cuff comments in a highly mathematical and technical disquisition entitled "Issues in the Philosophy of Cosmology".[6] He raises a series of major themes, two of which are of particular relevance to us here: that of "explaining the universe – the question of origins", as he puts it; and that of "the explicit philosophical basis" for any such explanation. The first is more or less synonymous with our quest for the origin of matter, since matter is constituted at some indefinite point in the course of the origin of the universe. Unfortunately, we cannot follow Ellis on this issue in its full depth and complexity since it is a highly specialized field of astrophysics. We shall have to content ourselves with simpler accounts later. As for the second, we can quote Ellis directly:

> Consequently, it is inevitable that (as is also the case for other historical sciences) philosophical choices will to some degree shape the nature of cosmological theory, particularly when it moves beyond the purely descriptive to an explanatory role – that move being central to its impressive progress in recent decades. These philosophical choices will strongly influence the resulting understanding, and even more so if we pursue a theory with more ambitious explanatory aims.[7]

As we shall see in what follows, this general conclusion also holds for the other "historical sciences" where questions of origins are raised.

There are, indeed, scientists who have expressed the view that asking for the fundamental nature of matter or the origin of the universe is asking a metaphysical question that is unanswerable in science. The molecular biologist, Gunther Stent, puts this view very forcefully when he argues that physics is different from all other disciplines for it is not limited or bounded, and consequently it can never be completed:

> There is at least one scientific discipline, however, which appears to be *open-ended*, namely physics or the science of matter. Whereas the goals of the bounded disciplines are, in the last analysis, defined in terms of physical concepts, the goal of physics of understanding matter must necessarily remain undefined and hence hidden from view. In other words, it is difficult to envision a set of statements that would "explain" the nature

6 George F.R. Ellis, "Issues in the Philosophy of Cosmology" (February 5, 2008). https://arxiv.org/pdf/astsro-ph/0602280.pdf.

7 Ibid.

> of matter. For such an explanation can be provided only by *meta*physics, in the true sense of that term…[8]

But even worse than launching into metaphysical answers, such questions might have no answers because they make no sense, as Stent also surmises:

> Insofar as I am able to judge the frontier disciplines at the two open ends of physics, cosmology and high-energy physics, seem to be moving rapidly towards a state in which it is becoming progressively less clear what it actually *is* that one is ultimately trying to find out. What, actually, would it *mean* if one understood the origin of the universe? What would it mean if one had finally found the most fundamental of the fundamental particles?[9]

One need not take scepticism as far as Stent to see that there are problems not only about what is involved in giving answers to such fundamental questions, but what they mean.

There are analogous problems raised by Stent's own discipline when asking the question: "what is the nature and origin of life?" Thus the philosopher Evan Thompson, for example, makes it quite clear that the question of the origin of life is, by implication, bound up with the problem of the nature of life, that is, the answer to the philosophical question "what is life?"

> The question how and when life originated is inseparable from the question of what a living system is. If the aim is to determine how and when life arose, then one needs to clear a way of characterizing what distinguishes living systems from nonliving ones. Such a characterization could also serve as a standard or criterion for recognizing life elsewhere on other planets, or for determining whether anything we might someday synthesize artificially would qualify as a living being.[10]

But this is precisely where scientists differ, for where the origin of life is concerned, the science of life or biology is inherently bound up with the philosophy of life, or at least with fundamental issues that scientists can only approach philosophically. From this it follows that these are questions on which they are

---

8 Gunther S. Stent, *Paradoxes of Progress*, op. cit., 50.

9 Ibid, 50.

10 Evan Thompson, *Mind in Life: Biology, Phenomenology and the Sciences of Mind* (Cambridge, Mass: Harvard University Press, 2007), 95.

bound to differ on philosophical grounds. And this is, indeed, the case as various biologist start from quite different philosophical presuppositions as to what is life, and with such divergent definitions it is little wonder that they differ about the origin of life as well, as we shall see.

The problem in asking the question "what is mind?" is almost the converse of that of "what is matter", for whereas the latter is commonly assumed to be a wholly scientific issue, the former is taken to be a wholly philosophical one. Throughout the ages it is philosophers who have discussed and provided answers to "what is mind?" So much so, that it is commonly held that science plays little part in this debate. But this is far from the case if we remember that "what is mind?" is inherently bound up with the correlative question "what is the origin of mind?" The latter is, of course, a scientific matter, for there are a number of sciences that are specifically designed to deal with it, ranging from neurology to psychology and anthropology. And these sciences undertake their research with quite different, usually implicit, philosophical definitions of "mind" in mind. Hence, the usual wrangles that have emerged between philosophers and scientists and between different schools of philosophy and of science. Disciplinary specialization in recent times has made this quarrel of faculties much worse and the confusion even greater. It is rare to find people like Evan Thompson or his erstwhile collaborator Francisco Varela who are competent in many fields. They have developed at once an autopoietic definition of life in their science and a closely allied enactment definition of embodied mind in their philosophy to which we shall return later.

One way out of such difficulties that has appealed to some philosophers and scientists of late has been to return to an almost forgotten theme first raised in the early part of the twentieth century mainly in relation to evolution, that of emergence. Morowitz avails himself of this key notion as do Thompson and Varela and many others for it has become a burgeoning field of late. It promises to be able to provide an overarching paradigm or general theory able to encompass both many different sciences and philosophies. To what extent this is a viable idea, we shall test in the next chapter. However, it is a most promising departure at present, even though there is no agreed on definition of emergence or what it amounts to or entails. Different schools of emergentists have arisen who differ on fundamentals; thus there is now a division between "strong" and "weak" emergence and other such oppositions abound. It has also aroused strong opposition from both scientists and philosophers who attack it for being merely a mode of speech, a jargon that explains everything and nothing. Later in this work, in Chapter 5, we shall offer a more rigorous and closely argued definition of emergence, but for the present we shall continue to use the term in a looser sense for we cannot do without it.

Emergence is at its most salient and significant when it serves to explain the origination of fundamental novelty in the universe. The three most crucial novelties are the emergence of matter, life and mind. The importance of the first requires no justification for it really amounts to the origin of the universe itself. If that can be explained at all to any extent in science then it provides some partial answer to the old metaphysical conundrum first raised by Leibniz, "why is there something rather than nothing". And, indeed, there are scientists who believe that they can answer this question scientifically. Thus John Horgan reports that "Lawrence Krauss in *A Universe from Nothing* claims that physics has basically solved the mystery of why there is something rather than nothing".[11] But George Ellis, to whom Horgan puts this proposition, strongly disagrees. He puts his unequivocal objections in strong terms:

> He is presenting untested speculative theories of how things came into existence out of pre-existing complex entities, including variational principles, quantum field theories, specific symmetry groups, a bubbling vacuum, all the components of standard model physics, and so on… And he gives no experimental or observational processes whereby we could test these vivid speculations.[12]

Indeed, Ellis believes that the very possibility of testing experimentally or observing phenomena in regard to theories at this fundamental level is just about over, and Horgan agrees with him:

> …many of the possible high-energy physics experiments and astronomy observations relevant to cosmology are now in essence almost complete… So what we can see at the largest and smallest scales is approaching what will ever be possible, except for refining the details.[13]

According to Horgan and Ellis, cosmology and physics have entered a final stage of mathematical formalism and philosophical speculation. Grand theories, such as multiuniverse theories and string theories, are of this type; and philosophical assumptions are unavoidable in such sciences as we have already seen.

It is perhaps an indication of the purely speculative character of theories on the nature of matter and the origin of the universe at this level that the most

---

11 John Horgan interviewing George Ellis, op. cit.
12 Ibid.
13 Ibid.

famous practitioner in this field, Stephen Hawking, has changed his mind at least three times in radical ways in the course of his career, not because any new evidence came to hand, but simply because of his efforts to arrive at consistency and beauty of the mathematics which he continually revised. As a young theoretician, he began with the idea of the Big Bang as a singularity beyond the reach of time and space, beyond the mathematics of Einstein's equations, the point where they broke down. This was followed by his so-called "no boundary" proposal, which was set out in a paper published jointly with James Hartle in 1983, according to which the universe curved back on itself like a sphere, so it was effectively eternal. This implied the possible existence of a large number of different universes, of which ours is but one, each with its own laws, which in our case happen to be Einstein's laws. Now in a posthumous publication, together with Thomas Hertog, has appeared the third version of origin as Hawking changed his mind yet again. Based on the speculative physics of string theory, the two authors have come back to the idea of a boundary after all, holding that the Big Bang of 13.8 billion years ago was a kind of starting point beyond which there is nothing. But according to this latest proposal reality gets flattened out to a two-dimensional hologram from which the illusion of other dimensions is projected.

But what if we simply go forward from the Big Bang onwards; can we then determine when matter originated? It depends, of course, what we mean by matter for which there is no one agreed definition. However, following the process as set out in the Standard Model it is possible to describe what then ensued, as the physicist Steven Weinberg does. According to Weinberg: "at about one-hundredth of a second, none of the components of ordinary matter, molecules, or atoms, could have held together. Instead the matter rushing apart in this explosion consisted of various types of the so-called elementary particles…"[14] But after three minutes and forty six seconds deuterium, an isotope of hydrogen (H^2), can form; and from then on, it seems, it is all elementary physics to proceed to the kind of matter we now have, though that will take another 700,000 years.[15]

But can we be so confident that this is how it was in the beginning that this was, indeed, the process that gave birth to matter as we now know it? This is where physicists part company. Orthodox physicists, like Weinberg and most others, hold to what is called the Standard Model, which has held up in most respects extremely well. Weinberg writes that "the standard model of the early universe has scored some successes, and it provides a coherent theoretical

14 Steven Weinberg, *The First Three Minutes* (New York: Basic Books, 1993), 5.

15 Ibid, 112.

framework for future experimental programs", though what these might be he does not specify. However, less orthodox physicists are not so sure. Thus Chris Impey writes: "the origin of matter is a frontier topic in physics and cosmology. We still don't know how it happened and the only plausible ideas involve *untested* extensions of the Standard Model."[16] And there lies the rub – are such extensions testable at all? Namely, can there ever be any empirical confirmation of what took place in the first three minutes and for some considerable time thereafter, all of which is so confidently assumed in the name of the Standard Model?

Thus the main apprehension, as Impey puts it, is "the worry that we've ventured so far from the realm of testable lab physics that we may not have the tools sharp enough to discriminate between competing theories. If we crawl too far along the branch of speculation it may break."[17] But if it does break, then very few of us will know it, for only those with exceptional mathematical skills can follow lines of reasoning couched in the symbolism of such higher mathematics. As Impey admits, "it is very unfortunate that the quest to understand the fundamental nature of matter has led to formalism that makes the field opaque to all but an elite cadre of theoretical physicists."[18] The problem, however, is not that we do not understand the elite physicist, but that they do not seem able to understand each other, for they cannot agree as to which of the many theories proposed about the origin of matter are true and which are false. They cannot even agree as to how this it to be established.

According to John Horgan, a scientific journalist who has interviewed most of the star practitioners in the field, what is at stake is whether this whole field is any longer science in the accepted sense:

> Much of modern cosmology, particularly those aspects inspired by unified theories of particle physics and other esoteric ideas, is preposterous. Or, rather, it is ironic science, science that is not experimentally testable or resolvable even in principle and therefore not science in the strict sense at all. Its principal function is to keep us awestruck before the mystery of the cosmos.[19]

16 Chris Impey, *How it began: A Time Traveller's Guide to the Universe* (New York: Norton, 2012), 203.

17 Ibid, 339.

18 Ibid, 296.

19 John Horgan, *The End of Science*, op. cit, 94.

It would seem that the mystery of the origin of matter is also one that will never be experimentally or observationally resolvable for it exceeds the power of science. All attempts to resolve it Horgan considers to be what he calls "ironic science", and he is anticipated in this view by David Lindley, the author of *The End of Physics*, who considers string theory "aesthetic science".[20] Horgan calls the late universally acclaimed cosmologist Hawking "a master practitioner of ironic physics and cosmology."[21] And so, too, is Edward Witten, one of the founders of superstring theory, the most fundamental of all fundamental theories of matter, who "may also be the most spectacular practitioners of naïve ironic science... Naïve ironic scientists possess an exceptional faith in scientific speculations in spite of the fact that these speculations cannot be empirically verified."[22] He points out that to test superstring theory would call for a particle accelerator 1,000 light years in size.[23]

A superstring is "some kind of mathematical ur-stuff that generates matter and energy and space and time but does not itself correspond to anything in our world."[24] The reason for that is because it inhabits many more than the four dimensions of our world. But as to how many more dimensions superstring theory requires, there is no agreement among the practitioners. Witten with his M-theory has limited himself to ten dimensions but Brian Green has gone one better to eleven. String theory has come to be combined with another untested and untestable cosmological theory of the beginning of the universe as a quantum event followed by chaotic inflation, which was proposed by Alan Guth and Andrei Linde. One thus ends up with the latest version of the multiverse, which "hypothesizes a vast number of parallel universes, unobservable by us, each with randomly different properties and laws of physics."[25]

Physics and cosmology have clearly entered into a state of crisis, as many of its outstanding practitioners have recognized. Books on this subject by Lee Smolin and Peter Woit have recently appeared.[26] But this is not the first time such a thing has happened. In some ways it parallels the situation in Quantum

20 David Lindley, *The End of Physics: The Myth of a Unified Theory,* (New York: Basic Books, 1993).
21 John Horgan, *The End of Science*, op. cit, 95.
22 Ibid, 65.
23 Ibid, 70.
24 Ibid, 71.
25 Chris Impey, *How it Began*, op. cit, 349.
26 Lee Smolin, *The Trouble with Physics: The Rise of String Theory, the Fall of Science, and What Comes Next* (New York: Houghton Mifflin, 2006). Peter Woit, *Not Even Wrong: the Failure of String Theory and the Search for Unity in Physical Law* (New York: Basic Books, 2006).

physics as regards interpretation. Bohr's so-called Copenhagen interpretation in terms of complementarity was only the first, since then many more have appeared. At times it almost seems, as if every leading interpreter has his own version. But this has not mattered to working physicists since they are all agreed as to the basic mathematical formalism, which has amply proved itself in application time and time again. The problem at present with regard to string theory is that there is no longer agreement on the basic mathematical formalism or how it can be tested since it seems to have no application. How all this will ultimately resolve itself, or if it ever will, is a matter for future science which is in principle unpredictable. We have to live with and accept the confused situation as it exists at present.

Thus, all in all, tracing the origin of matter right back to the origin of our universe if not beyond is a many layered affair. There are aspects to it that occur within the limits of observation, either the astronomical of the macrocosm, looking up at the very large expanse of Hamlet's majestic canopy of stars we can see in the heavens above, or the physical observation of the microcosm, looking down into the very depth of matter, into Hamlet's nutshell of infinite space housing the particles we can find in our accelerators. Beyond either of these limits of the observable we enter into *terra incognita* where "we are into the more philosophical rather than physics area", as David Schramm has put it.[27] But even on the observable side of this great divide there is great uncertainty for the origin of matter is ultimately, as Ellis states, a historical issue, one of determining "*wie es eigentlich gewesen war*", as Ranke the nineteenth-century Roman historian expressed it in another context. And in history chance and accident always play a part. And this is what Howard Georgi, a particle physicist, has also expressed:

> I think you have to regard cosmology as a historical science, like evolutionary biology… You're trying to look at the present-day universe and extrapolating back, which is an interesting but dangerous thing to do, because there may have been accidents that had big effects. And they [cosmologists] try very hard to understand what features are accidental and what features are robust.[28]

Georgi recommends that cosmologists should read the works of his biological colleague at Harvard, Stephen Jay Gould. Gould is notorious for arguing that

27 Quoted in John Horgan, *The End of Science,* op. cit, 103.
28 Ibid, 104.

the origin of life is a historical matter in which chance and accident played a large role.

It might be supposed that once the primordial origin of the universe and matter is granted, all the rest of science is plain sailing, namely, that there is smooth continuity from then on without any hard to explain gaps in the further evolution of everything, as, indeed, Morowitz and many others assume. But this is far from the case. The universe did not proceed on a uniformly deterministic course from that point on. There are at least two major stumbling blocks on the way to this comfortable conclusion. Two other problems of origin arise that are at least as difficult to resolve as that of matter itself: the emergence of life and mind, which by no means simply follow from the origin of matter as such, as Terrence Deacon makes evident:

> 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 the cosmic transitions contorted the causal fabric of things as radically as did the appearance of life or of mind. Even though these living transitions only took place on a comparatively insignificant scale compared to other cosmic transitions, and even though no new kind of matter or energy came into existence with them, what each lacks in scale or cosmic effect they make up in their organizational divergence from the universal norm…[29]

The origin of life as we know it was certainly a momentous event in the universe, and we know that it took place on earth approximately four billion years ago. What we do not know is how this took place and whether it had already occurred elsewhere.

Is the origin of life a historical and more or less accidental event of a chance concatenation of favourable conditions that happened to come together on our planet, as Gould maintains? Or is it a permanent possibility of self-organizing matter that inevitably arises throughout the universe, as the geneticist Stuart Kauffman holds? Darwin's theory of how life evolves is of no help in this matter for the reasons that Horgan gives:

> Darwinism cannot tell us why life appeared in the first place, or why, once it emerged, it took the course it did. Scientists have proposed various auxiliary mechanisms to make life appear more probable and robust,

29 Terrence Deacon, *Incomplete Nature: How the Mind Emerged from Matter* (New York: Norton, 2012), 144.

> including group selection, Gaia, and complexity theory, but none are very plausible. The particle physicist [Steven Weinberg] once wrote, “The more the universe seems comprehensible, the more it seems pointless.” The history of biology suggests a corollary aphorism: the more life seems comprehensible, the more it seems improbable. The most wildly improbable organism of all is the one that can fret over its improbability.[30]

Most biologists would disagree with this conclusion, and if life was found to have also originated on other planets outside the earth, on Mars, say, then this would be evidence against it. But at present the debate goes on as proof of life elsewhere is lacking.

Even if life was found elsewhere, this would still not resolve the problem of how it could arise from matter in the first place. How can a self-replicating molecule come into being? There are at present numerous theories which differ radically as to the when, where and how this fundamental emergent event took place. The “how” is the most important aspect of the explanation, for if we knew that then the “when” and “where” could easily be deduced. How this process of chemical evolution took place has been approached in two ways, either bottom-up, as in the Miller-Urey experiments, starting with simple chemicals and seeking to build up increasingly more complex ones; or top-down as in the work of J. Craig Venter who began with the simplest life forms and progressively deducted genes so as to arrive at the most essential elements necessary for self-replication. Neither effort has gone very far or succeeded in providing very much towards elucidating the mystery of how life arose.

Most other attempts at an explanation are largely theoretical chemistry or computer driven algorithms. They usually begin with RNA molecules which are considered more basic than the more complex DNA ones. Autocatalysis or the self-assembly of catalytic sets is one of the most favoured approaches to this type of research. Manfred Eigen, Stuart Kauffman and Geoffrey Hoffmann have been most active in this endeavour. In his book *The Ancestor's Tale*, Richard Dawkins reports some interesting experimental findings on how autocatalysis might work in a proto-evolutionary manner to generate complexity.[31] There are numerous other such theories because there are so many aspects of life to consider; as well as self-replication, there is a cell formation, metabolism, proto-evolution and many more. Even to list the numerous proposals in their barest generality would take many pages, the Wikipedia entry on this

30 John Horgan, *The End of Science*, op. cit, 195.

31 Richard Dawkins, *The Ancestor's Tale: A Pilgrimage to the Dawn of Evolution*, (Boston: Houghton Mifflin, 2004).

topic is 28 pages long. Here we can merely review some of the issues that are in contention.

Another contentious approach to the origin of life is that which follows from the autopoiesis theory in biology developed by Humberto Maturana and Francisco Varela.[32] According to this approach, the originating phenomenon of life is neither a fully formed cell nor a self-replicating molecule, but is rather a certain self-sustaining form of organization, an autopoietic system. Autopoiesis characterizes not just biological life as we know it, but any form of minimal life that satisfies the conditions of the autopoietic system. These conditions are set out by Varela as follows:

> For a system to be autopoietic, (i) the system must have a semipermeable boundary; (ii) the boundary must be produced by a network reaction that takes place within the boundary; and (iii) the network of reactions must include reactions that regenerate the components of the system.[33]

How such a system might chemically constitute itself is the subject of computer studies of autocatalytic sets somewhat in the manner of Kauffman, but distinguishing itself from his proposal by insisting on "an autocatalytic network housed within and interdependently linked to a semipermeable membrane boundary",[34] for that alone can qualify as meeting the minimal conditions for life according to this theory.

As we have already seen by reference to Thompson, what is meant by "life" in the autopoiesis theory is very different to what is meant by "life" in other approaches. There is no agreed definition of what is life and there cannot be any as long as there are fundamentally different theories of the origin of life and how life constitutes itself in the first place. And behind every theory of life there is a different philosophy of life. Thompson identifies three such major philosophical approaches: those who look on life in the mechanistic terms of molecular genetics, of DNA replication and evolution by mutation, such as Dawkins; those who take an ecological environmental approach focussing on the interrelations between living beings and their niches, such as Lynn Margulis and James Lovelock; and, finally, those such as Maturana and Varela who take the autopoiesis route whose philosophical implications Thompson is intent on spelling out. To these one might add the previously referred to

---

32 Maturana, H.R. and Varela, F.J., *Autopoiesis and Cognition: The Realization of the Living,* Boston Studies in the Philosophy of Science, (Dordrecht: Reidel, 1980).

33 Quoted in Evan Thompson, *Mind in Life,* op. cit, 101.

34 Ibid, 105.

philosophy of those such as Gould for whom contingency and accident play a historical role in evolution. This is by no means a complete listing; and even apart from the philosophical differences that exist at present, others are bound to arise in the future; that is provided there are no major experimental breakthroughs or discoveries that will finally enable science to establish a complete and comprehensive theory of the origin of life. Such a theory is in principle possible, but how and when, if ever, it will eventuate, we cannot know.

The origin of life is perhaps less of a fundamental difficulty than the origin of matter because it can, at least in theory, be overcome. The scientifically most satisfying way of doing so would be to produce a living organism in the laboratory starting from scratch, from the bare chemicals that we know make up living cells. This is indeed what many are attempting to do starting way back with the work of Stanley Miller who first succeeded in synthesizing organic compounds in 1953. Thus far, however, there has been little further success in this endeavour. However, scientists such as Miller have gone on with synthesizing laboratory work and one day they hope to produce a self-replicating molecule. If such a program should succeed, it might lay bare the mystery of the "how" of the origin of life, for under laboratory conditions it might be possible to trace every step in the way that the basic chemicals come together to constitute an organism.

The discovery on other planets of organisms more primitive than the ones that have survived on earth might be a step in facilitating the recreation of life. But even if it were not, even if it did not show us how to assemble the constituents for life and make them literally spring to life, it would still be a major stage in understanding how it all might have happened in the first place. At present the most favourable location where life might have arisen and continued to exist outside the earth is on one of Saturn's moons, Enceladus. As reports indicate, analysis of data from the Cassini spacecraft has revealed extremely complex molecules in spumes of water, "orders of magnitude larger than anything that's been seen before", according to Frank Postberg at the University of Heidelberg.[35] As Kate Craft at Johns Hopkins University goes on to say, "we astrobiologists get excited about large molecules and that sort of thing because something is building upon itself and making itself more complex."[36] This does not itself indicate life forms, but it might indicate some organic state closer to it than anything still existing on earth. So it could provide vital clues as to how

35 Report by Sarah Kaplan in *Washington Post*, reprinted in The Age (Melbourne), 26 June 2018.

36 Ibid.

life arose in the first place. But whether it will in fact do that remains to be seen, and is still a long way off.

Vico's admonition that we can only understand what we can make, the so-called verum-factum thesis, might hold for life, but it need not necessarily do so. The mystery of the origin of life might not be resolved even if by sheer luck we succeeded in creating life forms. Thus in principle we could come to know the origin of life, but in practice we might never in fact do so, for there is so much against it. As Margulis puts it:

> Biological systems are effectively historical accumulations. So I don't think there is ever going to be a packaged recipe for life: add water and mix and get life. It is not a single-step process. It is a cumulative process that involves a lot of changes.[37]

As we have already noted, the origin of life is an extremely improbable occurrence according to some scientists. Fred Hoyle has famously declared that "the spontaneous generation of life on the earth would be as likely as the assemblage of a 747 aircraft by a tornado passing through a junkyard".[38] He thought that life would have to have originated elsewhere in space, which only postpones the question of how. Gould takes it for granted that it did occur on earth, not somewhere else in outer space, but for accidental causes that might not ever repeat themselves elsewhere. Gould believes evolutionary biology to be a historical science which cannot always retrospectively trace back the antecedent circumstances, the initiating causes. He is quoted as saying "if you're missing the evidence of antecedent sequences, then you can't do it at all. That's why I think we will never know the origins of language. Because it is not a question of theory; it's a question of contingent history."[39] Presumably he believes that what holds for language also holds for life. Gould strongly disagreed with those biologists and palaeontologists whom he castigates as "naïve inductivists": "they actually think that… if we keep at it long enough we really will know the basic features of the history of life and then we'll have it."[40]

Gould is expressly arguing against the views and scientific practices of biological theoreticians such as Kauffman and many other mathematical determinists who hold that the conditions for self-organization, up to and including life itself, are inherent in matter as such, and consequently that life is necessary

37 Quoted in John Horgan, *The End of Science*, op. cit, 139.
38 Ibid, 106.
39 Ibid, 124.
40 Ibid, 124.

and ubiquitous. Kauffman is working on autocatalytic sets which, he thinks, are the precursors to living cells. He is also building computer models to demonstrate that order arises spontaneously from certain kinds of network arrangements, which he calls "order for free". He has expressed himself explicitly in these terms against Gould:

> But at least in terms of mathematical models, here you have a body of models that says that the emergence of life might be a natural phenomenon, in the sense that given a sufficiently complicated set of reacting molecules, you'd expect to crystalize autocatalytic subsets. So if that view is right… then we're not incredibly improbable accidents… And, therefore, we are at home in the universe in a different way than we would be if life were this incredibly improbable event that happened on one planet and one planet only because it was so improbable that you wouldn't expect it to happen at all.[41]

Many other biologists are, understandably, contemptuous of Kauffman's efforts at demonstrating the necessity of life by computer modelling. Miller calls it "paper chemistry": "running equations through a computer does not constitute an experiment."[42] John Maynard Smith and Margulis, who have also written extensively on this topic, have been equally dismissive. They have their own deterministic theories. The problems they all confront is put by Morowitz as follows: "the ultimate emergence of metabolism seems embedded in the laws of chemistry, but the reactions are a tiny subset of all possible organic reactions. We must search for the primary algorithm."[43] It remains to be seen whether any such will ever be found.

There are many views and many theories on the origin of life. None as yet commands the universal assent of an established scientific body, and perhaps none ever will. There are numerous parties to this dispute, but they line up approximately in two groups, which Morowitz sets out as follows:

> If one believes that the steps from protocells to prokaryotes compose a series of frozen accidents, then we shall never recover the processes and can only study the end states. If one believes that these emergences are rule driven and highly deterministic, then we can look forward to an ever-increasing understanding of how the laws of the universe, including the

41 Ibid, 133.
42 Ibid, 139.
43 Harold Morowitz, *The Emergence of Everything,* op. cit, 76.

> laws of emergence, have led to the molecular biology that we have come to know and understand.[44]

The differences separating these antagonistic positions are more than purely scientific. They are deep divisions of *Weltanschauung* that are primarily philosophical, but perhaps also ideological. Horgan puts it in these terms when he outlines what separates Kauffman, Gould and Margulis. Kauffman "feels that accident alone cannot have created life; our cosmos must harbor some fundamental order-generating tendency."[45] He holds this view as much on the basis of "philosophical conviction about how things must be as by scientific curiosity about how things really are."[46] Margulis favours a holistic ecological approach that goes against neo-Darwinist reductionism. Gould also opposes neo-Darwinism, but for very different reasons, he is against all ideas of inevitable order and progress in nature and stresses instead randomness, contingency and the role of historical accidents in evolution and the origin of life. And so the debate goes on.

Thus far we have briefly outlined the debates surrounding the question of the origin of matter and life, now we turn to the main subject of our work – the origin of mind. In relation to mind, the question of origin is perhaps even more crucial than in the two other cases, as the neurologist Gerald Edelman points out:

> Ignoring the origin of things is always a risky matter. It is even more risky in any effort that purports to explain mental events. But this is exactly what has happened in much of the history of psychology and the philosophy of mind. I guess this is so because thought is a reflexive and recursive process. It is therefore tempting to think that the nature of thinking can be uncovered by thinking alone.[47]

We shall pay heed to Edelman's admonition and devote much of this book to the sciences of mind focussing on the origins of mind, which is primarily a matter of evolution. For without such an investigation of origins we cannot arrive at the nature of mind relying on speculative or philosophical thought alone, as philosophers are wont to do.

---

44 Ibid, 83.
45 John Horgan, *The End of Science,* op. cit, 35.
46 Ibid, 35.
47 Gerald M. Edelman, *Bright Air, Brilliant Fire* (New York: Basic Books, 1992), 33.

However, the science of how mind originates is itself a highly contentious and partly philosophical matter; here we shall merely present a preliminary outline of some of the issues involved which will be taken up further in Part 2 of this work. On the issue of origin, there are multiple theories and hugely varying accounts and interpretations. How mind emerges can be studied scientifically from numerous points of view, which are far from coherent with each other. There is the anthropological perspective which considers how the species *homo sapiens* of which we are a branch evolved and developed the kind of mind which we now possess and which distinguishes us from our animal ancestors. Then correlative with that, there is the approach which considers the development of the mind in the human individual, which embraces numerous ways of studying the capacities of neonates and the changes they must undergo to become children and eventually adults. The anthropological approach and the individual developmental have been in a highly contentious relation to each other ever since Ernst Haeckel coined the memorable formula that ontogenesis recapitulates phylogenesis, which many since, above all Gould, have hotly disputed.[48]

This is only the start of the disputations that surround the sciences of mind. As soon as one enters the arena of the relation between mind and brain then one is in the thick of another battle. How does the mind emerge from the brain? What does it even mean to say that it does so? There is now general agreement that the mind originates from the brain and that it depends on it at least in this earthly life. Earlier views, such as those we encounter in *Hamlet*, and even in later religion and philosophy, that souls enter bodies from the outside, as it were, are now mostly discredited; but not completely so, as vestiges of them are still maintained by some eminent scientists, such as the Nobel prize neurologist John Eccles, who held that a creationist immaterial impulse must act on the brain if the body is to undertake freely willed actions. But even setting aside such out-dated ideas, there still remain a host of scientific and philosophical theories about how mind emerges from brain.

The physiologists and neuroscientists approach the issue, as it were, from below starting with what they can discover about the functioning of the neurons in the brain and how the different parts of the brain are connected to each other. The psychologists start from above by studying the mental capacities and competences of human subjects both in the laboratory and under normal conditions. Like two groups of tunnellers boring through a mountain, they hope to meet and join up somewhere in the middle. But so far this juncture is very far from accomplished, the gap between the two sides is huge and there

48 Gould, S.J. *Ontogeny and Phylogeny*, (Cambridge: Cambridge University Press,, 1977), 168.

are only very tentative and unsubstantiated theoretical suggestions as to how it can be spanned. Such proposals differ depending on the various schools on both sides of the divide from which they derive. There are models of the brain such as that first proposed by Hebb which seem to be capable in principle of generating structures that correspond to mental phenomena, but in practice they invariably prove to be too general and useless in detailed research. There are psychological models – such as the electrical Gestalts in the brain hypothesized by Kőhler at about the same time as Hebb but from an opposite point of view – which seem to tell us something about how the brain works, but these, too, have proved incapable of being experimentally demonstrated. There are, of course, many more up-do-date proposals along such lines by outstanding scientists, both neurologists and psychologists whom we will consider later in this work, but the two ends are still far from meeting, and perhaps they never will.

Indeed, there is a school of philosophers, derisively called "mysterians", who maintain that the human mind is in principle incapable of knowing itself and that only some "superior" mind is capable of understanding us as we can understand some lower creatures. But before we succumb to any such "mysterian" scepticism, it were best to examine how far science has already gone in broaching the mystery of mind and how much further it might reasonably still be expected to go. For there is no doubt that we are continually gaining more and more knowledge about the emergence of mind from all the points of view that we previously adumbrated, the phylogenetic, ontogenetic, neurological and psychological, to list but the main sciences involved, leaving out for the moment such more recent arrivals on the research scene as computer simulation and artificial intelligence. Whether these sciences will ever reach some kind of consensus about the mind which will be labelled the "truth" is impossible for us to predict. We can only maintain a sceptical *epochḗ* about any such conclusion and neither affirm nor deny it, but suspend judgement.

One of the main reasons why the origin of mind is such a difficult and contentious issue is because most of the evolutionary evidence to arrive at any firm conclusion is missing. The missing link between apes and humans cannot be filled despite the few tantalising material clues by way of skeletal shapes, stone tools and graphic signs that are continually being unearthed. Each new discovery transforms the picture, but it only helps to complete a small corner of the total puzzle for which all the other pieces are still missing. The complete picture might never be recoverable.

The crucial evidence is simply no longer there for us to be able to survey how the human mind originated. The soft tissue from the skulls and bones of prehistoric people has disappeared without trace; the behaviour of the people who shaped the stone tools, the social structure that enabled them to transport and trade them can barely be surmised, and the cognitive capacities that enabled the remarkable paintings on cave walls to be created can only be guessed at. Did these early homo sapiens possess the kind of language we have or some primitive precursor to it? We cannot answer this; and, as Gould avers, we might never be able to answer it, for we do not know the origin of language, which is so crucial to the origin of mind. Chomsky refuses even to consider the matter, merely insisting that our language is fundamentally different from any animal communication, and so, by definition, is our mind from animal mind. But in what the difference consists, about that there is little agreement, as we shall see in Chapter 3.

But if we do not know where we come from, how can we know what we are? As with life, the problems of what is the origin of mind and what is the nature of mind are closely related. As we shall see repeatedly, those who differ about the former also differ about the latter; as Terrence Deacon puts it, "assumptions about the nature of language and the difference between non-human and human minds are implicit in almost every philosophical and scientific theory concerned with cognition, knowledge, or human social behaviour."[49] Deacon goes on to explain what is at stake in this question of origins:

> Knowing how something originated is often the best clue to how it works. And we know that human consciousness had a beginning. Those features of our mental abilities that distinguish us from all other species arose within the handful of million years since we shared a common ancestor... If we could identify what was different on either side of this divide – differences in ecology, behaviour, anatomy, especially neuroanatomy – perhaps we could find the critical change that catapulted us into the unprecedented world, full of abstractions, stories and impossibilities, that we call human... Biologically, we are just another ape. Mentally, we are a new phylum of organisms. In these two seemingly incommensurate facts lies a conundrum that must be resolved before we have an adequate explanation of what it means to be human.[50]

Deacon doubts whether we are anywhere near such an adequate explanation. And it is to be doubted whether we will ever arrive at a scientific consensus that

49 Terrence W. Deacon, *The Symbolic Species: the Co-Evolution of Language and the Brain,* (New York: Norton, 1997), 14.

50 Ibid, 23.

commands universal rational assent. There might always be huge theoretical differences just as at present, so it is likely to remain a disputed matter in which philosophy plays as much of a role as science. Indeed, the two can barely be distinguished in this field. Deacon notes that many scientists are confident that they can solve these problems without the benefit of philosophy: "advances in the study of human evolution, the brain and language processes have led many scientists confidently to claim to be closing in on the final clues to this mystery. How close are we?"[51] In fact, as Deacon himself admits, we are still not even close:

> Despite all these advances, some critical pieces of the puzzle still elude us. Even though neural science has pried ever deeper into the mysteries of brain function, we still lack a theory of global brain function. But the most critical missing piece of the puzzle is access to the brains in question: ancestral hominid brains… With respect to fossil brains, we will never find the "smoking gun" – the first brain capable of language.[52]

According to Deacon, language is key to the origin and nature of mind: "ironically, then, the problem of language origin may actually offer one of the most promising entry points in the search for the logic linking cognitive functions to brain organization."[53] Many palaeoanthropologists and neurophysiologists are likely to agree with this; but, as might be expected, there are some who do not. Merlin Donald is one such. He holds that the importance of language for cognition and for consciousness has been vastly exaggerated, by no one more so than Chomsky, as we shall see. But he is critical even of Deacon's more moderate claims for the role of language in the origin of mind:

> Deacon proposes a theory of language evolution that tries to finesse the need for a mimetic preadaptation. He argues that hominids made the transition from primate "indexical" representation (as shown in operant conditioning) to fully symbolic representation (as shown in language) in one continuous evolutionary progression. This idea does not account for the metaphoric nature of language, the existence of mimetic skill, or the fact that language normally develops in a mimetic framework and

51 Ibid, 23.
52 Ibid, 24.
53 Ibid, 25.

> operates by metaphoric or mimetic principles. Moreover, it cannot scale the wall set up by the primate zone of proximate evolution.[54]

What Donald's last point amounts to we shall see later in Chapter 3, when we will undertake a more detailed comparison of the various theories of language origin and mind in general.

Deacon does not reply to or even refer to Donald's criticisms in his subsequent book, *Incomplete Nature*, written more than a decade later.[55] The reason for that omission is perhaps one that he had already prefigured in his earlier book:

> I have not attempted in any systematic way to review or compare the many alternative explanations proposed for the phenomena I consider… An exhaustive review of competing explanations would require another book at least as long as this one.[56]

Indeed, this points to the huge diversity of scientific and philosophical positions on almost every point raised on this issue of the origin of language and mind. However, Deacon does not forebear to spare Chomsky's contentious view, which is far too influential to be ignored, having been "echoed by numerous linguists, anthropologists, philosophers and psychologists". He sets out his objections as follows:

> Interpreting the discontinuity between linguistic and non-linguistic communication as an essential distinction between humans and non-humans, however, has led to an equally exaggerated and untenable interpretation of language origin: the claim that language is the product of a unique one-of-kind piece of neural circuitry that provides all the essential features that make a language unique (e.g. grammar). But this does not just assume that there is a unique neurological feature that correlates with this unique behaviour, it also assumes an essential biological discontinuity. In other words, that language is somehow separate from the rest of our biology and neurology. It is as though we are apes *plus* language – as though one handed a language computer to a chimpanzee.[57]

54 Merlin Donald, *A Mind So Rare: The Evolution of Human Consciousness*, (New York: Norton, 2001), 341.

55 Terrence W. Deacon, *Incomplete Nature*, op. cit.

56 Terrence W. Deacon, *The Symbolic Species,* op. cit, 14.

57 Ibid, 34.

Does Chomsky deserve such stringent criticism? Chomsky is certainly highly sceptical of Darwinian or evolutionary explanations of language in general. In an interview with Horgan, he has expressed the opinion that "given the enormous gap between human language and the relatively simple communication system of other animals, and given our fragmentary knowledge of the past, science can tell us little about how language evolved."[58] He maintains that language may not have arisen in response to any direct selection pressure, it "may have been an incidental by-product of a spurt in intelligence that only later was co-opted for various uses. The same may be true for other properties of the human mind."[59] This is obviously anathema to any evolutionary biologist or palaeoanthropologist like Deacon or Donald. According to Chomsky, we can explain on evolutionary grounds those features of the human anatomy or physiology where there are some homologous structures in other animals. But in the case of language and the human mind, these are missing.[60] Hence Chomsky dismisses evolutionary psychology as "a philosophy of mind with a little bit of science thrown in."[61]

Chomsky does not hesitate to appeal to a "language organ" in the brain, a Language Acquisition Device, for which there is no evolutionary precedent. It seems to Chomsky that without some such mechanism able to house "Universal Grammar" there would be no way of explaining how it is possible for children to learn language with such speed at such an early age given such a "poverty of stimulus", as Chomsky calls it, that is, so little language material provided by their environment. This is what Deacon has castigated as a "hopeful monster" theory of language, "as if one handed a language computer to a chimpanzee". And it does seem like a kind of "deus ex machina" in the brain, an ad-hoc resolution of the awkward predicament that Chomsky has got himself into by denying the constitutive role of evolution.

Thus far we have explored the various kinds of mysteries surrounding the scientific attempts to explain the origins of matter, life and mind. And in each case we have seen that there are problems in explaining the emergence of these fundamental entities to which science at present has no unequivocal answers; instead there are numerous theories that contend with each other almost like the proverbial thousand schools of philosophy in traditional Chinese thought or like those among the Greeks at the very start of philosophy. As more knowledge is gained, only more such scientific "schools" arise to dispute the

58 John Horgan, *The End of Science,* op. cit, 151.

59 Ibid, 151.

60 Ibid, 177.

61 Ibid, 178.

issues that this opens up. Whether they will ever converge, as we have already indicated, is problematic and unknowable. But science does go on, and we have no reason to believe that it will ever cease. The "end of science" theses now being mooted which we have already touched on, such as the views of Stent and Horgan, even if true, which is far from the case, do not necessitate an end to scientific research, only at most how that research is evaluated, and what might be expected to eventuate from it. But, as we shall see, that pessimistic outlook can at present not be vindicated, for there is still so much that science can accomplish.

2 Condensed Matter and Animal Minds

The three issues of the origins of matter, life and mind are the fundamental ones in science, the ones which establish the starting and finishing points of the Emergence of Everything, the alpha and omega of the Great Chain of Being on which Hamlet relies in placing Man at the apex of terrestrial creation in the manner of Robert Fludd's *Utriusque Cosmi Historia* (1617) published just prior to Descartes' works. We have slightly rearranged the old model in the light of modern science, but basically it is still intact. We can thus proceed to elaborate it further in the traditional manner by placing two more emergent punctuations above and below the origin of life in the middle: the first between life and human mind and the second between life and matter. Between the origin of life as such and the origin of human mind lies the upper intermediate realm of animal mind. And between the origin of matter as such and the origin of life lies the lower intermediate realm of solid-state or condensed matter which incorporates the numerous types of material substance of both physics and chemistry. Altogether this gives us the magic number of five to satisfy Hamlet's idea of quintessence or five essential staging posts of emergence from the dust of mere sub-atomic particles till Man, the crown of evolution. But, as we shall see later, this is in some respects an arbitrary play with numbers since evolution is a continuum that can be sliced up in many ways. But for our present purpose five plus or minus one is the smallest possible and therefore right number.

In this section we proceed to briefly explore the two intermediate modes of emergence – solid-state or condensed matter and animal mind. Both present problems of origin. The problem we confront with the emergence of animal minds is that we have no clear and non-arbitrary way of specifying when it began. In other words, there are no markers separating animals that in some incipient sense have minds from those that do not. Yet it is obvious at least to common sense that there are lower species that by no stretch of the imagination can be said to have minds, such as bugs and worms, and higher species that could be denied minds only on spurious philosophical grounds, such as Descartes entertained. But where lies the demarcation between the two? As we shall see when we return to this issue in Chapter 3, different scientists give widely varying answers to this problem, though mostly implicitly so for few even realize that it is a problem. Philosophers, too, have paid scant attention to this matter.

An overview of the positions adopted on this issue shows that there are three distinct answers given. There are those who maintain that all or nearly all animals have minds or at least some elements of mental activity or even incipient consciousness. At the opposite pole, there are those who hold that no animals have mind, apart from humans. And in between the two extremes, most uphold the view that only some animals have minds, but they do not specify or tend to disagree as to which these are. In the first category fall ethologists like Donald Griffin and scientists like Morowitz who go even further and believe that "mental activity is universally distributed throughout the animal kingdom…"[62,63] On the diametrically opposite side are the classical Cartesians who hold that animals are machines or the more up-to-date Behaviourists according to whom animals, including humans as well, are nothing but stimulus and response mechanisms, or even Chomskian for whom language is the precondition for mind. In the moderate mean are thinkers like Gregory Bateson who wrote that "in this matter, I prefer to follow Lamarck who, setting up postulates for a science of comparative psychology, laid down the rule that no mental function shall be ascribed to an organism for which the complexity of the nervous system is insufficient."[64] But what is sufficient complexity for mind he did not specify, and neither do most others.

Which of these points of view on animal mind is correct and how this can be assessed is a crucial issue in the philosophy of mind which few philosophers have seriously tackled. For if we wish to understand how mind emerges from matter we should begin by dealing with it at the point of origin when mind first arises, that is, when the first animals that might be said to have minds or mental capacities originate. What is it that distinguishes such animals from those that do not have minds? Why do we need to ascribe mental attributes to them? Why is it impossible to treat them purely in a Behaviourist manner? What does emergence mean in this context? It is by seeking to answer such questions first,

62 Donald Griffin, *Animal Minds* (Chicago: Chicago University Press, 2012).
63 Harold Morowitz, *The Emergence of Everything,* op. cit, 138.
64 Quoted in ibid, 104.

before tackling the much more daunting issue of human minds that we might hope to arrive at a better grasp of the latter as well.

Human mind did not arise through a special dispensation or creationist fiat, it had to go through a natural evolutionary process of emergence from animal mind. As most scientists and philosophers recognize, there is a huge disparity in both quality and quantity of mental powers separating humans, the species *homo sapiens*, from its nearest animal relatives, the species of apes, such as chimpanzees. There are, of course, those who minimize this difference or discount it altogether; even Darwin himself was somewhat inclined in this direction, as is the current ethologist David Premack. Nevertheless, most accept that there is a huge divergence between these species of apes and humans.

Obviously there once existed intermediate species that have disappeared which is known as the "missing link" that naturalist and palaeontologists have been trying to fill since Darwin. To some extent they have been successful, at least as far as the physical features of these intermediate hominin species are concerned and something of their material cultures is also known. But almost nothing is known about their minds. We know a great deal about our own minds and quite a lot about the minds of chimpanzees, but how the former emerged from the latter is largely hidden in the missing link. Hence, in that anthropological sense we do not know the origins of our own minds.

But that does not mean that science is completely baffled or that it is confronted with an insuperable mystery. We know for certain that an evolutionary process of emergence took place and we have considerable grasp of the two ends of this process, of the *terminus ad quo* and the *terminus ad quem*. However, filling in the intermediate sequence is not the same kind of scientific undertaking as in other fields of evolution where there might also be missing links, for these can be filled by hypotheses about physical continuities of development which can be established with a high degree of precision. And even missing links between one type of animal mind and another – such as, for example, how dinosaurs evolved into birds and how the extraordinary intelligence of some bird species, such as ravens, arose – can also be hypothetically closed because we are dealing at both ends with the same kind of mind, animal mind. But it is different when it comes to filling the huge gap between chimpanzees and humans for a different kind of mind has emerged, not just a mind that is conscious and intelligent, which chimpanzees certainly are, but one that is self-conscious and rational, which chimpanzees are not. Even such a minimal capacity as language learning is way beyond the competence of chimps, not to speak of higher faculties.

Nevertheless, despite these extraordinary difficulties, science does not come to a halt before the mystery of human mind but seeks to fill the evolutionary lacuna with probabilistic hypotheses as to what might the intermediate stages have been. Unfortunately, there are numerous such theories none of which can be conclusively ruled out or demonstrated. Each is, nevertheless, backed up by plausible and well-reasoned arguments which have to be judged according to the criteria of sound reasoning such as we employ in other fields where no experimental proof or disproof is available. In these respects, then, the situation is closer to philosophy than to most other fields of science. And, indeed, as we shall see, there are always philosophical or other speculative assumptions inherent in the various theories that have been proposed by palaeoanthropologists and evolutionists.

It is clear from the previous account that we have no single, unambiguous and well established theory of the origin of the human mind. Instead we have a huge plethora of possible scientific theories, even though the data available is the same for all theorists. This means that we do not know the origin of the human mind in any purely scientific sense even to the extent to which we know the origin of the animal mind or better still the origin of anything else where a clear evolutionary lineage of emergence can be traced.

But if we do not know with any degree of scientific certitude the origin of human mind, then how can we know the nature or essence of mind? For as we previously saw, knowing or not knowing the origin of something is not separate from knowing or not knowing the nature of its being. Knowing the stages through which something emerges is always a vital lead in determining how it functions and, therefore, what it really is and how it manifests itself. This is particularly true of the evolutionary processes of the brain on which mind depends. The brain is an organ of such extraordinary complexity that we cannot simply study how it is constituted and functions without some comparative sense of its earlier stages and how it relates to these. Thus the brains of the lower organisms give physiologists and neurologists the starting point for understanding and explaining the functioning and nature of the brains of higher organisms. And this in turn provides the crucial leads for explaining the faculties and capacities of animal minds.

Animal minds can be studied and their functions and natures determined with considerably greater scientific exactitude than human minds not only because they are inherently simpler or because a much greater variety of experimentation is possible since there are fewer ethical restrains which restrict similar procedures on humans, but also because there are clear continuities between lower and higher species. There are no such missing links in which rapid and drastic evolutionary changes occurred as we have seen is the case in the emergence of the human mind. The difference between one species and another are small, incremental and continuous. It is for all these reasons that we can have a far better grasp of animal minds than of the human mind.

We have thus far confronted four questions concerning origins and there is one more which is located between matter as such and the emergence of living matter; it concerns the origin of macroscopic from microscopic matter. Or putting it in modern scientific terms, how does the matter of solid state physics arise from the matter of particle physics? As we saw in the previous section, we know something concerning the emergence of matter at some short interval beyond the Big Bang when the universe had cooled sufficiently to permit the emergence of the first primary substances, hydrogen and helium, from elementary atomic particles. We also know how the other elementary substances were, so to speak, "cooked" in the stars and released in their disintegrating explosions into space, there to coalesce into planets such as our own solar system in which our earth is situated. We also know the chemical laws by which these substances combined to form the organic matter which is the precondition of life. However, there still remain the vexed and unresolved problems of the relation between the two states of matter, atomic or sub-atomic and solid state or condensed. The main problem is that they are governed by quite different types of laws in physics, the former by Quantum mechanics and the latter by Classical mechanics.

The problem of relating Quantum laws to Classical laws is an ongoing field of research in physics that is both theoretically and technically extremely advanced. Hence we cannot deal with it here except in the vaguest generality. It requires an understanding of concepts such as quantum superposition, entanglement and decoherence; these are explained in an article by Erich Joos "The Emergence of Classical from Quantum Theory".[65] As might be noted from that title, this is a fundamental form of emergence. The key to this form of emergence is decoherence which transforms objects governed by Quantum rules into ones governed by Classical rules. There are experimental techniques to control the rate of decoherence such that quantum effects gradually vanish as decoherence takes on. It is now possible to do this with much larger objects than single particles through creating superpositions of large numbers of particles to create states of complex entanglement. But the larger the objects entangled, the less it becomes possible to deduce their quantum behaviour, for this is a case of emergence.

Schrödinger's cat paradox is governed by this unpredictable feature of emergence, as Jack Cohen and Ian Stewart explain:

65 Erich Joos, "The Emergence of Classical from Quantum Theory", in *The Re-emergence of Emergence*, eds. Philip Clayton and Paul Davies (Oxford: Oxford University Press, 2006), 53–78.

> And there is no simple mapping from wave-function space to cat space. The specification of the animal's wave-function is almost completely useless if you want to know any common sense features of the cat – for example, what colour is its left ear. Imagine the computer calculation you would have to perform in order to answer such a question! You would have to process the cat's entire wave-function, somehow extract information about the atoms in its ears (tricky, because in quantum mechanics everything interacts with everything else), find out which atoms are on the surface of the hairs of the ears (equally tricky for the same reason), analyse their response to incident light of various wavelengths…[66]

Clearly, this is an undertaking, which though in principle possible, is in practice so difficult as to be unthinkable and meaningless. This is the kind of "in principle possible" which tells us nothing about what makes any sense whatever. We shall encounter it over and over again where issues of reduction are involved, for if we are told that something is in principle reducible to something else this tells us almost nothing about what deduction from this is actually possible. Usually it serves as a gesture, like waving the reductionist flag and claiming victory in the battle for or against reduction.

What is at issue in these matters is the dispute that has arisen between two kinds of physicists, those who work on sub-atomic matter, such as paradigmatically Murray Gell-Mann the discoverer of quarks, and those who work on solid state matter, such as most eminently Philip Anderson. The former hold that reduction in principle is all that matters; the latter believe that such reduction tells us almost nothing about higher levels of complex matter. The difference between them is more than just an intellectual dispute in theoretical physics, for issues of prestige, finance for research funding, and even politics are also involved. This became evident in the late 1980s during the hearings in the US Senate sub-committee on constructing a Supercharged Super Collider to delve ever deeper into the fundamental constitution of matter and possibly arrive at an answer to the question of the origin of matter. This was opposed by the opposite school of physicists who held it to be a relative waste of scarce resources compared to the problems that were still unsolved at higher scales of physics, where work could be much more usefully done and bring better practical benefits. We shall return to this in the next chapter.

It is understandable that after his great reductionist triumph in showing how the veritable zoological host of sub-atomic species can all be derived from

66 Jack Cohen and Ian Stewart, *The Collapse of Chaos: Discovering Simplicity in a Complex World,* (New York: Penguin, 1940), 269.

a few underlying sub-sub-atomic entities he called quarks, Gell-Mann was justifiably proud of his achievement, but at the same time he maliciously sneered at the work of solid state physicists like Anderson. He called is "squalid" state physics. He was echoing Dirac's earlier statement, after he had discovered the positron, that all the rest was mere chemistry. And just as Dirac did not anticipate the vast array of sub-atomic particles that would subsequently emerge, so Gell-Mann did not foresee the squalid state into which fundamental physics would sink with string theory. Anderson, not unreasonably, felt aggrieved by this kind of contempt heaped on his science by the doyens of high-energy physics and exponents of reductionism. He fired back in reply with his epoch making papers of the 1970s with which we will deal in the next chapter. These have since revealed themselves to have been some of the earliest intimations of the re-emergence of emergence in science after a half century of hibernation. Emergence, which had been unmentionable among respectable scientists and utterly derided among philosophers, once more came back on the scene with a vengeance.

What is at issue in the debate between Anderson and Gell-Mann, as representatives of the two types of physics, is precisely the issue that concerns us here in determining the nature of the emergence that takes place from the state of matter at the microscopic level to the macroscopic level, that is, the emergence of solid-state substances from atoms and elementary particles. Is this simply a matter of reverse methodological engineering, as it were, of inverting reduction into emergence, as Gell-Mann seems to have believed, or is it a much more complex relation, as Anderson argues? Anderson points out that it is not possible even in principle to deduce macroscopic entities from what we know about microscopic ones. To take a very simple example that Anderson refers to, it is not possible to derive the substance water and its chemical properties or the laws governing its behaviour from our knowledge of the atoms of hydrogen and oxygen, even though reductively speaking we know that water is simply H_2O. It is perhaps conceivable that there is some vast set of Schrödinger quantum theoretic equations that would enable such a deduction to take place, but not only are such equations difficult to formulate, but it is impossible that they should ever be solvable. All this takes us back to the initial problem of relating macroscopic to microscopic phenomena in physics, which, as previously explained, is highly theoretical and technical and so cannot be any further gone into here.

We have now completed our survey of the fundamental types of emergent phenomena and we find that there are five such in total. Three concern fundamental origins: the origins of matter, life and mind; and two concern the intermediate origins: the origins of condensed matter and animal mind. All of the five forms of basic emergence present different problems that are at present unanswerable in various ways and for very different reasons. Some might or might not be answered by future science; we have no way of knowing. The problems that they raise are all *sui generis* and have almost nothing in common – which shows that emergence is a very varied enterprise. The long preached unity of science and the uniformity and ubiquity of scientific method throughout the sciences is largely an illusion brought about by the success of reductive science which can in principle explain every entity in terms of a more fundamental entity, until it seems as if the basic laws of sub-atomic entities explain everything. But as soon as we reverse the reductive methodological process and proceed in the converse emergentist direction then problems that are insoluble arise, such as, above all, the five kinds of origins.

In terms of temporal progression, the five instances of fundamental emergence line up as follows: original matter beginning with the Big Bang; solid state matter starting at some later date with the cooling of the universe and later still with formation of stars and eventually planets; living matter which began on our planet some four billion years ago and perhaps also on other planets or satellites before or subsequently; animal minds emerged, though it is not definite at what evolutionary stage this occurred; and finally, self-conscious and intellective mind which only came with the evolution of homo sapiens, that is, Man and possibly some related species of hominins that arose less than half a million years ago. In this hierarchy of emergence, the later forms depend on the earlier ones; so that working backwards, we can say that human mind could not have emerged without animal mind and that without life and life without solid matter and so on. In that sense it all constitutes the one continuum of matter in its multiple forms.

Dividing up the continuum of matter into five major levels might seem like a purely arbitrary matter. But it is no more arbitrary than dividing up the colour spectrum into the four primary colours. There are caesuras and breaks in nature and the five we have identified are unavoidable. Obviously, as with colours, one can add many more differentiations and shades of difference as painters are bound to do with colours or philosophers of science with the whole continuum of emergence in nature. Hence different authors, basing themselves on different considerations, have come up with different numbers of subdivisions. Morowitz has separated as many as 28 levels of emergence and other authors have equally larger numbers.

But it is surely no coincidence that George Ellis has also settled for five basic levels of emergence. Though his subdivision does not quite coincide with our own because he surprisingly omits the emergence of matter itself, on which subject he is a world authority. Hence, his schema actually amounts to six basic types and we might settle on that if pressed.

Ellis' first level is that where "bottom-up action leads to higher level generic properties but not to higher-level complex structures or functions."[67] This more or less corresponds to our second level of the emergence of condensed matter. Ellis' examples are gasses, liquids and solids, and "at this level statistical physics applies, and entropy represents hidden information due to coarse graining."[68] He also insists that though "reduction is in principle possible", it fails in practice because of "the inability to derive by reductionist means the full complexity of behaviour of substances as simple as water", as we shall confirm in the next chapter.[69] The other more basically theoretical issues that preclude reductionism are due to the entropy principle (the arrow of time), "quantum measurement issues, and the divergences and incorrect predictions of the value of the cosmological constant [which] means that we do not properly understand quantum field theory…"[70] – all of which are unresolved problems in physics.

Ellis' second level is one that we have not noted, and that might be a serious omission in our account thus far if we aimed for completeness. It is where "bottom-up action plus boundary conditions leads to interesting higher-level structures not directly implied by boundary conditions."[71] Examples are the structural phenomena that arise purely out of a play of physical forces, such as snow crystals. Ellis gives a list of such material occurrences: sand piles, convection patterns, cellular automata, inorganic and organic molecules, and many more. However, he goes on to comment that "such structures are not truly complex, since they do not have the key element of goal-seeking that characterize living systems."[72]

Ellis' third level is that of living systems which correspond to our origin of life. This he believes is characterized not just by bottom-up but also by top-down action. This idea of top-down causation is a major feature of Ellis' approach in general both to life and mind, and it will constitute a major issue that we will debate in Chapter 5. According to Ellis, top-down causation in such "systems enables an element of teleonomy – and thus represents an effective

67 George F.R. Ellis, "On the Nature of Emergent Reality", in *The Re-Emergence of Emergence*, eds. Philip Clayton and Paul Davies, op. cit, 99.
68 Ibid, 99.
69 Ibid, 99.
70 Ibid, 99.
71 Ibid, 99.
72 Ibid, 99.

physical element of information."[73] However, such innate goals are inherent in the species and determined by evolution, and, therefore, no learning occurs.

Living systems, in turn, make possible the evolutionary emergence of animals with inherent control systems, such as nerve networks and brains. In such animals, unlike plants, "there exist feedback control systems directed by explicit goals related to memory, i.e. influenced by specific events in the individual's history."[74] Thus learning occurs "related to some form of consciousness leading to goal choice related to the remembered past."[75] Thus Ellis' fourth level corresponds exactly to our fourth emergence or the origin of animal mind.

The fifth level in both classifications is the highest mode of emergence, the quintessential form, as Hamlet puts it. This is where symbolic learning occurs and where language and culture play the crucial role in consciousness. It is where mere animal consciousness becomes human self-consciousness, as we shall see in Chapter 4. This is, of course, the human mind, the main subject that concerns us in this work.

In the very same volume in which George Ellis' article appeared, there is also one by Terrence Deacon which makes almost the very same classification of levels of emergence. Deacon, too, leaves out the emergence of matter as such and begins with the emergence of higher-order hydrodynamic properties such as those of water: "laminar flow, surface tension, viscosity…"[76] His second level is the emergence of order or self-organized behaviours out of instabilities, such as Bénard cells in a heated liquid. At the third level of emergence are living organisms which "involve some form of information-memory processes (as represented in nucleic acids)…"[77] The fourth and fifth levels also correspond closely to those of Ellis. Thus, it can be readily seen from both these authors and many others on whom we can draw that something like our five plus or minus one stage categorization is fairly standard.

The thinker who was perhaps the first to develop the tripartite order of matter, life and mind into a general philosophy was Maurice Merleau-Ponty in a little book entitled *The Structure of Behaviour* published as early as 1942.[78] He did so in conscious opposition to the Cartesian Dualism of thought and

73 Ibid, 100.

74 Ibid, 100.

75 Ibid, 100.

76 Terrence W. Deacon, "Emergence: The Hole at the Wheel's Hub", in *The Re-Emergence of Emergence*, eds. Philip Clayton and Paul Davies, op.cit, 126.

77 Ibid, 137.

78 Maurice Merleau-Ponty, *The Structure of Behaviour*, trans. Alden L. Fisher (London: Methuen, 1963).

extension or consciousness and matter, as well as its Sartrean derivative of the *pour soi* and the *en soi*. As against any such bipolarity, he asserts that "in reality, matter, life and mind must be understood as three orders of significations."[79] He develops each of these into the three dimensions of the "physical order, the vital order and the human order", and insists that each has its own irreducible reality. He is thus equally averse to any Materialist Monism, even such as advocated by the Gestaltists:

> But can the originality of biological and mental structures be really conserved, as Gestalt theory tries to do, while at the same time founding them on physical structures? A physical explanation of behaviour supposes that physical forms can possess all the properties of the biological and mental relations for which they serve as substrate… this amounts to saying that there is no difference between the three orders and that life and mind are different words for designating certain physical forms. Precisely if one is thinking in terms of structure, to say that physical forms in the final analysis account for human behaviour amounts to saying that the physical alone exist. Then consciousness *will* be what happens in the brain, and, as a matter of fact, we see Koffka defining consciousness, following the tradition of materialism, as that property "which certain events in nature have of revealing themselves".[80]

For Merleau-Ponty this amounts to a refutation not only of Gestaltist dreams of reducing mental forms to brain formations but also serves as a *reductio-ad-absurdum* of materialism in general.

Form is the key concept which he deploys in an overarching sense throughout his account of the three orders of the physical, vital and human. It is, of course, derived from Gestaltism but generalized much further beyond what Gestalt psychology assumed. It takes on an almost metaphysical dimension at the hands of Merleau-Ponty as each of the three orders is held to have its own types of forms which are hierarchically arranged such that the higher ones emerge from the lower ones and are based on them. Physical form is the most basic: it is "a physical system, that is, an ensemble of forces in a state of equilibrium or of constant change such that no law is formulable for each part taken separately and such that each vector is determined in size and direction by all

79 Ibid, 137.
80 Ibid, 136.

the others."[81] Form is thus akin to what is now called system, so that a kind of systems theory emerges out of Merleau-Ponty's work.

This is even more apparent when he turns from physical form to vital form distinguishing between the two in terms of their systemic structural properties. In physical forms "action which is exercised outside the system always has the effect of reducing a state of tension, of advancing the system towards rest."[82] It is otherwise with vital forms:

> We speak of vital structures, on the contrary, when equilibrium is obtained, not with respect to real and present conditions, but with respect to conditions which are only virtual and which the system itself brings into existence; when the structure, instead of procuring a release from the forces with which it is penetrated through the pressure of external ones, executes a work beyond its proper limits and constitutes a proper milieu for itself.[83]

The relevance of this to biological organisms is clearly evident.

Forms with respect to the human order or meaningful and mental formations are different again. Here Merleau-Ponty introduces the notion of symbolization or signification. Thus in perception what is perceived are not the qualities given to the senses or objects as entities but rather that which is of significance or has meaning for the human perceiver. In the first place, as with infants, these are maternal faces. From this he concludes as follows:

> The human signification is given before the alleged sensible signs. A face is a centre of human expression, the transparent envelope of the attitudes and desires of others, the place of manifestation, the barely material support for a multitude of intentions. This is why it seems impossible for us to treat a face or a body, even a dead body, as a thing.[84]

We can already see in this early work with its references to the body the lineaments of his later celebrated work on perception and mind in general.

Merleau-Ponty's work, cast from the start in the phenomenological tradition, has been a path breaking endeavour for phenomenologists ever since. However, his orientation towards science, as revealed in the tripartite division

81 Ibid, 137.
82 Ibid, 145.
83 Ibid. p. 145.
84 Ibid, 167.

of matter, life and mind, makes it more suitable for scientifically minded thinkers than Husserl's original phenomenology with its aversion to science as revealed in the exclusion of the naturalistic attitude through *epoché*. It is little wonder therefore that scientists such as Varela and philosophers such as Thompson have turned to Merleau-Ponty in their quest to naturalize phenomenology. In fact, Thompson takes great pains to render Merleau-Ponty's somewhat dated terminology into the terms of contemporary science, that is, into the concepts of much later theories than those current in Merleau-Ponty's own time. This might be taken as a somewhat anachronistic reading; however, to criticize it on that score is otiose and involves going into far more detailed exposition than we can allow ourselves here. It is much simpler and clearer to look directly at Thompson's own ideas themselves rather than seeing them refracted through Merleau-Ponty's dense language and at times obscure text.

A rendering of Merleau-Ponty's philosophy of matter, life and mind becomes the point of departure for Evan Thompson's book *Mind in Life* published as recently as 2007. Thompson uses this as the structural principle of his own work, though he mostly deals with life and mind and hardly at all with matter. Life and mind are both considered in terms of the "body", a concept that takes pride of place in his account of "mind in life". He bases himself on both Husserl's and Merleau-Ponty's phenomenology of the body, one that utilizes "a variety of terms for the spectrum of differentiated ways the body can be disclosed, from the objective physical body to a subjective life-flow, with various notions of the body as living or lived in between."[85] His ultimate aim is to combine biology and phenomenology around this vastly extended concept of "body", which we shall critically consider in Chapter 7.

Thompson believes he can eliminate the Cartesian mind-body problem or the so-called "hard problem" as expounded by David Chalmers – namely, that consciousness escapes all scientific accounts of how the mind functions and opens up an explanatory gap – with a much more tractable "body-body" problem where the gap is merely the distinction between "the lived body and the living body"[86] or *Leib* and *Körper* in Husserl's terms. He thinks that this can be accomplished by providing an "account of the lived body that integrates biology and phenomenology, and so goes beyond the 'gap'".[87] As he goes on to say:

> The scientific task is to understand how the organizational 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

85 Evan Thompson, *Mind in Life*, op. cit, 236.
86 Ibid, 237.
87 Ibid, 237.

> 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.[88]

Thompson's reference to "a special kind of autonomous system" invokes the biological theory of autopoiesis developed by Francisco Varela and Humberto Maturana. As we have seen, this is an extensive and far-reaching scientific approach to both the nature and origin of life that runs counter to all the other previously mentioned approaches, such as the divergent views of Kauffman featuring autocatalysis, Margulis featuring ecological systemic wholeness and Gould's historical contingencies in evolution. Varela seeks to combine this kind of scientific orientation with the established phenomenological school of Paris in order to bring about what he calls a naturalization of phenomenology that would go counter to and correct Husserl's bias against the naturalistic attitude and the natural sciences. A whole movement subsequently developed, one of international scope, based on this idea of naturalized phenomenology as an approach to the science and philosophy of mind. We shall consider it further in Chapter 7.

Thompson's book, which he began in collaboration with Varela before his death, reflects the orientation of this movement. It is structured along these lines for after an introductory first part, its second part deals with life based on the science of autopoiesis and the third part deals with mind based on a naturalized phenomenology. Almost throughout emergence features as a key concept, though it is not the main focus of Thompson's attention and is treated only in very general terms: "in complex systems theory, an emergent process is one that results from collective self-organization."[89] We shall consider in the next chapter whether complex systems theory can provide a comprehensive and adequate account of emergence.

Unlike Thompson, in this book we shall not be much concerned with life, either in its biological or phenomenological versions. Hence most of the issues he deals with will not be touched on or discussed. Nevertheless, there is considerable overlap and commonality between the two books. Biology and, in particular, the theory of autopoesis has much to teach us concerning emergence, though we must recognize that the emergence of mind is nothing like the emergence of life, a point that Thompson sometimes seems to overlook. Nevertheless, the sciences that deal with the emergence of mind are prepared for and explained by the sciences that deal with the emergence of life, as Morowitz's book bears out.

---

88 Ibid, 237.

89 Ibid, 60.

Thus, of the five levels of emergence we shall be concentrating almost exclusively on the last two, the others are only there to fill out the scientific picture, so to speak. The first level, that of matter itself, is a highly theoretical issue, both mathematically abstract and abstruse, which will be of no further concern to us in this work. The second and third levels, those of solid-state physics and biology, will play a minor supporting role in the next chapter, but hardly at all after that. Nevertheless, it is necessary to stress that what we have to say about the fourth and fifth levels of emergence, mind in animals and humans, is based on what has been said about the earlier levels. For emergence is raised to a higher power from one level to another through at least five stages, which can be symbolically represented as E^1, E^2, E^3, E^4 and E^5. This is an exponential nested hierarchy in which the lower levels are contained or encapsulated in the higher ones. How this takes place in respect of mind, how it is that human mind envelops and embraces animal mind, will be a key problem to be discussed in Chapter 3.

If what we have established about the continuum of matter and the five primary levels of emergence holds, Hamlet's dictum that Man is the quintessence of dust has been more or less vindicated by modern science. He was, thus, by contemporary standards, far more insightful about that than Descartes who, soon after Hamlet, declared that Man is a compendium or union of thought and extension, that is, mind and body. Unlike Descartes, who expresses himself in conventional religious terms, Hamlet is disgusted by his own vision of things, "man delights not me – nor woman neither", he hastens to add, to blunt the sexual or sexist point. The contemporary physicist, Steven Weinberg, takes a similar attitude to the facts of life and human existence in general. He writes that "the effort to understand the universe is one of the very few things that lifts human life a little above the level of farce, and gives it some of the grace of tragedy."[90]

Writing a tragedy like *Hamlet* or doing physics like Weinberg are certainly noble human endeavours. But whether only such supreme accomplishments lift us above the level of farce is to be doubted. For as a counterpoise to Hamlet's disgust and Weinberg's pessimism we might point to the cosmic and comic, though not farcical, perspective of a *Deus sive Natura* that forms Man out of mere dust and dirt, for with Man there comes the human mind that is not exclusive to prince or physicist but shared by all human beings alike. It is true, as the Bible teaches, that "dust thou art and to dust thou shalt return"; nevertheless, every person can use the mind to make some sense of life and find moments of gladness to be alive. As for the rest, the rest is silence, as Hamlet said.

90 Steven Weinberg, *The First Three Minutes*, op. cit, 155.

CHAPTER 2

On the Sciences of Emergence

1 Emergence and Reduction

We have arrived at mind through five stages of emergence: from elementary matter to condensed matter to living matter to animal mind and, finally, to human mind. Each of these stages is the subject matter of one or more sciences, starting with sub-atomic physics, going through solid-state physics, chemistry, biology, neurology and concluding with psychology, linguistics and sociology. All of these sciences rely on different forms of emergence, as we have already seen in our study of the different problems of origins. But is there or can there be one general theory of emergence to cover them all? This is the key question for this whole chapter. As we shall see, there have been many attempts to formulate such a general theory or science of emergence, usually presented as a systems theory or theory of complexity or theory of adaptive matter and many other such designations. We shall attempt to determine with what success such an aim has been pursued.

The re-emergence of emergence after its long hibernation has often been hailed as a new turning in the sciences, as an end of reductivism and the start of a counter current in the opposite direction, not from the complex to the simple but from the simple back to the complex. At least that is one way of presenting it. Another is to treat it as a paradigm revolution in the Kuhnian sense. Some, indeed, see it in still grander terms as another Scientific Revolution comparable to that which occurred in Descartes' day. Are such epochal changes in store for us now, or are we dealing with something far less momentous than that?

When *Hamlet* was written just prior to the Scientific Revolution the reigning natural philosophy, as it was then called, was a compendium of Aristotelian physics and Ptolemaic cosmology, and it was coming to an end. Perhaps its last gasp was the work of Robert Fludd which became popular not only in London but also in Paris in the 1620s and prompted Descartes, with the aid of Mersenne, to devise a mechanistic conception of matter in opposition because, they believed, it would be more in keeping with their Catholic faith; though, ironically, the Church preferred to stick to the old Aristotelian-Ptolemaic system for a long time, especially after it condemned Galileo's Copernican thesis in his book on the two worlds systems. Yet in the long run Galileo and Descartes prevailed and a new science was born. It proved to be an outstanding success. But its fate is now at stake.

There are thinkers, such as Stent and Horgan, to whom we have already referred, who hold that the great project of reduction furthered by this Cartesian science has now been accomplished and that there is nothing more for it to do. And since they identify what we can now call Classical science with science *per se*, they come to the conclusion that science is now at an end. This is not the view we reached in a previous work, *The Ends of Science*, where it was argued that it is the ends or goals of science that have drastically changed rather than that science as such was coming to an end. It is this far less apocalyptic conclusion about ends which we shall seek to vindicate in what follows.

In a way, these pessimistic conclusions about the end of science were already prefigured in the reductivist philosophy of science advocated by the great luminaries of twentieth century science and followed more or less unthinkingly by the great majority of scientists. Einstein, whom we quoted previously, declared that "the supreme test of the physicists is to arrive at those universal elementary laws from which the cosmos can be built up by pure deduction."[1] And once these universal elementary laws are known, that is the end of science, or so it seems. Much later he was echoed by Weinberg who said that "one of man's enduring hopes has been to find a few simple general laws that would explain why nature with all its seeming complexity is the way it is."[2] Weinberg speaks advisedly of "seeming complexities", as if these so-called "complexities" are neither insuperable nor intractable but only apparent, thereby implying that complexity is not any fundamental hindrance to reduction. Once the underlying laws are known, then the problems of applying these to real phenomena is a mere matter of "complexity", and complexity is always resolvable, if not in practice, then at least in principle. For Weinberg and reductionists like him the latter outcome is just as good as the former. Even if the complexity involved might never be overcome in practice, it is good enough that it should be possible to surmount it in principle.

It seems to such reductionist thinkers that if we can proceed from complexity to simplicity then we must also be able to reverse the process and go just as easily from simplicity back to complexity; that is to say, if we can reduce the complex phenomena of biology, chemistry and solid-state physics to the simplicity of the laws governing elementary particles, then we should be able to

1 Albert Einstein, *Ideas and Opinions*, op. cit, 225.

2 Steven Weinberg, "Newtonianism, Reductionism and the Art of Congressional Testimony", *Nature* 330 (1987), 433–7, in *Emergence: Contemporary Readings in the Philosophy of Science*, eds. Mark Bedau and Paul Humphreys (Cambridge/Mass.: MIT Press, 2008), 348.

start with these laws and derive all these complex phenomena; or, as Einstein intimated in our earlier quotation, that the cosmos can be built up by pure deduction. Thus, taking his cue from the reductionists, Morowitz seeks to build up the cosmos by assuming that emergence is merely the opposite course to reduction. On this view, if we can show that entity X is reducible to entity Y, then it follows that X emerges from Y, and consequently that X is deducible from Y. Hence, proceeding on this assumption, Morowitz starts with the fundamental entities of sub-atomic physics and step by step shows how higher entities emerge from these in a smooth and continuous process until he manages to reconstitute the whole cosmos following the upward course of emergence from the lowest to the highest entities. According to this view, there is nothing forbidding about emergence, it is simply the inverse of reduction, as he puts it:

> The reductivist approach leads us continually to seek solutions at lower and lower hierarchical levels. To move conceptually in the other direction, we must apply primary algorithms and seek for emergent properties or entities that become the agents for advancing another hierarchical level. Emergence is both a property of computer models and the systems being modelled. And so nature yields at every level novel structures and behaviours, selected from the huge domain of the possible by pruning, which extracts the actual from the possible. The pruning rules are the least understood aspect of this approach to emergence, and understanding them will be a major feature of the science of the future.[3]

Thus emergence is only a matter of "primary algorithms" and "pruning rules" and these will be arrived at through programs in computer models. It seems that a computer theory for the emergence of everything is what is to be aimed for in a science of the future.

Unfortunately, as we saw in the last chapter, it is not as easy as that to recreate the cosmos. This is so because emergence is not the complementary opposite inversion of reduction. Putting it in the most general terms, if entity X can be reduced to entity Y then it is true that X emerges from Y, but it does not follow that X can be deduced from Y. This can be so for many reasons. At its simplest this might be because the complexities involved are so great as to be intractable and therefore insuperable. The physicists Robert Laughlin and David Pines argue forcefully that there can be no such thing as deducing higher level emergent phenomena from a knowledge of the basic laws and initial conditions:

3 Morowitz, H.J., *The Emergence of Everything,* op. cit, 14.

> The emergent physical phenomena regulated by higher organizing principles have a property, namely, their insensitivity to microscopics, that is directly relevant to the broad question of what is knowable in the deepest sense of the term. The low-energy excitation spectrum of a conventional superconductor, for example, is completely generic and is characterized by a handful of parameters that may be determined experimentally but cannot, in general, be computed from first principles.[4]

Laughlin and Pines bewail the fact that emergent phenomena in physics tend to be treated with incredulity, as if this constituted some kind of "magic", like pulling a rabbit out of a hat or extracting something from nothing. It seems to be "fundamentally at odds with the reductionist beliefs central to much of physics…"[5] Thus, "the fact that the essential role played by higher organizing principles in determining emergent behaviour continues to be disavowed by so many physical scientists is a poignant comment on the nature of modern science."[6] As they go on to point out, the difference between reductivist and emergentist science is not merely a matter of a philosophical differences about the nature of science, but goes deep down to working practices as well:

> Living with emergence means, among other things, focusing on what experiment tells us about candidate scenarios for the way a given system might behave before attempting to explore the consequences of any specific model. This contrasts sharply with the imperative of reductionism, which requires us never to use experiment [in this way], as its objective is to construct a deductive path from the ultimate equations to the experiment without cheating. But this is unreasonable when the behaviour in question is emergent, for the higher organizing principles – the core physical ideas on which the model is based – would have to be deduced from the underlying equations, and this is, in general, impossible.[7]

Their crucial statement concerning the future of science is that "the Theory of Everything is not even remotely a theory of every thing."[8] What this means is that even when we know all the fundamental equations, this tells us nothing

4 Robert B. Laughlin and David Pines, "The Theory of Everything", in Mark Bedau and Paul Humphreys, eds., *Emergence: Contemporary Readings in Philosophy and Science* (Cambridge, Mass.: MIT Press, 2008), 261.

5 Ibid, 264.

6 Ibid, 264.

7 Ibid, 264.

8 Ibid, 260.

about what emerges out of them. The great physicist Richard Feynman put this very graphically when he said:

> The next great awakening of human intellect may well produce a method of understanding the qualitative context of equations. Today we cannot. Today we cannot see that the water-flow equations contain such things as the barber-pole structure of turbulence that one can see between rotating cylinders. Today we cannot see whether Schrödinger's equation contains frogs, musical composers, or morality – or whether it does not.[9]

Jack Cohen and Ian Stewart, who quote Feynman, demonstrate with numerous examples drawn from a wide variety of sciences that derivability of higher phenomena from lower level ones is neither necessary nor in most cases possible:

> Crystal lattices are not just phenomena that emerge from quantum mechanics. They have a universal aspect; they are phenomena that will emerge from any theory sufficiently close to quantum mechanics that involves identical roughly spherical atoms and energy minimization. This kind of universality is common to many, perhaps all, emergent phenomena; it is why they are able to emerge at all… But in order to exist at all, the phenomena that we observe on a human scale must be independent of the detailed organization at the lower level… Phenomena that are independent of detailed substructure are *universal* – "nature's theorems." They arise for general widely applicable reasons, not because of special details.[10]

Cohen and Stewart go on to argue that biological phenomena are of this emergent kind, and that they are independent of "special details". Thus biological phenomena that are observed at the macroscopic level of organisms are in that sense "independent" of the constituent factors at the molecular level of individual cells. This means that biological traits cannot be completely equated with or reduced to the chemistry of DNA molecules. They emerge from complex interactions between the genetic information contained in DNA and the epigenetic processes in the organism as a whole:

> What we are arguing is that there is no simple universal connection between genes and organisms. Instead there is a rich, fascinating and largely

9 Richard P. Feynman quoted in Jack Cohen and Ian Stewart, *The Collapse of Chaos,* op. cit, v.
10 Ibid, 236.

> unexplored joint dynamic… the versatility of genetic kits drives evolution in directions that make sense on the level of organisms…[11]

Their argument seems specifically directed against Richard Dawkins' reductionist thesis of the "selfish genes" which rules out organisms as relevant to the course of evolution. The organism is just the conveying vessel containing genes and has no other function. As against that, they contend that "natural selection does not operate on DNA sequences directly but through their organisms", and "there is no nice map from DNA to organisms, it makes no sense to talk of the phenotype determined by a given bit of DNA".[12] Thus biology or life as such cannot be reduced to molecular genetics:

> If the manner by which DNA code is transformed into creatures is ignored, we have no idea whatever of the possible complexity of the creature that results from a given segment of DNA. It is possible to change an organism entirely by changing only a few bits worth of DNA… On the other hand, changing twenty million DNA bases might do no more than alter the colour of an animal's ear tufts…[13]

Indeed, as they point out, animals with identical DNA might, nonetheless, develop quite differently depending on environmental conditions, as in the fish species Bonellia where males and females have the very same DNA.

The dispute between reductionists and emergentists is now turning into a battle royal which takes place on all levels. It is as much an issue in physics at all scales as in chemistry and biology. In this context, however, we are not interested in debating the matter beyond what is necessary to establish the concept of emergence as a viable approach in science. It will, of course, form the foundation for the account of mind developed in subsequent chapters. Nevertheless, it is worth pointing out how deep-seated the issue has become and how controversial, for it concerns all of science.

The fundamental dispute between the reductionists and the emergentists is not just about science as such, at one point it became directly political because issues of funding were involved. As we mentioned previously, the controversy raged before Congressional Senate sub-committees as it concerned the financial issue of whether to invest huge sums in order to build a Superconducting Super Collider that could generate the high energies necessary to test

11 Ibid, 318.
12 Ibid, 319.
13 Ibid, 354.

the theories that would come closer to the *Theory of Everything.* The main proponent for this proposal was the reductionist Steven Weinberg; his opponents were the emergentists Philip Anderson and James Krumhansl. What was at issue both financially and scientifically was a matter of priorities, whether the lion's share of government funding was to go to one project in elementary subatomic physics or was to be more widely distributed in solid-state physics and other sciences of complex phenomena. The latter won out, and the outcome was that the Super Collider was not built. Whether this really mattered to the physics of fundamental entities is difficult to establish, since even Weinberg concedes that the energies necessary to demonstrate the unification of all the fundamental forces "seem hopelessly beyond our reach":

> Beyond the questions about the standard model that we expect to be answered by the Super Collider, there is a level of deeper questions having to do with the unification of strong, electroweak, and gravitational interactions, questions that cannot be directly addressed by any accelerator now conceivable. The really fundamental Planck energy where all these questions could be explored experimentally is about a hundred trillion times higher than the energy that would be available at the Super Collider. It is the Planck energy where all the forces of nature are expected to become unified. Also, this is roughly the energy that according to modern string theories is needed to excite the first modes of vibrating strings, beyond the lowest modes that we observe as ordinary quarks and photons and the other particles of the standard model.[14]

It is facts of this nature that make is impossible for there to be any experimental proof for superstring theory and are behind the belief, held by Horgan, Lindley and others, that reductivist physics has reached its end as an empirical science and can only continue in a mathematical speculative mode.

Even though Weinberg is well aware of such problems, he nevertheless believes that fundamental research in physics should proceed for that provides the basis for all other types of explanation. Directly himself to his critics, Anderson and Krumhansl, he replies as follows:

> …even though new concepts "emerge" when we deal with fluids or many-body systems, we understand perfectly well that hydrodynamics and thermodynamics are what they are because of the principles of

14 Steven Weinberg, *Dreams of a Final Theory: The Search for the Fundamental Laws of Nature* (London: Random House, 1993), 187.

> microscopic physics. No one thinks that the phenomena of phase transition and chaos (to take the two examples quoted by Krumhansl) could have been understood on the basis of atomic physics without creative ideas, but does anyone doubt that real materials exhibit these phenomena because of the properties of the particles of which the materials are composed?[15]

The answer to Weinberg's rhetorical question is that previously provided by Cohen and Stewart, it is that "the phenomena that we observe on a human scale must be independent of the detailed observation on the lower level… independent of the detailed substructure," and certainly so of the lowest levels of all. Hence investigating the lowest levels of matter is of no relevance to hydrodynamics or chaos, which manifest themselves on the human scale.

This is also the general reason why mental and behavioural phenomena are not reducible to the detailed processes of neurology. They are independent of and indifferent to the precise details of neural connectivity, though, of course, they could not exist without them. What goes on in the brain in fine grained detail is so complex as to be beyond over-all observation; we can only observe and follow single neural connections and firing details, not the whole of the brain processes even at any one moment since it is so huge. The brain has 100 billion neurons and each of these can be connected to upward of a thousand others, hence, the number of possible connections is more than just astronomical, it is super-astronomical for the mathematics is just mind boggling. Which of these are in play at any one moment in any one brain is beyond reckoning. And what goes on in any individual brain when something is experienced is almost certainly utterly different in detail from what goes on in any other brain having the same experience. Hence, any idea of reducing experiences to neural events must be abandoned for reasons of variability alone, if not for other reasons to do with emergent phenomena, which is what experiences are, as we shall argue in the following chapters.

This is the kind of argument that the renowned neuroscientist Gerald Edelman presents against the reduction of mental phenomena to neural events. He asks the rhetorical question: "what is the detailed evidence that there is a large amount of variation in the nervous system from individual to individual, and at what levels does it occur?"[16] And he proceeds to answer the latter part in

15 Steven Weinberg, "Newtonian Reductionism and the Art of Congressional Testimony," Mark A. Bedau and Paul Humphreys, *Emergence*, op. cit, 350.

16 Gerald M. Edelman, *Neural Darwinism: The Theory of Neuronal Group Selection*, (New York: Basic Books, 1987), 57.

great detail. He provides a table setting out eight "sites and levels of neuronal variation" at various scales. We shall merely list his general categories of sites of variation: in genetic traits and developmental primary processes, in cell morphology, in connection patterns, in cytoarchitectonics, in transmitters, in dynamic response, in neural transport and in interaction with glia. In a subsequent work he offers this general conclusion:

> Individual nervous systems (particularly those of vertebrate species) show enormous variability. This important characteristic was explicitly noted by Lashley, who with rare restraint offered no explanation for it. Variability occurs in both space and time at many levels: molecular, cellular, anatomical, physiological and behavioural. Despite the commonality of neural structures within a species, the degree of variability in each individual far exceeds that which could be tolerated for reliable performance in any man-made machine. Yet characteristic species behaviour can be described, and this behaviour is obviously adaptive or current species would not have survived.[17]

Edelman deploys this argument specifically against the computer model of the brain, but it also works against other forms of reduction as well, as we shall see when we return to Edelman in Chapter 7.

The complexities involved that make reduction impossible are not just a contingent, incidental matter but go to the very heart of such problems, and can never be discounted for "in principle possible" reasons. Mathematically or logically considered everything that is numerical is in principle possible, but that does not mean that it makes any sense at a human level. The invention of computers has given the illusion that complexities are always resolvable if sufficient computer power is available. But for some problems, the really interesting ones, sufficient computer capacity will never be available. This illusion was prevalent in the early days of Artificial Intelligence when it seemed that everything that human beings can do can be simulated by computers. But then it was found that certain types of problem rapidly lead to the computational catastrophe of a combinatorial explosion of possibilities. Another such is the so-called "frame problem", which now by general consent is considered insuperable, as we shall see in Chapter 7.

This is also the reason why deduction of emergent phenomena from fundamental principles can never be carried out. For even if, *per impossibile*, we did

17 Gerald M. Edelman, *The Remembered Present: A Biological Theory of Consciousness* (New York: Basic Books, 1989), 40.

have the computing power necessary, we could not start the process for lack of sufficiently detailed knowledge of the initial conditions from which the deduction is to take place. Thus, for reasons of so-called sensitivity to initial conditions, we could never deduce the weather even if we knew all the laws of how the innumerable factors that can influence the weather operated. The so-called butterfly effect of Chaos Theory makes it impossible to know the initial conditions in sufficient detail to be able to predict deductively how the weather will develop. Of course, for limited purposes, given sufficient observations, the weather can be predicted from day to day with reasonable certainty. So weather forecasting is an exact science, but not a deductive one.

Hence, to argue that emergence cannot be deductively explained is not to maintain that science fails and that no scientific explanations at all are possible in these matters. The failures of reductivism must not be taken as the break-down of science. As we showed previously, the mysteries of origin, which are the prime problems of emergence, can be illuminated by science, they are not impenetrable. However, this is not the kind of science that the reductionists expect all science to be. It is not the science that predominated during the previous era of Classical science when the great reductive achievements were accomplished. Nor does it have the kind of historical progression to which Kuhn's model of paradigm revolution followed by normal science applies. Dealing with origins calls for a very different kind of scientific approach, as we saw in the previous chapter.

Indeed, as we shall presently proceed to show, the Kuhnian model does not even apply to all sciences during the Classical era. For example, the crucial science of hydrodynamics had a quite different non-Kuhnian history, as revealed in the detailed scholarly account by Olivier Darrigol to which we shall return at the end of this chapter.[18] And the same is true of many other sciences. In those sciences which did experience drastic changes – such as physics at the hands of Newton, thermodynamics beginning with Clausius and electrodynamics through Maxwell – the changes that subsequently occurred were at least as important. It did not happen all in one fell swoop by paradigm revolutions, but through a number of successive transformations, gradually in an incremental fashion, as Darrigol argues.[19]

There is no doubt that the great reductive achievements and the major theories they generated covering large fields were of crucial importance for the progress of science. Without that basic knowledge science could scarcely

18 Olivier Darrigol, *Worlds of Flow: A History of Hydrodynamics from the Bernoullis to Prandtl*, (Oxford: Oxford University Press, 2005).

19 Ibid, 323.

proceed. However, it is now impossible for science to proceed in the same way or for thinkers about science to assume that since science can no longer advance in that way it must come to an end, or continue only with the humdrum task of applying the known laws and theories to new phenomena. The latter is what Victor Weisskopf contends when he states that "once new and fundamental laws are discovered, a large and ever increasing activity is to apply the discoveries to hitherto unexplained phenomena."[20] Anderson, who quotes this, supplies the necessary corrective:

> The main fallacy of this kind of thinking is that the reduction hypothesis does not by any means imply a "constructionist" one. The ability to reduce everything to simple fundamental laws does not imply that ability to start from those laws and reconstruct the universe.[21]

This is the gist of our previous argument that emergence is not simply the upward inverse of downward reduction. Emergence is a concept that Anderson, unfortunately, does not avail himself of, though his main argument is in that spirit. And, indeed as he goes on to show, a variety of different sciences have arisen that might collectively be called the sciences of emergence which proceed in a fundamentally different way from the sciences of reduction.

In the sciences of emergence we can no longer always expect to find a single unequivocal solution to every problem, the so-called "true" answer. We can no longer rely on a commonly accepted Kuhnian paradigm which provides the model for on-going research. This does not mean that research does not proceed in the rigorous and controlled scientific way that it has always done, only that it is no longer designed to test or falsify a single theory. Rather, for every major emergent phenomenon there are always multiple possible theoretical explanations each of which stakes an equal claim to be the correct one, at least in the initial stage. What ensues is a controversy or dialectic such as is usually associated with philosophy. And, indeed, philosophers invariably enter the fray and there is no way to keep them out for this kind of science invariably calls for the help of some other not strictly empirical criterion. As we have already seen, this is particularly the case in the sciences dealing with the three fundamental problems of emergence in the sciences of the origination of matter, life and mind.

---

20 Victor Weiskopf, quoted in P.W. Anderson, "More is Different: Broken Symmetry in Nature and the Hierarchical Structure of Science", *Science*, 177 (1972), 393–6; in Mark A. Bedau and Paul Humphreys, *Emergence*, op. cit, 222.

21 Ibid.

Such an intrusion of philosophy into science will no doubt come as a surprise or even nasty shock to scientists and philosophers brought up in the tradition of Classical science with its emphasis on experimental or at least empirical validation alone, science from which all metaphysical assumptions or other such so-called value biases are supposedly strictly excluded. This has been the dictate of a so-called Scientific Method that Positivists have pursued in vain. They will argue that it has taken centuries for science to free itself from the burden of philosophy, at least from Newton's definitive pronouncement "*hypothesis non fingo*", meaning that he refused to indulge in philosophical speculation about underlying causes, that he was merely concerned with observable phenomena and the laws governing them. The Positivists took their cue from that.

As a result, at least since Newton there has been an ever growing separation between philosophy and science, which was described in the opening chapter of *The Ends of Philosophy*.[22] Newtonian physics was the first major science to divorce itself from philosophy at the end of the 17th century; it was followed by chemistry with Lavoisier and Dalton at the end of the 18th century; and eventually also biology gradually freed itself throughout the 19th century and finally in the mid-20th century when vitalism or the *élan vitale* was finally laid to rest by Crick and Watson's presumed reduction of life to DNA chemistry. The other sciences have tried to follow suit in establishing themselves on a purely empiricist reductive basis, in accordance with Positivist dicta about Scientific Method.

But now with the emergence of emergence in the very heart of science, philosophy has once more a role to play in science. This was already intimated in physics in the various philosophically diverse interpretations of the equations of quantum mechanics; but for working physicists this did not really matter as long as they could rely on the equations themselves, and did not have to concern themselves about how they were interpreted. But with emergent phenomena it does matter which philosophical version is adopted for that determines the kind of explanation one will offer, as we saw in the previous chapter concerning fundamental origins. And in general, the scientific approach to matter in its various emergent forms can no longer be reductively mechanistic as it was during the era of Classical science, and neither can it be dictated by Scientific Method.

The new approach to matter in the contemporary sciences is well captured by the physicists Robert Laughlin and David Pines:

22 Harry Redner, *The Ends of Philosophy*, op. cit.

> The central task of theoretical physics in our time is no longer to write down the ultimate equations but rather to catalogue and understand emergent behaviour in its many guises, including potentially life itself. We shall call this physics of the next century the study of complex adaptive matter.[23]

But matter comes in multiple forms and it is not possible to devise a single formula or equation that covers every one of them. The main burden of their argument is that a "Theory of Everything is not even remotely a theory of every thing."[24] Such a theory can only be utilized when a small number of elements are involved, "it cannot be solved accurately when the number of particles exceeds about 10."[25] Any assembly larger than that is impossible to calculate, and "no computer existing, or that will ever exist, can break this barrier because it is a catastrophe of dimensions."[26] Hence, as they conclude that "there are many notorious failures of alleged *ab initio* computational methods" in physics and that expecting them to work in biology or in a higher science is "patently absurd."[27] Once more we arrive at the basic insight that reduction cannot be reversed by deduction, we cannot derive something from that which it can be reduced to. This is particularly so in the emergent properties and behaviour of matter for that involves "higher organizing principles" and the deduction of such principles "from the underlying equations is, in general, impossible."[28]

If this is so, then emergence cannot be explained away or eliminated by reference to reduction. The fact that something can be reductively explained tells us very little about how it emerges or how it behaves as an emergent phenomenon. To take a simple illustration, the fact that the composite substance water is H_2O, that is, it can be chemically reduced to atoms of hydrogen and oxygen, does not explain any of the macroscopic behaviour of water, how it runs down a plug hole or why the hydrodynamic laws apply, as Anderson and others have so cogently argued.

But that still leaves open the question of whether the various forms of emergence can themselves be unified under the one general law or model or whether each is, in fact, *sui generis*. The attempt to develop a general theory of

---

23 Robert B. Laughlin and David Pines, "The Theory of Everything", Proceedings of the National Academy of Science 97 (2000), 28–31, in Mark A. Bedau and Paul Humphreys, *Emergence*, op. cit, 265.

24 Ibid, 260.

25 Ibid, 260.

26 Ibid, 260.

27 Ibid, 260.

28 Ibid, 260.

emergence has been going on almost throughout the twentieth century. A number of movements designed to account for emergent phenomena have succeeded each other. Perhaps the first was General Systems Theory, so we will begin with that first then proceed to the others.

2 In Search of a General Theory of Emergence

General Systems Theory was founded by Ludwig von Bertalanffy as early as 1928 in a book by that name, but it was not published until 1968. The origins of his approach go back to the work of Alexander Bogdanov which was published in Russia in the years from 1912 to 1917. It was in America after the Second World War that von Bertalanffy gained his major following, attracting such key supporters as Kenneth Boulding, Anatol Rapoport and Ross Ashby. What they sought to show was that similar systems featured across all fields of science. Hence they sought to develop a set of general concepts which would indifferently apply to all the sciences. Among these are such now well-known terms as open and closed systems, boundaries, homeostasis, adaptation and equifinality. The General Systems Theory movement is still current but has been largely overtaken by many others which can be seen as its later offshoots.

Among the most widespread and influential of these was the Cybernetics and Information Theory movement. The former owes its inception and name largely to the work of Norbert Wiener and the latter to Claude Shannon, but both were from the start closely integrated and both emerged from the MIT school which, as we shall see, was the institutional setting for so many American scientific and philosophical developments in the second half of the twentieth century. This conjoint movement arose out of the problems encountered in war-time developments of automatic aiming and firing weapons and electronic communication. However, it soon went way beyond technology and was touted as a general theory of all systems in machines, organisms and organizations. According to John Horgan, Wiener "proclaimed that it should be possible to create a single, overarching theory that could explain the operation not only of machines but also of all biological phenomena, from single-celled organisms up through the economies of nation-states."[29] Wiener and his fellows in the movement believed that such an all-embracing general theory of all systems was possible because "all these entities process and act on information,

29 John Horgan, *The End of Science*, op. cit, 207.

they all employ such mechanisms as positive and negative feedback and filters to distinguish signals from noise."[30]

The Cybernetic-Information movement took off rapidly after the war and soon gained traction throughout all the sciences in America. In large part this was due to a whole series of annual meetings sponsored by the Macy foundation called the Macy Conferences which were attended only by special invitation and involved a whole bevy of outstanding scientists drawn from all the major scientific disciplines.[31] It was at the 1946 conference that Wiener first outlined the new general systems theory model. The key to his proposal was that information was a physical measure which he called negative entropy, thus combining thermodynamics and the information theory, as he put it:

> In fact … it is not surprising that entropy and information are negatives of one another. Information measures order and entropy measures disorder. It is indeed possible to conceive all order in terms of message.[32]

As Jean-Pierre Dupuy comments, "what he was aiming at was a physics of information. He saw the concept of information as the means by which biological and even social questions could be reduced to problems of physics."[33] Much against the intention of Shannon's original theory, he took the idea of "information" way beyond signals transmissions and turned it into a physical measure totally independent of coding, interpretation or meaning and not peculiar to humanly devised systems. It was a momentous move that soon reverberated within all the sciences. Information was present everywhere in nature since it was synonymous with order and organization. It is prevalent in biology as the coupling of organism and outer environment is a matter of the reception of messages which determine how it would behave. This view aroused strong controversy at the Macy conferences, as Shannon insisted that information, as he defined it, was totally distinct from the meaning of a message. What was at issue was to ascertain whether there was any objective way of determining which part of a message was signal and which was noise, and in fact, no scientifically objective way of separating the two has been discovered. As we shall see, this dispute has continued till our own time. In Chapter 7 we shall present at greater length the critique of the relevance of Information Theory or

30 Ibid, 207.

31 See Steve Joshua Heims, *The Cybernetic Group* (Cambridge, Mass.: MIT Press, 1991).

32 Quoted in Jean-Pierre Dupuy, *The Mechanization of the Mind: The Origins of Cognitive Science*, trans. M.B. De Bevoise (Princeton: Princeton University Press, 1999), 113.

33 Ibid, 113.

Cybernetics to biology or any human science, as strongly argued by Gerald Edelman.

What made the Wiener view of information as a physical measure so attractive in the immediate post-war years was that it was utilized metaphorically to interpret the spectacular discoveries by Watson and Crick of what seemed like a genetic "code" that determined heredity. It was very quickly taken in a literal way and the language of Information Theory – terms such as code, program, transcription, translation and above all message – was projected into the molecular structure of the cell. After that there was no stopping the so-called Information Revolution, and even though many of the original concepts of Cybernetics fell away, those of information remained. Based on the idea of information processing, a new Cognitivist movement arose in psychology. There were cognate developments in economics and the social sciences aided by the formalized theories of games launched by John von Neumann and Oscar Morgenstern, and the decision theory of Leonard Savage, all associated with the Macy Conference circle. And behind it all there were the developments in computing, also presented in these terms by von Neumann and others.

Not surprisingly there was strong opposition from specialists in all these fields. In biology, Paul Weiss "opposed the reductionism and determinism of the theory of genetic information while resisting the challenge to biology posed by cybernetics as well".[34] In opposition to the terms of molecular biology, he proposed the idea of self-organization, particularly for the specialist science of embryology. It was to feature prominently in all kinds of other contexts later, as we shall see. He was not alone in being averse to the generalizing assumptions of Information and Cybernetics. Cohen and Stewart have pointed out that the idea of genetic information is only a metaphor that does not hold up very well and does not explain very much:

> Strictly speaking, the genetic code isn't even a code. Codes must be written as well as decoded. It is true that DNA codes for – that is, determines – proteins, but there is no converse process encoding proteins into DNA. The DNA message is not transmitted but copied… the metaphor of DNA as a message transmitted by the parent and received by the offspring does not hold up under scrutiny. When the message is transmitted there is no receiver. The message, indeed, is supposed to describe how to construct a receiver![35]

34 Ibid, 131.

35 Jack Cohen and Ian Stewart, *The Collapse of Chaos*, op. cit, 290.

They are adamant in insisting that the DNA molecule does not contain all the information necessary to constitute an organism, since the base sequence of DNA only comes into effect in the course of biological development which permits its expression. Hence, there is no simple and direct correlation between the genome and the phenome.

Cybernetics had a long lease of life even after the passing of its founders. One of the younger participants at the Macy Conferences, Heinz von Foerster, went on to establish what has come to be called a Second Cybernetic. He tutored many followers including the biochemist Henri Atlan and the philosopher Jean-Pierre Dupuy. According to Wolfgang Krohn, Günter Küppers and Helga Nowotny, von Foerster "elaborated a theory of self-organization systems by drawing on the findings of biological systems theory, cybernetics, and automation theory".[36] As they further explain:

> ...he develops, following Shannon's Information theory, a measure of order by formulating the conditions of its increase: a system can achieve a state of higher order if and only if the relative increase of the maximally possible disorder within the system is greater than the present prevailing actual disorder, since under such conditions the relative order increases automatically. This may occur in two ways that are independent of each other and cooperate in real systems. In the first process, environmental perturbations (undirected energy) are selected by the system to be subsequently incorporated into the structure of the system. In the second case, the system expands. This expansion does not occur arbitrarily but is due to an import of "building blocks" that are compatible with the systems structure. Heinz von Foerster called the first of these principles "order from noise", while the second, according to Erwin Schrödinger, is "order from order".[37]

On this view, self-organization develops out of the interaction of order and disorder or redundancy and variety. This approach is distantly related to Prigogine's work on systems far from equilibrium or so-called dissipative systems, and at a further remove to the Maturana and Varela conception of autopoiesis. According to Dupuy, "there was a strong conviction that the entropy of

36 Wolfgang Krohn, Günter Küppers, Helga Nowotny, eds. *Self-Organization – Portrait of a Scientific Revolution, Sociology of the Sciences Year Book,* vol. xiv, (Dordrecht: Kluwer, 1990).

37 Ibid, 2.

chemical processes and the complexity of living things were connected, and that success in elucidating this connection did not in the least depend on applying categories borrowed from mechanistic engineering [i.e. cybernetics] or information theory."[38] However, according to Prigogine's collaborator and coauthor, Isabel Stenger, von Foerster's teaching and that of his followers was far too general for what Prigogine sought to convey:

> In Stenger's view, the kindship between the embryologists' theory of self-organization [e.g. Weiss] and that of the Brussels school of thermodynamics [Prigogine] did not extend to the Chilean school of Maturana and Varela nor Atlan and his circle in France, the latter two groups, she argues, remaining hostages to the mechanistic schemes of second-order cybernetics.[39]

Clearly what is at issue is whether a generalist approach such as that of the Second Cybernetic can encompass the more specialized work in discrete fields such as that of embryology or complex chemical or biological systems in general.

After Cybernetics and Information Theory, after the Second Cybernetic, the next great wave of generalist theories on systems, self-organization and emergence arose out of computer generated work. There have been two closely related branches to this whole endeavour that go under the separate names of Chaos Theory and Complexity Theory, though they are so close to each other that Horgan has coined the term "chaoplexity" to cover both. But whether in fact such a general all-embracing science exists or can ever exist is put in doubt when Horgan goes on to point out that the terms "chaos" and "complexity" have been "defined in so many overlapping ways by so many different scientist and journalists that the terms have become virtually synonymous if not meaningless."[40]

It all began with the almost experimental discovery that working with computers on simple recursive equations, repeating them countless times over, produces highly unexpected and frequently counter-intuitive results. One could almost call this kind of work experimental mathematics, for one allows a computer to proceed on simple mathematical instructions and then observes what results from the procedures over a period of time, frequently something strange and wonderful, revealing an order that no one had ever suspected

38 Jean-Pierre Dupuy, *The Mechanization of Mind*, op. cit, 130.
39 Ibid, 184.
40 John Horgan, *The End of Science*, 190.

before. This is how Benoit Mandelbrot discovered his fractal sets with their amazing self-similarity on different scales. This is also how Edward Lorentz hit on the butterfly effect or deterministic chaos. And thus Chaos Theory was born.

Deterministic chaos goes back historically to the work of Henri Poincaré on the three-body problem: the gravitational attraction of two bodies can be exactly calculated using Newton's equation, but the gravitational effect of three bodies acting on each other is no longer calculable using such linear system equations. It constitutes a non-linear system that cannot be dealt with mathematically in this linear way, but it can be computed using high-speed computers. Poincaré did not have computers at hand, so such problems were considered unsolvable. Computers opened up the possibility of finding solutions to equations of non-linear systems.

The difference between linear and non-linear systems lies in the kind of physical behaviours they describe. The former govern smooth and continuous changes, such as the laminar flow of a liquid, the latter govern chaotic flow or turbulence. Paul Humphreys lists a number of crucial differences: "in linear systems, a small change in parameters or small external perturbation will lead to small changes in behaviour. For nonlinear systems, such small changes can lead to very large qualitative changes in motion."[41] The latter is now known as sensitivity to initial conditions and it is what explains the butterfly effect. Another difference is that "linear systems exhibit dispersion, as in the decay of water waves moving away from a central source, whereas the stability of eddies in turbulent flow exhibits nondispersion. This can be associated with pattern emergence…"[42] The last point is crucial for it seems to open the way to a study of emergent order arising out of chaos. As Humphreys goes on:

> In recent years much attention has been paid to relations among nonlinear systems, complexity theory and emergence; and it is often claimed that nonlinear systems form a subset of the set of complex systems, and that complex systems give rise to emergent phenomena.[43]

This suggests that there can be a general theory of emergence. Is such a theory in fact possible? There are many undoubtedly distinguished scientists who are working towards such a theory. They confidently expect a Newton or Darwin to arise who will do for emergence what these superior, all-encompassing minds

41 Paul Humphreys, *Emergence: a Philosophical Account*, (New York: Oxford University Press, 2016), 260.

42 Ibid, 260.

43 Ibid, 259.

did for gravity or evolution. But there are also equally distinguished scientists who contend that this is an illusory quest for no such general and comprehensive theory is possible. Every major case of emergence is sui generis; and though there are analogies and likenesses that can be drawn, it is not possible to devise one overarching formulation to comprehend them all. Each must be studied and explained in its own way. This difference of opinion and outlook is one of the crucial parting of the ways in contemporary science.

Those who are dedicated to finding a general theory of emergence, such as the denizens of the Santa Fe Institute, are undeterred by any of the criticisms thus far, which they see as mere carping, or by the difficulty of the task itself. They conceive such a theory as a general theory of complexity or what Holland has called "a unified theory of complex adaptive systems." In a speech to the Santa Fe Institute he elaborated what such a theory would embrace:

> In the natural world such systems include brains, immune systems, ecologies, cells, developing embryos, and ant colonies. In the human world they include cultural and social systems, such as political parties and scientific communities. Once you learned to recognize them, in fact, these systems are everywhere. But wherever you found them, they all seemed to share crucial properties.[44]

He then went on to expound what he held these crucial properties to be:

> First, each of these systems is a network of many "agents" acting in parallel. In the brain the agents are nerve cells, in a cell the agents are organelles, such as the nucleus and the mitochondria, in an embryo the agents are cells, and so on. In an economy, the agents might be individuals or households.
>
> Second, a complex adaptive system has many levels of organization, with agents at any one level serving as building blocks for agents at a higher level.
>
> Third, all complex adaptive systems anticipate the future… From bacteria on, every living creature has a prediction encoded in its genes: "In such and such an environment, the organism specified by this genetic blue-print is likely to do well."

44 Quoted in W. Mitchell Waldrop, *Complexity: The Emerging Science at the Edge of Order and Chaos* (New York: Viking, 1992), 82.

> Finally, complex adaptive systems typically have many niches, each one of which can be exploited by an agent adapted to fill that niche.[45]

This ambitious program was enthusiastically received by most of those engaged in this kind of work at the Santa Fe Institute.

Holland is not chary in pointing out that this quest is not only a theoretical issue but one with huge practical implications:

> Many of our most troubling long-range problems – trade imbalances, sustainability, AIDS, genetic defects, mental health, viruses – centre on certain systems of extraordinary complexity. The systems that host those problems – economies, ecologies, immunity systems, embryos, nervous systems, computer networks – appear to be as diverse as the problems. Despite appearances, however, the systems do share significant characteristics, so much so that we group them under a single classification at the Santa Fe Institute, calling them complex adaptive systems (CAS). This is more than terminology. It signals our intuition that there are general principles that govern all CAS behaviour, principles that point to ways of solving the attendant problems.[46]

Much was promised at the start; but since then very little has been delivered, despite the tremendous amount of work on computer simulation.

To encompass the ends set out by Holland and others, algorithms and other computer based techniques have been developed. Holland has invented genetic algorithms whereby new programs can spontaneously arise out of stretches of computer code shuffling and rearranging themselves in order to meet the requirements of a given problem. This is supposed to mimic artificially the natural way in which the genome in living entities undergoes spontaneous change as the organisms evolve through natural selection. This is what Holland means in saying that complex systems can be adaptive.

Christopher Langton has developed programming techniques for the computer modelling of biological complexity in the firm belief that this is the route that would lead to the holy grail of artificial life. Langton sets out this approach in the following terms:

> Its credo is that life is not a property of matter per se, but the organization of matter. Its operating principle is that the laws of life must be laws

45 Ibid, 145–6.

46 Ibid.

> of dynamical form, independent of details of a particular carbon-based chemistry that happened to arise here on Earth four billion years ago. Its promise is that by exploring other possible biologies in a new medium – computers and perhaps robots – artificial life researchers can achieve what space scientists have achieved by sending probes to other planets: a new understanding of our own world through a cosmic perspective on what happened in other worlds. Only when we are able to view *life-as-we-know-it* in the context of *life-as-it-could-be* will we really understand the nature of the beast.[47]

Langton holds that such computer models constitute virtual forms of life, or perhaps even mind for that matter, that are in principle no different from the, so to speak, natural counterparts. Kauffman is perhaps not quite as *outré* as that, but he, too, believes that mathematical models, such as those that computers generate, point the way to reality, to the way in which emergent phenomena arise. As Horgan remarks, "Kauffman seems to see all phenomena, from bacteria to galaxies, as manifestations of abstract mathematical forms that undergo endless permutations."[48]

Little wonder that specialist scientists are sceptical of such generalist ambitions couched in mathematical computer language. Anderson has expressed doubt as to whether computer models can explain real natural phenomena. He points out that the fact that a model matches a phenomenon is no proof that it is its explanation because another model might do so equally well. Drawing on such criticism, Horgan comes to the conclusion that "so far chaoplexologists have created some potent metaphors: the butterfly effect, fractals, artificial life, the edge of chaos, self-organizing criticality. But they have not told us anything about the world that is both concrete and truly surprising either in the negative or positive sense".[49]

It is perhaps too early to call it a day on the attempt at a general theory of emergence by means of systems theory or complexity theory or complex adaptive systems or any other such mathematical or computer method. As Horgan notes, "the field of chaos, complexity and artificial life will continue. Certain practitioners will be content to play in the realm of pure mathematics and

47 Ibid, 277.
48 John Horgan, *The End of Science,* op. cit, 136.
49 Ibid, 225.

theoretical computer science."[50] However, he hastens to add, "they will not achieve any great insights into nature."[51]

Even amongst those at the Santa Fe Institute there were critics and sceptics of such grandiose computer generated projects. Murray Gell-Mann doubted the possibility of the one theory being able to cover all the various diverse phenomena of complexity: "there are huge differences among these systems, based on silicon, based on protoplasm, and so on. It's not the same."[52] Gell-Mann argued, in opposition to Holland and many of the others, that complexity per se is not the same as adaptation, nor is it necessarily order creating:

> Turbulent flow in a liquid is a complex system. But it can't be called adaptive. In turbulent flow there are eddies that give rise to smaller eddies and so on, and certainly eddies have properties that allow them to survive in the flow and have offspring, while others die out. There's information in the system, no question. But it does not produce a schema, a compression of information with which it can predict the environment.[53]

According to Gell-Mann, "biological phenomena stem from so many random, historical, contingent circumstances that it is not possible to account for them all by the one general law of emergence."[54] Hence, no one such law or theory can explain why the universe has generated so much order of different kinds despite the Second Law of thermodynamics that mandates a drift towards disorder, that is, entropy.

Gell-Mann is not the only critic at the Santa Fe Institute, Anderson, usually his opponent, is another. He holds that at each stage of emergence "new laws, concepts, and generalizations are necessary" such that no one general theory can capture all these various levels at once, for each is independent to some extend of that below it and that above it.[55] Nor does he believe that computer models can capture all these various emergent systems: "I know that they don't work! I always wonder whether global climate models or oceanic circulation models or things like that are full of phony statistics and phony measurements";[56] and even if they did work, they would tell you nothing

50 Ibid, 225.
51 Ibid, 225.
52 Ibid, 225.
53 Quoted in Roger Lewin, *Complexity: Life at the Edge of Chaos*, (Chicago: University of Chicago Press, 1992), 15.
54 Quoted in John Horgan, *The End of Science*, op. cit, 213.
55 Philip Anderson, "More is Different", op. cit, 222.
56 Quoted in John Horgan, *The End of Science*, op. cit, 210.

about what is really going on in the system: "at some point, the computer is not telling you what the system is doing."[57]

These are merely the insider critics, outsider ones have been still harsher. Biologists have attacked Kaufmann's theories on emergence of order out of complex systems per se, his claim that "self-organization is a natural property of complex genetic systems. There is 'order for free' out there, a spontaneous crystallization of order out of complex systems, with no need for selection or any other external forces."[58] Thus Robert May, an Oxford ecologist, stated in typical British put-down terms, that what the Santa Fe Institute scientists do is "mathematically interesting but biologically trivial".[59] Kauffman's work with Boolean networks, such as his NK networks to model genetic on-off activation, is said to "provide yet another example of a mathematician ignoring real time."[60] But perhaps even more critical of these computer models is Stanley Miller, who called them "paper chemistry". He was specifically referring himself to Kauffman's theory of autocatalytic sets as an explanation of the origin of life, which he criticized as "running equations through a computer does not constitute an experiment."[61] John Maynard Smith has also been a stern critic of what Kauffman envisages in the name of a general theory.

But no critic has been more scathing of this whole brave new world of NK networks, CTRNs (Continuous-Time Recurrent Networks), cellular automata, spin glasses, genetic algorithms, chaos theory, dynamical systems, and all the rest of what has been held out as the promise of a new scientific age than Horgan. He states his objections as follows:

> For such hypotheses to be meaningful, the proponents must tell us what, exactly, complexity is and how it can be measured… Unless this problem is solved, all these hypotheses about laws of complexity or complexity-generating forces are meaningless. I doubt (surprise, surprise) that the problem can be solved.[62]

He goes on to list around 45 different definitions of complexity propounded by 45 distinguishes scientists. He quotes Gell-Mann who states that "any definition

57 Ibid, 210.
58 Quoted in Roger Lewin, *Complexity*, op. cit, 24.
59 Ibid, 184.
60 W. Mitchell Waldrop, *Complexity: The Emerging Science at the Edge of Order and Chaos* (New York: Viking, 1992), 312.
61 Quoted in John Horgan, *The End of Science*, 139.
62 Ibid, 277.

of complexity is context-dependent, even subjective."[63] He pours scorn on the attempts by the chaoplexologists to construct "alternative universe and life histories in computers and determining which features are robust and which are contingent or ephemeral."[64]

Chaoplexity is by no means confined to American science, there are analogous developments in Europe, especially in France and Germany, though far less so in Britain, for curious reasons to do with the more traditional orientation of British post-Second World War science when even Turing's work in biology, a pioneering endeavour, was not continued, much to the chagrin of Margaret Boden and others.[65] However, in France there were a number of developments along these lines, starting with the fact that Benoit Mandelbrot originally came from France, though he did most of his innovative work on fractals in America.[66] It was in France, too, that another mathematician, René Thom, developed his ill-starred Catastrophe Theory, from which so much was expected and so little eventuated. Albert Libchaber made a significant contribution to Chaos Theory in Paris with his delicate experiments on the onset of turbulence, which predated Mitchell Feigenbaums' computer-generated work on bifurcations leading to chaos.[67]

Major work on self-organization with application to biological phenomena was undertaken by Henri Atlan, who derives from what came to be called the Second Cybernetic based on Heinz von Foerster's idea of "order from noise", which Atlan presents as follows:

> In general beyond the formalism and mathematical techniques used in different theories, we have been able to characterize self-organization as an optimum between a rigid and immutable order that cannot be modified without being destroyed, such as that of crystal, on the one hand, and ceaseless renewal that knows no stability, such as the swirling chaos of smoke, on the other. This intermediate state is not frozen and can react to unexpected disturbances by changes that are not simply the destruction of the pre-existing organization, but rather a reorganization that permits the appearance of new properties. These properties may take the form of new structure or behaviour. Nothing a priori makes it

63 Ibid, 303.

64 Ibid, 277.

65 Margaret Boden, *Mind as Machine,* vol. II, op. cit, 1267.

66 See James Glick, *Chaos: Making a New Science* (New York: Viking, 1987), 81–118.

67 Ibid, 194.

> possible to predict these properties in specific detail – this is precisely their novelty.[68]

Atlan has obviously emergence in mind, though he does not invoke the term. Atlan seems to be aiming for a general theory of emergence based on self-organization, and that, he believes, he can explain in terms of the general thesis of "order from noise" or "the effect of perturbations that tend to produce a disorganizing effect, [so that] some systems can reorganize themselves and take new structural and functional properties."[69] But such "disorganization followed by organization" can only occur if it "results from the creation of new meanings in the information transmitted from one part to another or from one level of organization to another."[70] Atlan's theory of emergence is based on Information Theory, for, as he puts it, "the creation of meaningful information is at the core of the phenomenon of self-organization."[71] But, unfortunately, Information Theory does not have any way of specifying "meaningful information", as Claude Shannon its inventor insisted, it has nothing to do with meaning.

Atlan's close associate in Paris, Jean-Pierre Dupuy, believes that all these developments that we have previously outlined cohere together and are parts of a general approach to self-organization and emergence. He puts it in the usual Kuhnian terms that they together constitute a new revolutionary paradigm in the sciences:

> The many physico-mathematical concepts and theories that have contributed to the upheaval fit together with each other in extremely complicated ways. One thinks of "catastrophes", "attractors" and bifurcations of nonlinear dynamical systems, critical phenomena and symmetry breaking, self-organization and critical self-organizing states; the physics of disorder systems, deterministic chaos, and so. The models of this in physics make it possible to understand the mechanisms of morphogenesis, which is to say the emergence of qualitative structures at a macroscopic level that organize themselves around singularities – or qualitative discontinuities – of underlying process at the macroscopic level.[72]

68 Henri Atlan, *Enlightenment to Enlightenment, Intercritique of Science and Myth*, trans. Lenn J. Schramm (Albany: State University of New York Press, 1993), 56.

69 Ibid, 56.

70 Ibid, 56.

71 Ibid, 58.

72 Jean-Pierre Dupuy, *The Mechanization of Mind*, op. cit., 7–8.

In order to pursue such researches in France, Dupuy founded a centre for interdisciplinary study, *Centre de Recherche en Épistemologie Appliquée* (CREA), in Paris. It housed mathematicians such as Jean Petitot and biologists such as Francisco Varela who were jointly engaged on a venture to account for mind through an attempt to "naturalize phenomenology", that is, to provide the phenomenological insights of Husserl with a basis in the sciences of chaoplexity, to which we previously referred and will come back to later in Chapter 7.

In the context of France, it would be remiss to fail to mention the most ambitious and all-embracing general theory of emergence covering all the sciences, Edgar Morin's four volume work under the general title *La Méthode*. But it would be presumptuous to pretend to be able to say something worthwhile about it in a few words. This is theorizing at its grandest and most speculative. The material dealt with embraces all that is contained in this book and much more besides. But the approach adopted is almost the opposite of that exemplified here. To deal with it fully would require a book of equal size, so a few words will have to suffice for the present.

Morin's approach can best be gleaned by listing the main sources from which it derives. Pride of place must go to Cybernetics and Information Theory as expounded by Norbert Wiener and Claude Shannon. Secondly, there is inspiration drawn from the previously referred to Second Cybernetic of Heinz von Foerster as conveyed to Morin by Atlan, as he himself acknowledges:

> At the origin of the ideas I develop herein, I find first of all Henri Atlan, who woke me from my experimental sleep by initiating me to the idea of creative disorder, then to its variants (chance as organizer, disorganization/ reorganization). Atlan introduced me to von Foerster… von Foerster made me discuss Gunther, Maturana and Varela. Each in his own way allowed me to look at the invisible, the notion of self, and to introduce the notion of subject.[73]

This makes it clear that the work moves in the one theoretical sweep from chance and disorganization to self and subjectivity. In fact it moves even further beyond the subject to the whole noosphere, a concept introduced by Teilhard de Chardin and invoked in volume four, *Les Idées*.

Morin's whole approach to the subject and to mind in general, as expounded in volume III, *La Connaissance de la Connaissance,* is suffused through and through with these influences, especially that of Cybernetics and Information

73 Edgar Morin, *Method: Toward a Study of Humankind*, vol. I, *The Nature of Nature,* trans. J.L. Roland Bélanger (New York: Peter Lang, 1992), 24.

Theory, which is exemplified in numerous neat little diagrams of recursive feedback loops.[74] It is based on the idea of computation as the key principle in a vastly exaggerated version of Cognitivism. As his translator Roland Bélanger explains:

> The self is cellular, and every cell computes. Therefore, *Computo, ergo sum* is and must be forerunner of *Cogito, ergo sum.* The preliminary stage, the *computo*, is essential to Morin's view of Man, viz. that Man is physical and biological, in other words, natural not super-natural.[75]

This is an extreme version of neo-Cartesian computational materialism as current in the late twentieth century. It is what this book is mainly directed against as we shall show in Chapter 7.

In Germany, too, there are analogous departures towards a general theory of emergence some of which are almost as ambitious as Morin's. They usually take the form of a theory of self-organization. Basing themselves on the prior biochemical work of Manfred Eigen and the synergetics of Hermann Haken, a number of scientists and thinkers have developed such general theories, among which the most ambitious is that by the mathematical physicist Klaus Mainzer.[76] Mainzer holds that such a general theory, taking the one form of a mathematics of non-linear complex systems, will apply to all of reality at all levels from matter to mankind, as indicated by the subtitle of his book, *The Complex Dynamics of Matter, Mind and Mankind.*[77] As he puts in his Introduction:

> The book shows that the theory of nonlinear complex systems cannot be reduced to the special natural laws of physics, although its mathematical principles were discovered and at first successfully applied in physics. Thus it is no kind of traditional "physicalism" to explain the dynamics of lasers, ecological populations, or our brain by similar structural laws. It is an interdisciplinary methodology to explain the *emergence* of certain macroscopic phenomena via the nonlinear interactions of microscopic elements in complex systems. Macroscopic phenomena may be forms of

74 Edgar Morin, *La Méthode,* vol III, *La Connaissance de la Connaissance* (Paris: Edition du Seuill, 1986).

75 Edgar Morin, *Method*, vol I, *The Nature of Nature*, op. cit, xxxi.

76 Hermann Haken, *Information and Self-Organization: A Macroscopic Approach to Complex Systems* (New York: Springer, 2000).

77 Klaus Mainzer, *Thinking in Complexity: The Complex Dynamics of Matter, Mind and Mankind* (New York: Springer, 1994).

> light waves, fluids, clouds, chemical waves, plants, animals, populations, markets, and central cell assemblies which are characterized by order parameters.[78]

Just like Morin, Mainzer believes he has a mathematical method that is the key to solving all such problems of complexity in science and some in philosophy as well. He holds that "the theory of nonlinear complex systems has become a successful problem solving approach in the natural sciences – from laser physics, quantum chaos and meteorology to computer modelling in chemistry and computer simulations of cellular growth in biology."[79] And he goes on to maintain that "the social sciences are recognizing that the main problems of mankind are global, complex and non-linear, too…"[80] Thus he arrives at his main position that "if this thesis in computational dynamics is correct, then indeed, we have a powerful mathematical strategy to hand for interdisciplinary problems of natural science, social science and the humanities."[81] It were a consummation devoutly to be wished.

Critics have not been slow to point out that general mathematical approaches that apply to many diverse phenomena explain little or nothing about what is going on in the specific cases and cannot account for the emergent phenomena that arise. For it is possible for varied emergent phenomena that have nothing in common, such as the ones that Mainzer lists, to be described in a purely formalistic way by the same mathematical equations yet be utterly different in nature. Painful lessons have been learned about this from the mathematics of Thom's Catastrophe Theory. This was supposed to account for catastrophic breakdowns ranging from attacking dogs to the fall of civilizations. There was nothing amiss with the mathematics but what it told us about canine psychology or real historical catastrophes was next to nothing. There have been similar problems with Chaos Theory which in fact has been found to have very limited applications in the natural sciences; but originally the claim made for it was that it was ubiquitously relevant in all sciences.

Mainzer has similar aspirations for the humanities which he extends to philosophy as well:

> It is obvious that the complex systems approach delivers solutions to the mind-body problem which are beyond the traditional philosophical

78 Ibid, 1.
79 Ibid, 1.
80 Ibid, 1.
81 Ibid, 1.

> answers of idealism, materialism, physicalism, dualism, interactionism, etc.[82]

Unfortunately, for most philosophers and neurologists this is far from obvious. Nevertheless, as a general strategy to account for neurological functioning, the complex system approach has something to be said for it. As Mainzer puts it,

> "the emergence of mental states (for instance pattern recognition, feelings, thoughts) is explained by the evolution of (macroscopic) interactions of neural cells in learning strategies far from thermal equilibrium. Cell assemblies with mental states are interpreted as attractors (fixed points, periods, quasi periodic or chaotic) of phase transitions."[83]

This is certainly a valid approach to neurology, but it is far from solving the problem of mind; for what is called for as he puts it later, is "to bridge the gap between the neurobiology of the brain and the cognitive sciences of the mind, which traditionally has been considered an unsolvable problem."[84] Sad to say, the problem is still unsolved, and although we can give bold investigators, like Mainzer, credit for helping to narrow the gap between brain and mind, they have certainly not bridged it.

Another such approach to a general theory of self-organization is that advocated by Bernd-Olaf Küppers, who together with his fellow editors Wolfgang Krohn and Helga Nowotny have considered such a theory, if accomplished, to amount to a veritable Kuhnian scientific revolution, a paradigm shift affecting all the sciences.[85] Küppers presents it as "a paradigm shift that, while having its roots in physics, manifests itself above all in all interdisciplinary connections. It concerns the so-called 'paradigm of self-organization' and its application in wide areas of the natural sciences, along with first indications of a possible application in the humanities."[86] This is more or less the tone adopted by all the contributors to this multidisciplinary edited work.

However, Küppers is cautious enough to point out that such a general theory of self-organization to explain emergence does not yet exist; it is merely a project for the future:

82 Ibid, 8.
83 Ibid, 7.
84 Ibid, 150.
85 W. Krohn, G. Küppers and H. Nowotny, eds. *Self-Organization*, op. cit.
86 Ibid, 52.

> …the paradigm of self-organization embodies the further difficulty that it does not yet contain a unified theoretical nucleus. We cannot yet speak of a theory of self-organization. The theoretical nucleus of a paradigm of self-organization is at present a conglomeration of widely differing approaches, and it remains to be seen whether and to what extent this nucleus will be capable of yielding a unified theory.[87]

At the conclusion of his article he puts this even more strongly: "here we sum up a fundamental problem, for… the present state of the paradigm of self-organization is characterized above all by the great heterogeneity of the individual theoretical approaches."[88] And this is where the problem has remained ever since, for no further progress can be seen towards the development of a general theory of self-organization and emergence.

Nevertheless, enthusiasm for some such outcome has not waned. Thus Margaret Boden has not lost hope for a general theory of self-organization. She writes as follows:

> The four concepts first mentioned – embryogenesis, adaptation, evolution and purpose – are fundamentally akin… All four are example of self-organization… By the middle of the twentieth century, the cybernetic movement has made more people willing to believe that these phenomena could be scientifically explained. Cybernetic explanations weren't mechanistic in the Cartesian sense. They focused not on matter or energy, but on abstractions, such as information, computation, adaptation, or equifinality. These were assumed to be instantiated in physical systems, notably in organisms.[89]

Unfortunately, these phenomena are still far from being fully scientifically explained, nor has it been shown that any one theory will hold for all of them. Towards the end of the twentieth century Gerhard Roth and Helmut Schwegler write as follows:

> How the fertilized egg (zygote), through cell division, differentiation and migration, develops into an adult organism is not understood; how the incredible complexity of the brain is produced, how neural networks

87 Ibid, 52.
88 Ibid, 61.
89 Margaret Boden, *Mind as Machine* op. cit, 1249.

> function in the context of perception, memory and control of behaviour is not understood either.[90]

It is doubtful whether adaptation, evolution and purpose are any better understood. The idea of a general theory to explain them all is certainly out of reach. Cybernetics is not such a theory and has not contributed all that much though at the time of its formulation it seemed full of promise.

Boden's was an exceptional trans-Atlantic voice and another was Stephen Wolfram, an exponent of cellular automata, for as might have been expected from their native scientific traditions, most British scientists were averse to grand schemes and overarching theories, preferring instead detailed empirical work. This showed itself in the opposition that Kauffman encountered in British biology, as in the previously quoted remark from Robert May that his kind of work was biologically trivial. Evolutionists in Britain were particularly opposed to Kauffman's statement that "self-organization is a natural property of complex genetic systems. There is 'order for free' out there, a spontaneous crystallization of order out of complex systems, with no need for selection or any other external forces."[91] The idea that order can emerge out of an abstract system, genetic or otherwise, is anathema to Kauffman's British friend Brian Goodwin, who objects as follows:

> Look at what we know about the structure of genes, how they are expressed, the incredible details of metabolic machinery that are now known. All that is true. I don't deny those achievements. I just insist they tell you nothing important about biological form, how form is generated… knowing the structure of H_2O gives you no knowledge as to why water goes down a plughole in a vortex. We need a concept of the whole organism as the fundamental entity in biology and then understand how this generates parts that conform to its intrinsic order.[92]

On this view, each organism is specific and realizes its own unique form; there are no "laws of self-organization and emergence", as the subtitle of Kauffman's book declares.[93]

---

90 Gerhard Roth and Helmut Schwegler, "Self-organization, Emergent Properties and the Unity of the World", in W. Krohn, G. Küppers, and H. Nowotny, eds. *Self-Organization,* op. cit, 37.

91 Quoted in Roger Lewin, *Complexity*, op. cit., 24.

92 Ibid, 35.

93 Stuart Kauffman, *At Home in the Universe: A Search for the Laws of Self-Organization and Complexity*, (Oxford: Oxford University Press, 1996).

According to the British biologist Jack Cohen and the mathematician Ian Stewart, "we do not have a good formal theory of emergence, but we can pin down some of the general mechanisms that come together to generate emergent phenomena."[94] The governing assumption behind that remark is that we might never have one, since emergence is such a varied phenomenon. This does not preclude the discovery of close parallels and analogies that constitute "general mechanisms" behind emergent phenomena.

Yet the unspoken conviction that there must be a general theory of emergence or of self-organization or at least a general theory of systems persists for it has such a weight of historical expectation behind it. Over nearly a century it has generated a number of movements, each of which has climbed on the back of the previous ones, while at the same time claiming to be new and original. In sequential order, there have been General Systems Theory, Cybernetics and Information Theory, Complexity Theory and Self-Organization theory. Each has had its day of fame and popularity, each has aroused immense hopes and kindled burning enthusiasm, and in due course each has petered out with little to show for all the effort expended.

3 The History of Hydrodynamics

If not much more is to be expected from the generalist approaches, then perhaps more is to be hoped for from the specialist ones. Thus instead of one all-embracing conception of emergence based on some general theory of systems or complexity or information we must look to a differentiated set of types of emergence based on a number of special fields. In each such field not only does emergence differ, but so, too, does the very idea of what constitutes complexity, system or even law. Clearly, as we shall see, what these notions mean in biology is very different from what they mean in physics, and what they mean in psychology is different again from what they mean in biology. It is not just, as Anderson puts it, "at each stage new laws, concepts and generalization are necessary",[95] but that furthermore at each stage another kind of emergence enters.

This leaves us with the key problem of what are the relevant fields or levels to be considered in this respect. Anderson's epochal article "More is Different" provides the first approximation to such a listing. He sets out what he calls a

94 Ian Stewart and Jack Cohen, *Figments of Reality: The Evolution of the Curious Mind* (Cambridge: Cambridge University Press, 1997), 72.

95 P.W. Anderson, "More is Different", op. cit, 222.

hierarchy of at least eight major levels starting with elementary particle physics and ending with the social sciences, in between those extremes there are the usual departmental divisions, such as solid-state or many body physics, chemistry, molecular biology, cell biology, physiology and psychology; and he goes on to say that there are many more, for "surely there are more levels of organization between human ethology and DNA than there are between DNA and quantum electrodynamics and each level can require a whole new conceptual structure."[96] Obviously, this is far more levels than we can consider here, though we made a start in the previous chapter. And the same applied to Morowitz's work where he lists even more levels of emergence; he presents "a catalogue of 28 observed instances that have emergence in common, but vary over an enormous range of agents, interactions, hierarchical levels, and character of the interaction rules and pruning rules."[97] He goes on to add that "these examples should illustrate the many meanings of emergence…"[98] and therefore, the many different approaches to it.

Twenty eight meanings of emergence are clearly far too many, and even eight are more than we can deal with. We shall restrict ourselves to just five, Hamlet's quintessences, the fields we outlined in the previous chapter and designated as E^1, E^2, E^3, E^4 and E^5. E^5 is the one that interests us most, and towards whose elucidation this whole work is striving: it is, of course, the emergence of human mind as the quintessence of dust. E^1 is the field furthest removed and most elementary and for those reasons of least concern to us here. It covers the emergence of matter which, as we saw previously, is a highly technical, if not outright speculative, issue in fundamental physics and cosmology. Problems of emergence occur at this elementary level, of course; they are of the kind that Humphreys treats in his work: "candidates for ontologically emergent phenomena include entangled states in quantum mechanical systems, covalent bonding in chemistry, and ferromagnetism in condensed matter physics."[99] We have not considered them in this work.

E^2 or solid-state physics generates a huge host of phenomena of emergence since most of the material things we observe or deal with in our practical work belong to this level. In fact, the stuff we are most concerned with is ordinary water and the science that deals with its behaviour, namely hydrodynamics, is a very old science that almost every major scientist or natural philosopher, as

96 Ibid, 228.

97 Harold Morowitz, *The Emergence of Everything*, op. cit, 25.

98 Ibid, 25.

99 Paul Humphreys, *Emergence*, op. cit, 42.

they were once known, was involved with going right back to Thales and Archimedes; the list includes Descartes, Newton, Daniel and John Bernoulli, Jean le Rond d'Alembert, Leonard Euler, Joseph Louis Lagrange, and almost every subsequent mathematician and physicist till the twentieth century. The great Victorian physicist, William Thomson, declared: "Now, I think hydrodynamics is to be the root of all physical science, and is at present second to none in the beauty of its mathematics". Darrigol states that he "wanted hydrodynamics to reign over all the physical sciences, from the intimate constitution of ether to the beautiful pattern of ship waves."[100]

Thomson was, of course, wrong about this leading role of hydrodynamics among all the physical sciences, as Einstein, Bohr and the whole subsequent course of physics which abolished the ether showed. Nevertheless, the history of hydrodynamics as a science presents a model historical lesson of how a scientific discipline develops, one that goes quite counter to Kuhn's model of paradigm revolutions, as Darrigol brings out. We might learn from this how the sciences of emergent phenomena might develop in the future when the whole period of the Classical reductive science is over and there are no more paradigm revolutions in store.

Looking ahead to Chapter 6 it is worth noting that hydrodynamic problems were Descartes' initial introduction to natural philosophy. He modelled his whole view of matter on hydrodynamic phenomena, and this enabled him to reject all the earlier approaches to matter such as the Aristotelian *hyle* or Gassendi's "atoms" or Gilbert's theory of magnetism, that is, action at a distance, or any of the more arcane alchemical views. His new conception of matter as mechanical served as the counterpoise to his own view of mind as thought, that is, as in all respects the opposite of matter. He began his work on matter, under Isaac Beeckman's tutelage, by attempting to solve Simon Stevin's paradox of water pressure. He went on from this to base his conception of matter, above all, his theory of vortices, on hydrodynamic models. Newton's refutation of that hydraulic approach to matter marked a crucial stage in the development of physics, as Stephen Gaukroger points out:

> Newton had examined the motion of bodies in fluids in Book II of his *Principia*, arguing in some detail that it is the density of the medium, and not the subtlety of its parts, that is the principal factor in determining the resistance it offers. On the basis of this he concluded that the vortex

100 Olivier Darrigol, *Worlds of Flow*, op. cit, v.

> theory was incompatible with Kepler's third law, and that vortices could not be self-sustaining…[101]

From that point on, it was from Newton that the physics of matter took off, and Descartes was effectively by-passed. His view of matter was soon refuted, but his correlative view of mind held on for the next three centuries. The explanation of this anomaly is a crucial issue in the history of philosophy which has as yet not been considered.

Another crucial stage in physics occurred during the late nineteenth century, when hydrodynamics, as a general science of liquids and gases, now called classical fluid mechanics, was translated and absorbed into the more embracing new science of statistical mechanics. It was then that it became apparent that water and its behaviour were emergent phenomena based on a more fundamental atomic reality. Morowitz notes that "the central concepts of emergence trace back to the statistical mechanics of Ludwig Boltzmann, James Clark Maxwell and Josiah Willard Gibbs." As he explains:

> The founders of statistical mechanics assumed the atomic molecular view of matter and further posited that the atoms and molecules obeyed the laws of mechanics. They were interested in showing how the macroscopic laws of thermodynamics and kinetic theory could be obtained from the mechanics of the reductionist agenda, the atoms and molecules. By dealing with ensembles of particles or ensembles of states and showing that the macroscopic observables were averages over microscopic states, they were able to deal with variables like pressure and temperature as emergent properties.[102]

In brief, the pressure of a gas is the average number of molecules inside a given region and the temperature is the average kinetic energy of those molecules. This furnishes the explanation of the law that pressure is proportional to temperature known as Boyle's law.

There is a huge variety or ordinary phenomena that this science covers, things like tides, waves, winds, waterfalls, eddies, atmospheric flows, smoke in the air, sound waves and detonations. There are mathematical equations governing all of these which are mostly difficult or impossible to solve except as approximations. Computers have been a great help in this respect but even the

101 Stephen Gaukroger, *Descartes: An Intellectual Biography* (Oxford: Clarendon Press, 1995), 383.

102 Harold Morowitz, *The Emergence of Everything,* op. cit, 9.

most powerful computer cannot overcome the inherent difficulties of complexity. The reason for this is given by David Potter:

> …in what may be described as many-body problems, the system of interest is an assembly of so many particles that it is inconceivable that the complete internal motion may be resolved exactly… Typically for a laboratory liquid, for example, 10^{23} particles are involved. Nevertheless, in assemblies of many particles, the macroscopic properties of the whole system may be rather simple and dependent on average properties of the particles. While statistical physics is of great importance in such problems, it is frequently the case that the problems become nonlinear and intractable. These systems are, however, amendable to the application of computational mechanisms, since we may infer macroscopic properties by studying the *self-consistent* motion of typical particles.[103]

However, it is important not to exaggerate what even the most powerful computers of the future, perhaps quantum computers, will be able to achieve in dealing with intractable problems of such vast numbers, for the reasons that Weinberg gives:

> If you knew everything about water molecules and you had a computer good enough to follow how every molecule in a glass of water moved in space, all you would have would be a mountain of computer tape. How in that mountain of computer tape would you ever recognize the properties that interest you about water, properties like vorticity, turbulence, entropy and temperature? There is in the philosophical literature a term, emergence, that is used to describe how, as one goes to higher and higher levels of organization, new concepts emerge these are needed to understand the behaviour at that level.[104]

Weinberg, who had been Anderson's antagonist in the Congressional hearings surrounding the great controversy about funding for the Supercollider, seems to have taken a leaf from Anderson's book and recognized the role of emergence in physics. As we have already noted, Anderson's key argument turns on the point that "the ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the

103 David Potter, *Computational Physics* (London: Wiley, 1973), 116.

104 Steven Weinberg, "Newtonianism, Reductionism and the Art of Congressional Testimony", 350.

universe… The constructionist hypothesis breaks down when confronted with the twin difficulties of scale and complexity."[105] This necessitates an emergent jump from quantity to quality. Anderson provides scientific explanation for this Hegelian dialectic in many-body physics:

> …we have begun to formulate a general theory of just how this shift from quantitative to qualitative phenomena takes place. This formula, called the theory of "broken symmetry", may help in making more generally clear the breakdown of the constructionist converse of reduction.[106]

The constructionist converse of reduction is the view espoused by most scientist until very recently as expressed by Victor Weisskopf, which we previously quoted: "once new fundamental laws are discovered, a large and ever increasing activity is to apply the discoveries to hitherto unexplained phenomena".[107] But it is one thing to discover the general fundamental laws covering a class of phenomena, it is quite another to apply it to specific cases or particular manifestations, which can at times be much more difficult. As we shall see presently, it is this discrepancy which dogged the history of hydrodynamics from the very start. The general laws of water flow were formulated quite early, mostly in the eighteenth and early nineteenth centuries by mathematicians and hydrologists, such as d'Alembert, Euler, Lagrange, George Gabriel Stokes and Claude Louis Navier. But these laws proved ineffective and almost useless in calculating and predicting the real behaviour of water. Between the "in principle" theory and "in practice" application there is always a large gap. It took much further work by many theoreticians and practitioners, above all by Ludwig Prandtl early in the twentieth century, both in theoretical elaboration and experimental refinement, before hydrodynamics could be turned into a useful science. This is the basic reason why science does not stop even if all the fundamental laws are known, which is far from the case, and why all talk of an end to science, as mooted by Stent and decisively affirmed by Horgan, is widely premature, as we have argued.

Explaining emergence in all its various manifestations is an overwhelming problem that science has to confront, and it is an issue not to be dealt with through the discovery of new fundamental laws, but through the detailed study of how such laws apply to particular cases and special classes of phenomena. This is especially so when the phenomena prove to be particularly

105 Philip Anderson, "More is Different", op. cit, 222.
106 Ibid.
107 Ibid, 221.

unexpected and surprising. This is the case with the so-called dissipative systems that Prigogine has made his own domain, even though he did not discover them. There are many such oscillatory reactions in chemistry which previously were ruled out on thermodynamic grounds that equilibrium states should rapidly ensue. This is clearly not so with the most renowned of these chemical processes, the so-called Belousov-Zhabotinsky reaction.

What makes these far from equilibrium systems so interesting is that they display the emergence of order out of chaos and seem to have close analogies with the emergence of life. Like life, they are macroscopic manifestations of orderly dynamics that do not depend on the specific microscopic elements which enter into them and constitute them. As Cohen and Stewart explain:

> The B-Z patterns are consequences of a meta-law, a law about laws, a common pattern shared by an entire class of rules. They are only a representative of a much more universal phenomenon, of a common pattern that applies equally to chemicals and to slime mould. Those phenomena depend not on the detailed properties of the underlying setup, such as which chemicals you use, or even whether you use chemicals at all. They depend on certain universals – a planar spatial distribution; local oscillations; certain kinds of short-range coupling. They are independent of any underlying chaos; indeed, they are independent of a great many regularities as well.[108]

In all these respects, these reactions resemble the chemical processes that constitute live forms. But is this any more than just a passing resemblance? Did life in fact emerge in some such purely chemical way? We have no answers to these questions since we do not know and are still far from knowing how life originated, as we argued in the previous chapter. Nevertheless, dissipative systems constitute a form of emergence that is perhaps half way between the emergence we have encountered in hydrodynamics and solid-state physics in general and the emergence of biological forms. The distinction of these two types of emergence can be expressed as that which arises out of wholly disorganized systems, clouds of countless atoms all totally identical, governed by thermodynamic statistical laws, and that which arises from organized systems constituted by dynamic processes and reactions of chemicals each of which is in its own way unique. It is the distinction between emergence arising out of mass amorphous formations and that of distinctive forms of individualised beings. It is the difference between what we have called E^2 and E^3.

108 Jack Cohen and Ian Stewart, *The Collapse of Chaos*, op. cit, 266.

Cohen and Stewart elaborate further on this distinction at great length in their book:

> We distinguish two kinds of emergence "regular" and "super". We call the regular emergence "simplexity" and the super version "complicity". The archetypal examples of complicity are evolution and consciousness, and the complicist view gives interesting insights into questions of mind and human spirit.[109]

Animal mind and human mind, that is, E^4 and E^5, are types of emergence we shall discuss later; and as we shall show, complicity is a crucial feature of their functioning. Here we will merely deal with evolution, namely, with biological systems. According to Cohen and Stewart, "the prime example of complicity is evolution, for which the two systems of rules are the chemistry of DNA and the systematic way in which organisms interact with their environment".[110]

It is worth pausing at this point to review once again the previous arguments launched against the generalist views concerning the possibility of general systems theories or theories of complexity. In the light of what Anderson, Cohen and Stewart, and many others maintain, it is clear that what is meant by system or complexity or for that matter information or any other such general term varies from context to context. In fact, we might say that each of them has a somewhat different meaning in each of the major fields of emergence. What is a system at the subatomic level is vastly different from what is a living system or organism; and these, in turn, are just as different from organizations. It is the same with complexity, for which Cohen and Stewart have coined such higher order terms as "simplexity" and "complicity". Whether "information" can escape this kind of multiplicity seems highly dubious, for what J.J. Gibson calls "information" in perception is very different from Shannon "information", as we shall see in the next chapter.

Cohen and Stewart make the startling pronouncement that emergence is itself subject to evolution: "as creatures evolved, the evolutionary process itself evolved, and both became more complex together."[111] This fits in well with our view that higher forms of emergence, or metaphorically expressed, emergence raised to higher powers, arise in the course of the development of matter in the universe, and eventually in the course of the evolution of life on earth. Thus we can speak of the emergence of emergence. It is the emergence of evolution as

109 Ibid, 397.
110 Ibid, 418.
111 Ibid, 320.

itself a novel and higher mode of emergence which in turn generated the still higher forms of emergence of animal mind and human mind. At each of these levels emergence is different and no one theory can encompass them all, and furthermore, as Cohen and Stewart contend, "they are necessarily inaccessible to reductionist explanation."[112] Only a Darwinian type of evolutionary explanation can account for them.

Evolution is, of course, the key aspect of Darwin's great theory, which he elaborated in the middle of the nineteenth century when the reductionist course in all the other sciences was in full swing. A mere century later, it was generally believed that the work of reduction throughout the sciences had been completed and that this included Darwin's theory as well, for it seemed that life, too, had been reduced to chemistry. But, in fact, this was far from the case, since Darwin's theory was based on emergence not on reduction. And, as we previously argued, the reductive sciences could not account for the emergence of life or that of mind either. The work of explaining emergence is still left to be done even when that for reduction has been completed.

Darwin's so-called "theory" is actually not a theory at all in the reductivist sense; as has often been noted, it is not falsifiable in the usual experimental way. It has always occupied a very anomalous place in the reductivist scheme of the sciences. It does not propound any fundamental laws, it has no predictive power, it cannot be deduced from any of the basic laws of physics and chemistry. There is nothing in the universal development of matter to predicate or predict that evolution should take place. We can read Darwin's discovery not as a theory at all, but as a methodological prescription of how emergence in evolution is to be studied. What it prescribes is that biological species should be arranged in systematic chronological orders of ancestors and descendants; that between any two such adjacent species that have survived there are always extinct intermediate ones or missing links to be discovered; that the survival of a species depends on its "fitness", that is, its adaptability to environmental conditions; and so on. This, indeed, is a prescription for how all the higher forms of emergence can be studied. As we shall show in the next chapter, the account of emergence at the higher levels must frequently takes this Darwinian form and it constitutes a mode of adequate scientific explanation.

Those who have argued for an "end to science" thesis, such as Stent and Horgan, have failed to grasp this. They believe that once the fundamental reductive laws are known then in principle all is known. All that is now left to do is to apply these laws, and mere application they consider to be "normal science",

112 Ibid, 440.

that is, boring science that is no longer graced with the excitement of discovery. Both of them follow a Kuhnian view of the history of science that considers only paradigm revolutions to be truly original work in science, what follows in the periods of mere normal science is the hackwork of refinement and application. Hence, once paradigm revolutions no longer occur because all fundamental laws are known, then science enters a period of ever diminishing returns, as Stent put it; and, according to Horgan, when it is not engaged in practical problem solving, which is useful science or technology, it sinks into the morass of ever more *recherché* speculation no longer based on experimental evidence. Post-paradigmatic science is, in effect, the moribund state of science.

Such a Kuhnian reading of the history of science as one of incessant paradigm revolutions which are bound sooner or later to come to an end is belied by the history of hydrodynamics. Darrigol's detailed account of this science goes quite counter to the Kuhnian model. In a brief note at the end, he extrapolates his conclusion to most of the major physical sciences:

> According to Thomas Kuhn, applications contribute to the smooth, gradual expansion, and consolidation of normal science. Significant conceptual change can only result from the accumulation of major anomalies, in which case a global revolution occurs and a new paradigm emerges. The above cited examples of application-induced change fit neither the smooth paradigmatic phase nor the revolutionary one. They do not lead to the overthrow of the theory, yet they entail transformations of such magnitude that the word "application" sounds inadequate. The phenomena are not passively subjected to a rigidly established theory, but instead react upon the content and structure of the theory.[113]

The examples he gives range from William Hamilton's attempt to apply mechanics to light rays, Delaunay's work on the perturbations of the moon, Saint-Vernant's theory of elasticity, Lorentz's application of Maxwell's electrodynamics, and Feynman's path-integral formulation. But most of his book is devoted to recording and explaining why it took almost two centuries from the Bernoullis to Prandtl to turn the basic mathematical theories of hydrodynamics into a science that hydrologists and engineers could actually apply in practice.

The sciences of emergent phenomena are much more likely to proceed on the model that Darrigol charts rather than the one that Kuhn presented. In this

113 Olivier Darrigol, *Worlds of Flow*, op. cit., 323.

sense it is true that paradigm revolutions are over, partly because there were no such things in the first place, for as Darrigol notes: "the myths that make Newton the sole creator of mechanics, Cauchy the father of the theory of elasticity, Clausius the founder of thermodynamics and Maxwell the unique inventor of modern electrodynamics do not resist historical analysis".[114] But mainly it is because the great Classical age of scientific reductivism is just about over, and a new age of emergentism has already begun.

We have now reviewed in general terms the science of matter and can turn to that most remarkable of all emergent phenomena, that of mind. It is what interests and concerns us most since we are minds and self-knowledge is what we are after. But in order to know ourselves we need to know our ancestral origins or where we come from in an evolutionary sense. And that takes us first of all to animals, that is, animal minds with which we shall begin in the next chapter.

114 Ibid, 323.

PART 2

The Science of Mind

CHAPTER 3

Animal Minds

1 Griffin and Edelman on Animal Minds

Man is the paragon of animals, Hamlet said. Indeed, by nearly all measures Man is the best of beasts. Noble in reason, infinite in faculties, in form and moving, express and admirable in action, and so on – but at bottom still a beast who fills Hamlet with disgust. Hamlet is weary of the humanist declarations of the grandeur of Man that have resounded down the ages. Man is the measure of things, pronounced Protagoras at the start of Western Philosophy. Whatever he meant by that grandiloquent utterance, it is true that Man is capable of taking the measure of things – so perhaps, better put, Man is the measurer of things. And that capacity, usually called reason, comes from the extraordinary mind that Man undeniably possesses. Whatever Man is or is not relates to this mind. Hence, what is mind and where does it come from, that is the question, which will be our concern in this and the next chapter. To answer it we must first turn to the animals, among which, as said, Man is the paragon.

Man's mind is, indeed, exceptional; but Man is not the only animal with a mind. There are other animals that also have minds. In fact, if there were no other species of animals with minds then Man could not have attained one, at least according to Charles Darwin. In an evolutionary sense this is undoubtedly true, for the human mind could not have derived from mindless creatures directly, only from those that already have minds. Hence, to understand the nature of the human mind we must know its origin in animal mind and study that first. Only that understanding will enable us to know what humans inherited from their animal ancestors; and also how they have surpassed their atavistic heritage to become the paragon of animals.

But do animals, apart from Man, have minds? There have been many philosophers, most notoriously Descartes, who have denied this; he maintained that animals are mere bodies and that bodies are machines or automata; a fateful pronouncement whose full force has not been felt till the present time when computing machines were invented. Descartes allowed, of course, that humans, too, have bodies, but he held that they are more than mere bodies since they have minds to initiate and direct their bodily motions, whereas animals lack any such volitional or intellective faculties.

The general term that Descartes uses for these faculties is the Latin "mentis", but mostly he employs the synecdoche of "*pensée*" or thought for the whole of the mind. Descartes has still not completely escaped the traditional Aristotelian conception which endowed the lower faculties of nourishment and sensitivity to the body, treating them as vegetative and sensitive psyches, and therefore as bodily and separate from *nous* or thought, the intellective faculty of mind. Hence, Descartes has no hesitations in ascribing sensations, perceptions and emotions to animals. However, he gives a purely physiological account of these in terms of material animal spirits that bring about bodily effects. Clearly, there is some confusion in Descartes as to what is or is not mental; he is not what we would now call a consistent Dualist, for that came only later with Locke. We shall return to this point in our account of Descartes in Chapter 6.

In more recent times the denial of mind to animals has come from the Behaviourists. Animals are treated as reflex organisms subject to stimulus and response reactions and only capable of changing their reactions through classical conditioning, such as Ivan Pavlov investigated, or operant conditioning such as J.B. Skinner studied. Behaviourism is, of course, equally a rejection of the presence and operative role of mind in humans as well, who are also taken to be stimulus-response organisms, in that respect no different from animals. At this point we shall not consider any further philosophic or scientific points of view that deny mind altogether. Whether this includes neo-Behaviourist philosophers such as Gilbert Ryle, Ludwig Wittgenstein and more recently Daniel Dennett is a moot point we shall also not debate here.[1] To what extent Cognitivist psychologists and philosophers in general are also implicated in a denial of mind we will leave for later consideration in Chapter 7.

If we grant minds to animals, at least as a working presupposition, then there are three difficult questions to be settled. The first concerns the question whether all animals have minds or only some. If it is the latter, then which are the animals with minds, and where and how is the cut off point between them and those without minds to be drawn? This is an issue we have already raised in Chapter 1 on the origin of animal mind and we will pursue it much further here. The second question has to do with the nature of animal mind: is animal mind the same as human mind or are there fundamental differences between them? If it is the latter, then what are these differences, and how do they manifest themselves? A third question has to do with the emergence of mind in animals due to evolutionary developments: how does evolution generate mind? What kinds of changes are relevant? Is it only a matter of the size and

1 Harry Redner, *The Tragedy of European Civilization: Towards an Intellectual History of the Twentieth Century* (New Brunswick, NJ: Transactions, 2015), Chapter 5.

complexity of the brain, or is much more than that involved? Does the animal's body and its relation to its environment also play a crucial role?

The approach we shall adopt in answering these questions is an evolutionary and ecological one in line with that which Darwin originally presented in his *The Descent of Man* (1871). It is evolutionary in that we shall consider the full range of animal capacities in a comparative way, seeking to order species in a hierarchy of abilities as well as one of phylogenetic descent; for the two need not necessarily coincide, as, for example, some cephalopods are more intelligent than many vertebrates. It is ecological in so far as any such capacities must be related to an animal's environment, its *Umwelt* to use a term coined by Johann von Uexküll. The nature of Umwelt – the environmental niche that an animal inhabits and to which it responds – is a crucial factor in determining its mental powers. Darwin, like many other ethologists since, tended to be unduly impressed by single cognitive feats found in certain animal species, often based on reports he had gleaned from purely anecdotal sources. He did not place these in the context of the animal's full range of activities in relation to its environment. He was right, however, in pointing to the continuities that hold between animals and humans, above all the abilities displayed by apes which are so close to the human level.

If we grant minds to humans, then we are bound to grant minds to the higher apes, especially to chimpanzees, our closest evolutionary relatives, who in most respects behave as we do, even though there are crucial differences, as we shall see. Such apes have the full panoply of faculties present in humans, not excluding thought and creativity, but at a much more primitive level. All the main faculties of mind are certainly there: apes perceive and act, they have sensations and feel emotions, they express themselves and communicate, they enter into complex and sophisticated social and even personal relations with their conspecifics as well as human beings. The work of Jane Goodall on chimpanzees in their natural environment is very convincing in all these respects. The work of all those trying to teach apes language, albeit unsuccessfully, such as David Premack, reveals how close apes can come to this archetypal human accomplishment without actually mastering it. The reasons for the failure of apes to acquire language has generated much controversy, advisedly so, for it is a key aspect in determining what separates animals minds from human minds. For the differences between these two types of minds are much more fundamental than Darwin supposed and as other ethologists who have followed in his footsteps maintain; which is not to deny evolutionary continuity between them.

Thus there is no doubt that apes have minds. But as we go lower down the evolutionary ladder doubts begin to arise. We can move from apes to dogs with a fair degree of assurance for these are social animals with whom many of us communicate and interact on a daily basis and some of us form strong attachments and bonds. But what if we descend from dogs to rats? Or what if we go down further from rats to frogs; or further still from frogs to fishes? What if we enter the realm of the invertebrates, that of cephalopods, such as octopi and cuttlefish with their remarkable behavioural abilities, or the etymological kingdom of social insects, such as bees and termites with their impressive navigating and building capacities? How far down the evolutionary scale must we go before all semblance of mind is lost?

Purely intuitively considered, most people would maintain that dogs obviously have minds and frogs just as evidently do not. Some who are more knowledgeable are inclined to say that rats are somewhere close to the boundary between the two realms of mindless and mindful creatures. It seems that with rats and such "small deer", as poor Tom in *King Lear* calls them, we first begin to perceive the lineaments of mind. If that is so, then it is at this liminal point that the emergence of mind first begins. In terms of the quintessence of dust, it is where the transition occurs from E^3, the origin of animal life as such, to E^4, the origin of mind. The next major transition, that from E^4 to E^5, lies much higher up the evolutionary scale, somewhere between chimpanzees and homo sapiens, but at that point there is an evolutionary gap as the relevant in-between hominin species are extinct. In what follows we shall see to what extent this purely intuitive sense of things can be vindicated by science and philosophy.

We begin considering these matters by taking the views of those who are at the opposite pole to those who hold that no animals have minds, these are the people who believe that all animals have minds. There are numerous ethologists, psychologists, evolutionary theorists, and a few philosophers who explicitly or implicitly maintain this view, or come very close to it. Certainly, hardly anyone believes that amoebas have minds, but even that apparently egregious opinion is upheld by those who think of mind as being present to various degrees in all animals, such that zero degree of mind is only to be found in non-living inert matter. But that is an extreme position that only very few in the West ever entertained, though it is common in the East due to the religious beliefs of Jains and Buddhists.

An important representative of those who adopt a very catholic view on animal mind is the ethologist Donald Griffin. He tends to be very impressed with single startling examples of behaviour and is all too ready to assign mind and all its attributes, such as consciousness, intentionality and thought, on the basis of remarkable communicative performances alone. One such example is the so-called "dance language" of honeybees as investigated by Karl von Frisch, whereby a forager bee can communicate a source of nectar to the other bees in the hive by means of a dance-like movements which indicate the precise location of the food trove.[2] Even more remarkable is that group-activity of bees described by M. Lindauer when the "waggle dances executed on a swarm lead to a group decision on the part of the colony about where they should move to establish a new colony".[3] Griffin comes to the conclusion that "these dances on the swarm lead to a sort of consensus whereby the colony selects out of many possible cavities the one that has been judged by the scout bees to be the best."[4]

Griffin calls this "communicative behaviour" and has no hesitation in ascribing mental concepts both to individual bees as well as to swarms. Thus he states that "all this communicative versatility certainly suggests that the bees are expressing simple thoughts, as I have argued elsewhere."[5] He confirms this in stating that "bees are thinking about cavities, and are able to change their 'allegiance' from one they have discovered themselves to a better one they have learned about as recipients of symbolic information from the dances of one of their sisters."[6] He caps off the idea that bees engage in "symbolic information" with a long quote from Carl Jung, with which apparently he agrees, and which contains the assertions that "this kind of message is no different in principle from information conveyed by a human being [and] we would certainly regard such behaviour as a conscious and intentional act and can hardly imagine how anyone could prove in a court of law that it had taken place unconsciously…"[7]

That might well be so if a bee could be put on oath and tested in a court of law for veracity. But not only can a bee not be put into the witness box or even put on trial, it cannot pass the simplest test that a rat can perform with ease; after all, bees cannot learn to run mazes. Griffin, however, likens the location capacities of bees to those of rats, as he puts it.

> Gould interprets these data as demonstrating that honeybees employ what are called "cognitive maps"; that is, they have some sort of internal representation of the geometric relationships of important major landmarks and use this to orient their flight when displaced to a novel location within the area with which they are familiar. Similar experiments

2 Donald A. Griffin, *Animal Minds,* op. cit, Chapter 9.
3 Ibid, 192.
4 Ibid, 192.
5 Ibid, 194.
6 Ibid, 193.
7 Ibid, 194.

> with rats and other vertebrates have been carried out by Tolman and others, but insects had not been believed capable of this level of cognition.[8]

If bees are as cognitively aware as rats, then it follows that if we grant minds to rats we must do so to bees as well. Griffin certainly believes so when he ascribes to them a capacity to think. Jung apparently goes much further than that and believes that they have minds that fall not much short of those of humans. Most probably Griffin would not wish to go as far as that, but frequently, in the language he uses, he expressed himself along the same lines. However, a purely mechanistic explanation for the behaviour of bees is not difficult to envisage.

Clearly, something has gone seriously awry with many peoples' idea of what is mind and on what basis it can be assigned. Mind, we shall contend, is a holistic concept that can only to ascribed on the basis of a syndrome of defining criteria, not all of which need to be present at once in every case, some might be absent, provided most others obtain; that is to say, mind is not any one thing, but a collection of mental attributes or faculties. The syndrome includes such obvious things as the faculties of perception, sensation, emotion, cognition, attention, perceptual learning and in higher species episodic memory; and overall the kind of behaviour that is purposive. That is to say, the creature must at least to some degree show problem-solving intelligence and learning capacity requiring flexible adaptability that is more than just conditioned responses. This is merely a provisional list of criteria, for only a textbook of animal psychology would be able to exhaustively enumerate them all. Yet we can intuitively apply such a syndrome of capacities to animals. To give a simple example, the experiments in problem solving carried out by Berndt Heinrich on captive ravens seem to be convincing enough for anyone to agree that ravens are highly intelligent birds and can coordinate all their mental faculties in efficient ways to solve difficult problems; which would make it very controversial to deny them minds, unless other contrary evidence is forthcoming.[9] The same conclusion applies to Alex the parrot trained by Irene Pepperberg.[10] As for other species of less capable and less intelligent birds, the evidence is equivocal, but only an exhaustive investigation of all their faculties would make a judgement possible. Bees would obviously fail such a testing procedure. The main differences between insects, reptiles, fish and some birds as opposed to mammals in general is, as Donald states, that the latter are "much faster learners, have more powerful memories, can map out and retain a large

8 Ibid, 139.
9 Ibid, 104–5.
10 Ibid, 171–3.

environment and make fine distinctions … carry out a protracted chain of actions to achieve a goal, focus on the complex objects of their perception, and solve tricky problems."[11]

The ascription of mind to animals is an interpretative procedure that might be called "mind reading" for it shares some of the hermeneutic procedures of reading a strange text. In both cases one formulates initial hypotheses on the basis of the immediately available evidence and then one tests this against the whole text or the animal's total performance and capacities. One then amends the initial hypothesis or makes a new one and then tests this against the holistic sense one has gained of the animal in the course of getting to know it. And the more one learns about what the animal can or cannot do, the surer becomes one's "mind reading" skill and the more reliable one's judgement. Hence, only those who work closely with given species of animals are in a position to exercise such a judgement. On the whole, one would prefer the judgement of ethologists and animal psychologists to that of animal trainers or the owners of pets, even though the latter might have something to teach the former.

If one does predicate mind of an animal, then all its faculties and behaviours must be referred to and described in mental terms; one must be in a position to say that it can exercise most mental faculties and not just one on its own. But conversely, if one denies mind to an animal, then it can do nothing mental and no one single mental attribute can be ascribed to it; all that it does can only be described behaviourally. Thus, strictly speaking, since a bee has no mind, it cannot be properly be said to see, only to detect and respond to visual stimuli; it cannot communicate anything by its location dance, only induce a direction-guiding response in other bees, and so on. It is just the opposite with a raven, since that does seem to have a mind. The birds and the bees cannot be coupled together as easily as this is done in primary school.

This holistic nature of mind is even more to be stressed when considering human minds, as Thompson affirms:

> When we describe experience we are describing it as belonging to the whole person, and our descriptions have a holistic and normative character. We describe the interrelations of perceiving, intending, feeling, imagining, and acting and try to make sense of these interrelations in various norm-governed ways. By contrast, when we describe the neural processes on which experience depends, we are describing subpersonal phenomena, and our descriptions do not have this holistic and normative character.[12]

11 Merlin Donald, *A Mind So Rare*, op. cit, 121.

12 Evan Thompson, *Mind in Life,* op. cit, 208.

Thus if a certain body is taken to have a human mind, then most such mental faculties must be at once attributed to it, for these are logically interrelated such that any one mental characteristic entails all kinds of others. The mind is a complex unity not a collection of characteristics. Of course, the kind of unity and coherence that is inherent in the human mind, with its self-conscious subjectivity and Ego agency, is vastly superior to the much lower grade integrity of animal minds, as we shall see in the next chapter. Nevertheless, some such holistic features obtain for animal minds as well, depending on their evolutionary level, for the more evolved and accomplished a species is, the better integrated is its mind.

What we have argued for in relation to animal minds applies all the more so to machines. There, too, the holistic nature of mind must be invoked before applying any one mental characteristic or epithet. No machine can be said to think, see, decide, reason or do anything mental since mind as a whole cannot be ascribed to machines, at least not for the present devised or for the foreseeable future ones. Whether that day will ever come when we will be able to speak of machines as having minds is not something which can "in principle" be established or refuted, it is an undecidable proposition. Though the question of whether a device that begins to approach a mind can any longer be spoken of as a machine is one we will consider more closely in Chapter 7. As we have argued in relation to animals, one must not be overly impressed with the single performances of which machines are capable, such as playing chess or recognizing faces, which, if performed by humans, who obviously have minds, attest to high intelligence and perceptual discrimination, but if performed by machines without minds do no such thing.

There are, of course, extreme cases of brain damage or degenerative diseases where an animal or a human being can partially or completely lose its mind, though still remaining alive and active. This happens to humans in the late stages of Alzheimer's when the brain shrivels away. In the case of animals, experimental studies in brain ablation have been carried out to demonstrate analogous phenomena. The psychologist Richard Morris in Edinburgh subjected rats to hippocampal lesions in order to study the deficits that would result in their learning and memory capacities as compared to normal rats, and thereby to elicit the function of the hippocampus. He trained rats to swim to a submerged platform in a tank of cloudy water. Normal rats were able to swim directly to the platform after 10 trials, but impaired rats were not able to do so. The difference between them is even more marked when the point of entry into the tank is varied, as Patricia Churchland reports:

> If the starting point is varied, normal rats swim directly to the submerged platform. If the platform site is then shifted, normal rats learn where to go

> in one trial, whereas hippocampal [deprived] rats require many trials, This one-trial learning of location is a rough analogue of declarative memory, and the deficits in the hippocampal [deprived] rats are good, if imperfect analogues of declarative-memory deficits in the hippocampal [deprived] patients.[13]

For people to completely lose their ability to remember anything and only be able to function from moment to moment is tantamount to losing a human mind, though perhaps still retaining an animal type mind. For an animal, such as a hippocampal deprived rat, to be so unmindful of what it has learned, so as to be unable to adjust its behaviour under changed circumstances and have to relearn everything again, is tantamount to losing its animal mind. It degenerates to a purely mindless behaviouristically determined creature.

But even in such extreme cases, the ascription of mind is still an either-or, all-or-none matter. There can be no such thing as degrees of mind, only kinds of mind, which, of course, vary immensely as between different species and ultimately between humans and animals as a whole. A species of animal either has a mind or it does not, it cannot have a little bit of mind, any more than a girl can be a teeny-weeny bit pregnant. There are, as we have just shown, such things as impaired minds both among humans and animals, but that is a matter of damage rather than degree, and it can only apply to individuals not to a whole species.

Thus far we have sought to specify the positive criteria for assigning mind, but we can also establish negative criteria for a denial of mind. What is characteristic of mindless creatures is that their behaviour is utterly rigid, inflexible and stereotypical and their sensory receptors are oriented to a narrow and highly specific range of stimuli. This is exemplified in the famous experiments carried out on frogs by Jerome Lettvin, Humberto Maturana, Warren McCulloch and Walter Pitts at MIT. The frog only reacts to a bug-like stimulus in its environment and to nothing else. Hence it cannot be said to perceive. Its behaviour in response to the stimulus of rapidly darting out its tongue can be wholly described and accounted for in Behaviourist terms of reflex conditioning. The same is true of the remarkable attack-strategy that spiders employ in their hunt for prey; but they, too, as Donald contends, have no consciousness or mind, because "the spider is too narrow, we might suggest, too rigid and too fixed in its repertoire. It seems blissfully unaware of the most significant objects in its large environment..."[14] Frogs, too, react purely automatically to a

13 P.S. Churchland, *Brain-Wise: Studies in Neurophilosophy*, (Cambridge, Mass.: Massachusetts Institute of Technology Press, 2002) 81.

14 Merlin Donald, *A Mind So Rare,* op. cit, 114.

very narrow range of sensory stimuli. They are utterly incapable of adjusting their behaviour in response to the changes in their environment, which, of course, they cannot perceive. Their learning capacities are equally limited to varieties of conditioned reflexes. Little wonder, then, that the apocryphal frog will not jump out of the pot but is slowly cooked to death.

The psychology of the frog can be fully accounted for on Behaviourist principles, but this is not true of the psychology of the rat or any other animal with a mind. Among the earliest American psychologists to point this out was Edward Tolman, even though he never spoke of mind for that would have been academically suicidal in American psychology departments, "only toward the end of his long and distinguished career did he confess to being a 'crypto phenomenologist'", as Griffin puts it.[15] In Europe, of course, psychologists had no such inhibitions in applying mind and mentalistic concepts to animals. However, it was Tolman who exposed the shortcomings of the Behaviourist model in accounting for the behaviour of higher animals starting at the very lowest with rats. As Jerome Bruner recounts much later:

> Tolman's big bang was getting rid of what he called a "telephone switch-board" theory of mind and substituting a "map room" model governed by instrumentalism of "means-ends-readiness", a kind of heuristic teleology.[16]

Bruner refers specifically to Tolman's article of 1948 "Cognitive Maps in Rats and Men" published in *The Psychology Review*. Whether one can so neatly bracket mice and men together is now questionable but at the time it seemed a step away from orthodox Behaviourism.

It is true, nevertheless, that for most lower animals the terms of Behaviourism provide an adequate scientific explanation. But even there, Behaviourism is not always enough since this approach refuses to consider what goes on in the brain and treats it as a black box. But without entering into brain neurology it is doubtful whether an explanation can be given for the location abilities of bees for which something like a "map" in the brain has to be postulated. Neurology must be invoked in all such cases where Behaviourism alone is no longer adequate. This is, of course, true for all schools of psychology, even for

15 Donald A. Griffin, *Animal Minds,* op. cit, 136.

16 Jerome Bruner, "Will Cognitive Revolutions ever Stop", in David Martel Jackson and Christine E. Eveling, eds. *The Future of the Cognitive Revolution*, (New York: Oxford University Press, 1997), 281.

those that deal with human minds, since in all cases mind/brain systems are involved.

In speaking of mind/brain systems, we are only resorting to a short-hand expression for an all-embracing complex that includes much more than what is contained in the hollow of the skull. The brain on its own is certainly the central part of the physical component of that system, but on its own it is only an organ that is inactive and impotent. It is only when it is linked to a body by means of the nervous system, and the body in turn is nestled within a physical environment through locomotion and the senses that the brain can begin to be active. This holds for all animals regardless of their particular niches, including Man; thought Man's environment is much more than a niche in the physical world, it is the whole of the known world in so far as we can know it.

Man's environment reaches beyond that of any other animal species, it goes further than the physical world and extends into the non-physical dimension of a historically constituted social, cultural and linguistic world. As we shall show, this means that the human mind also goes further than the body, even though it is inseparable and totally dependent on the body for its existence. It also means that one must not identify the mind with the body either in the way that identity-thesis exponents of Materialism advocate or the way in which certain schools of Phenomenologists come close to advocating when they identify mind with living body, that is, with *Leib* as opposed to *Körper* in the way that Husserl uses these terms.

This is true not only for the mind, but for the brain as well, for no human brain can develop independently of the non-physical objective world of society, culture and language, that which Hegel called "objective mind." This is evidenced by the fact that the infant's and child's brain can only develop in the context of the human world, for it is constituted as a human brain through socialization and enculturation, including the acquisition of language, which must take place in a window of opportunity between birth and puberty, for if that is missed then no characteristically human brain will arise. Children brought up completely outside human society do not have human brains, but something quite other for which we have no name for such cases are historically very rare. Thus the human brain is as much the product of language and culture as is the human mind, as Deacon and many others have argued.[17]

But even when a human brain has been constituted as in an adult, it is still dependent in its functioning on its connection with the body and environment without which it ceases to be properly active. Sensory deprivation experiments have shown that the brain shuts off when deprived of such stimuli

17 Terrence W. Deacon, *The Symbolic Species,* op. cit, passim.

from the outside. And even deprivation of human contact, as in long periods of solitary confinement, has a very deleterious effect both on the mind and brain for it can lead to debilitation, derangement and death.

This approach to mind and body as comprising a mind/brain system governed by the logic of emergence is what is explicitly or more usually implicitly inherent in much contemporary neuroscience and psychology. An explicit statement is that by the neurologist Antonio Damasio:

> When I say that body and brain form an indissoluble organism, I am not exaggerating. In fact, I am oversimplifying. Consider that the brain receives signals not only from the body but, in some of its sectors, from parts of itself that receive signals from the body. The organism constituted by the brain-body partnership interacts with the environment as an ensemble, the interaction being of neither the body nor the brain alone.[18]

Damasio is well aware that most such brain-body organisms do not have minds; they are not mind/brain systems. He is adamant that for mind to arise much more is called for than mere behaviour:

> But complex organisms such as ours do more than just interact, more than merely generate the spontaneous or reactive external responses known collectively as behaviour. They also generate internal responses, some of which constitute images (visual, auditory, somatosensory, and so on), which I postulate as the basis for mind… Brains can have many intervening steps in the circuit mediating between stimulus and response, and still have no mind, if they do not meet an essential condition: the ability to display images internally and order those images in a process called thought.[19]

It may well be that the capacity to form such internal images or representations or "maps" is necessary for mind, though it is a matter of dispute among neuroscientists and philosophers, but it is far from sufficient. If it were enough to constitute mind then it would be difficult to deny mind to bees, who most probably constitute some kinds of representations or cognitive "maps" of food sources which they then outwardly represent in their dances. Unfortunately, this is not enough to grant them mind for they fail such obvious tests as being

18 Antonio Damasio, *Descartes' Error: Emotion, Reason and the Human Brain*, (New York: Harper Collins, 2000), 88.

19 Ibid, 88.

able to be taught delayed reactions to stimuli or any "insightful" problem solving abilities. Thus Donald's negative conclusion about bees holds:

> There are many nervous systems of great complexity (for instance honeybees) that, by any criterion I know do not possess consciousness. Thus one cannot extrapolate directly from neural complexity or neural activity, to awareness. The physical foundations of awareness cannot lie upon some vague general principle that applies to all nervous systems, but rather upon some very specific design features of particular kinds of nervous systems.[20]

The same conclusion, we would maintain, applies to frogs. However, what are these specific design features that apply to mind? That is not yet known, neurology has still a long way to go.

Nevertheless, all neurological work points to the conclusion that neural complexity or neural activity is the essential precondition in any such design features peculiar to minds. For without an inordinate degree of complexity, one so large as not to be susceptible to surveyability, even in principle, a purely reductivist conclusion would follow that the system in question is merely mechanistic and not a mind/brain system. With a purely mechanistic or behaviouristic brain-body system, there can be no emergence of mind and the logic of emergence, that of an indissoluble non-identity, cannot be applied. This will form the nub of our argument about emergence in Chapter 5. This is the reason that we cannot ascribe an emergent mind to organisms whose behaviour can be explained solely by reference to neural processes alone or by reference to sequences of linear stimulus-response chains, namely, one dimensionally.

The scientific study of the brains of animals has now been going on for some centuries at an ever increasing rate. Tremendous progress has been made since the development of new high technology resolution means of observation and recording and since the riddle of genetic inheritance has begun to be solved by molecular biology. But there is still a long way to go even with the simplest organisms. Decades of research have by now been undertaken on one of the very simplest of all animals, C. elegans a small barely visible worm that has but 959 cells altogether; that began with Sydney Brenner and his team in Cambridge, England in the mid-1960s and has been continued in Cambridge, Massachusetts and many other places all over the world.[21] This worm has always exactly

20 Merlin Donald, *A Mind So Rare*, op. cit, 98.

21 Siddhartha Mukherjee, *The Gene: An Intimate History,* (London: Bodley Head, 2016), 191.

302 neurons arranged in the same way and around 5,000 interneuron connections. These are very few when compared with humans who have several billion neurons and trillions of connections, but even these few have not been completely elicited and work on them still goes on. Perhaps more is known about the neural net of the Ascaris worm, a parasite that has only 162 neurons designed to control the worm's wriggling movements. As for C. elegans, Torsten Wiesel admitted that "we still don't understand how C. elegans works". He allowed that to understand the brain "we need at least a century, maybe a millennium."[22] But perhaps even that is too short a time.

One of the best examples of work in unravelling the brain mechanism of a single function is Eric Kandel's work on so-called "memory" and "learning" in the sea-snail Aplysia California. As Horgan reports:

> Kandel's group showed that both habituation and sensitization produce molecular changes in neurons controlling Aplysia's withdrawal reflex. In the case of habituation, neurons discharge fewer neurotransmitting molecules into the synapse connecting them to adjoining neurons; sensitized neurons, conversely, discharge more neurotransmitter. These experiments provide evidence for a proposal, first advanced in the 1950s by Donald Hebb, that learning varies the strength of connections between neurons.[23]

The Hebbian "fire-wire" rule for the connection of neurons doubtlessly works for all brains, including human ones as well. But that is still very far from showing how these neurons link up to form neural nets, and how such nets interact to constitute behaviour, let alone mental phenomena. As Jerome Kagan remarks "…the big prize is understanding the relation between molecular and behavioural events."[24] But when that prize will be awarded is beyond our predictive powers to determine.

The presence or absence of mind is not directly proportional to the size of a creatures brain or the sheer number of neurons it possesses. This is made evident by Donald in the case of the octopus:

> It carried the invertebrate system to its ultimate expression, having somewhere in the neighbourhood of eight hundred million neurons… It has remarkable perceptual intelligence with an exquisite sense of touch and

22 Quoted in John Horgan, *The Undiscovered Mind,* (New York: Broadway Books, 1999), 20.
23 Ibid, 40.
24 Quoted in ibid, 37.

> very good vision. But it does not have very many options as far as action is concerned. It can either approach and eat or run away and hide.[25]

The reason for this great disparity between brain size as measured in number of neurons and poverty of behavioural repertoire, as Donald explains, is that the octopus "uses up to 80 percent of its central neurons to arrive at such decisions."[26] Otherwise, its "memories are triggered only passively by associative cues. Thus the creature is a slave of its environment unable to conceive of an agenda of its own."[27] On this assessment, it is doubtful whether the octopus can be ascribed a mind. But there might be strong disagreement on this score, especially from those who work with octopi.

In what follows we shall make use of the geometric analogy and speak of one dimensional, two dimensional and three dimensional brain systems. The one dimensional systems are the purely linear physical brain-body systems of mindless organisms. The two dimensional are those where another dimension above the purely linear has to be invoked to account for the observed behaviour of an organism. These are the animals, starting approximately with rats and ravens, to which we assign the emergent attributes of mind and consider them as mind/brain systems in the extended sense previously described. Such creatures must, therefore, be studied psychologically, since they have a psyche in the Aristotelian sense, as well as neurologically and behaviouristically according to the laws of Behaviourist psychology. As we shall presently see, the higher laws of Gestalt and Gibsonian psychology also apply to them. In the next sections we shall seek to show that human beings occupy a still higher dimension of mind, namely, that they are three dimensional mind/brain systems. The grounds for considering these paragons of animals to be in this way above animals and closer to the angels – as Hamlet declares, "how like an angel in apprehension" – we shall examine in the next chapter.

As we shall show, the three kinds of brain systems relate to each other in a hierarchical nesting manner, such that the higher types encapsulate within themselves much of what belongs to the lower types. So that the three dimensional brains also encompass aspects of the two and one dimensional brains. This is what makes the metaphor of dimensions apposite. It is also in keeping with evolutionary developments, for these reveal that earlier structures are never completely expunged but built on and put to different uses in a more evolved context. So it is with the evolution of the brain, where earlier forms

25 Merlin Donald, *A Mind So Rare*, op. cit, 159.

26 Ibid, 159.

27 Ibid, 159.

survive because they are incorporated into higher ones. Comparative anatomy and physiology reveals this in an unmistakeable way. However, as we shall show in the next chapter, this does not mean having to accept Paul MacLean's evolutionary model of the so-called triune brain.

The neurological theory that fits best with our account of one, two and three dimensional brain systems is Gerald Edelman's theory of neural group selection (TNGS) which he develops in a number of books of which *The Remembered Present* is the most relevant in this context.[28] In this work Edelman distinguishes between the operations of the brains of three classes of animals: those that exhibit mere adaptable behaviour without consciousness, those with primary consciousness, and those with higher order consciousness, which corresponds exactly with our three dimensions inherent in the three types of brain systems nested within each other.

As we cannot go into the complexities of brain neurology or the technicalities of TNBS, we shall merely excerpt some of Edelman's major findings, especially those that are less likely to be disputed. The theory as a whole has been subject to considerable controversy from rivals in the field; when it was first presented it was scathingly reviewed by Francis Crick and others. We cannot go into these arguments here, and given the rapidity of theoretical changes in this fast-moving field, it is possible that they are no longer relevant. It is perhaps ironic to note that Edelman's collaborator Giulio Tononi and Crick's collaborator Christof Koch have joined forces in advocating a new Integrated Information theory (IIT). We are not in a position to discuss this here; but it will be interesting to see when that theory is in its turn superseded, for no theory in this field lasts very long, and this tells us much concerning the present state of science.

The crucial turning point in Edelman's account of the evolution of the brain is the emergence of consciousness which we can identify with the origin of mind. This has enormous evolutionary survival advantages for the animals that first accomplish it, as Edelman puts it: "primary consciousness permits an animal to regulate the salience of various parts of a stimulus complex in terms of its own individual adaptive needs and, above all, to guide its actions and behaviour to reach particular goals."[29] For "without a means of developing a composite 'image', relieved to some extent of the immediate flux and variation of signals, an animal would be at the mercy of simultaneous but disparate environmental happenings."[30] Edelman's main concern as a neurologist is to

28 Gerald M. Edelman, *The Remembered Present,* op. cit.

29 Ibid, 92.

30 Ibid, 92.

account for how such an "image" is formed and thereby to locate the neurological processes of primary consciousness. The necessary and sufficient conditions for this are the following factors:

(1) perceptual categorization by selective systems in primary and secondary cortices of each modality.
(2) memory including sequential as well as temporary unitary elements.
(3) learning and
(4) a biological self-nonself distinction.[31]

How all this takes place in the brain is illustrated by the accompanying diagram in Edelman's book which we cannot explain in detail or reproduce, but must refer the reader to the original source.

Primary consciousness, which develops in what we have called a two dimensional brain system, "may be briefly described as the result of the ongoing discrimination of present perceptual categorization by a value-dominated self-nonself memory." This is what Edelman calls "present memory". Memory as recategorization is a specific dynamic process that is crucial to primary consciousness because it involves what Edelman calls a "discriminative bootstrapping – the self-nonself memory must be built of previous perceptual categorizations by a conceptual system. Only after a considerable amount of experience accumulates can detailed conceptual discrimination of current categorization occur."[32] By the term "conceptual discrimination" Edelman means something as basic as perceptual discrimination and categorizations which all sentient animals perform.

Memory as recategorization is explained in terms of generalization of an initial category discrimination which transforms it:

> Categories are not immutable but can be altered by the current state of the animal. Indeed, one of the most extraordinary properties of animals capable of perceptual categorization is their ability to generalize: after encounter with a few instances of a category under learning conditions, they can recognize a great number of related but novel instances.[33]

This results in a kind of memory that is not a fixed reproduction of stored items (as in a computer) but a constantly transforming activity. It transforms the neural processes themselves: "recall, under the influence of constantly changing contexts, changes the structure and dynamics of the neural population

31 Ibid, 97.
32 Ibid, 101.
33 Ibid, 110.

that were involved in the original categorization".[34] Even though this is a non-mechanistic process, Edelman believes he can build mechanical surrogates that to some limited degree mimic how it takes place. These are his Darwin automata that "alter their synaptic efficacies and can classify, associate, and generalize at least in a limited fashion."[35] The important thing for Edelman is that they can be made to work without a prescribed program, solely by virtue of their encounter with the environment. We shall take up the issue of mechanical models later in Chapter 7.

Higher order consciousness, which depends upon a three dimension brain system, develops out of primary consciousness. It is, of course, the consciousness characteristic of humans for one of its main preconditions is language – "a uniquely human activity in its full symbolic and syntactic manifestations, but one foreshadowed in higher primates such as chimpanzees."[36] Language with its "rich syntax and grammar is inconceivable without the ability to produce *concepts and primary consciousness*, on both of which a subject-predicate relation will ultimately be based."[37] This is as much as to say that higher order consciousness depends on primary consciousness, or in our terminology that a three dimensional brain system is based on and encapsulates a two dimensional one. In this sense, the animal mind is still there within the human mind, and in all kinds of circumstances it can revert back to it and act purely unconsciously as far as higher order consciousness is concerned.

Through having language humans become self-conscious, "they are (with the possible exception of chimpanzees) the only self-conscious animals."[38] Hence they can report directly on their own conscious states, whereas the consciousness of animals "can only be arrived at only by correlating morphology and behaviour … it also rests on the hypothesis that their behaviour is regulated by some sort of image."[39] Unfortunately Edelman provides no specificity as to which animals are conscious and which are not. Thus he provides no answer to our problem of where in the evolutionary continuum of animal species mind begins. Perhaps he considers this a problem for zoologists and ethologists and not for neurologists such as himself. Perhaps, for similar reasons, he is also not too clear as to why chimpanzees might constitute the only non-human exception to the possession of self-consciousness, which, as we shall show in the next chapter, is one of the defining features of humanity. If

34 Ibid, 110.
35 Ibid, 110.
36 Ibid, 103.
37 Ibid, 103.
38 Ibid, 22.
39 Ibid, 22.

language is necessary, then how can chimpanzees be self-conscious since they do not speak? And if language is not necessary, then why deny self-consciousness to other species of ape? And once one begins to depart from humans, where does the process of finding exceptions end?

2 Gibson and Gregory on Perception

The essential distinctions between mindless animals and those with minds, on the one hand, and between animal and human minds, on the other, can be made across the whole gamut of mental faculties, all those mental attributes apart from which it is not possible to ascribe mind to an organism. For as we argued previously, mind is a holistic concept, it either exists as a whole or not at all, but that whole is made up of mental parts, such as, just to mention a few key faculties, perception, action, motivation, cognition, emotion, imagination, memory and many more. Any one of these cannot exist on its own but only in the context of the whole syndrome of mind. Furthermore, there are intricate entailment relations between them, such as, for example, that between desire, belief, expectation and so on depending on what is desired. This logical integrity of mind is, of course, best exemplified by the human mind. Animal minds are far less integrated, but these too must display some such inner coherence or they cannot be considered minds at all.

As previously explained, "reading" the mind is a kind of hermeneutic activity: we can attribute any one mental characteristic to a body and then proceed to confirm that most of the others of the mental syndrome are there as well. If the test fails then we must withdraw the original attribution or modify it accordingly. Nevertheless, this does not make it impossible to concentrate on one faculty and provide a psychological theory for it alone without having to theorize about all the others as well, provided that it be understood that other faculties have to be also accounted for at least implicitly. This is, in fact, how most of the major schools of psychology proceed. With that proviso in mind, we shall concentrate in this section on just one of the key faculties of mind – perception.

Perception is a crucial faculty for it is what puts an organism in touch with its environment; and sight is the outstanding sense for it enables more of the environment to be perceived than any other. A mindless animal, such as a frog, has sense organs which interact with various stimuli impinging on them from the environment; a frog registers and reacts to bug-like flying objects, but not much else. Hence, it does not perceive, though in a colloquial sense it does "see". An animal with a mind does perceive for it apprehends a wide variety of objects and events specific to its habitat or *Umwelt*. A human mind is unrestricted in its perceptual capacities within the range of its sense organs. It has access, therefore, to the whole world with all its objects, properties and relations that fall within the scope of human perception. And, in such perception, the whole mind is engaged so that desires, emotions, meaningful relations, and much more of the mental gamut of faculties is involved.

This is only partially the case in animal perception since the animal mind is much more limited than the human mind. It is even less so the case for mindless creatures, which cannot even be properly spoken of as perceiving, since that is a mental act, but in common parlance nobody would deny that an animal with eyes can see. This is a mere terminological problem which should not hold us up as we will differentiate one mode of seeing from another, and come to separate mindless from mindful seeing.

For terrestrial animals sight is the most important of all the senses for it enables them to locate things at a distance, which is crucial for survival. As is well known, different eyes have evolved among a number of species quite independently of each other. The crucial question is not what kind of eyes a creature has or where they are located, which is a physiological issue, but what they can register or see, which is an ethological and psychological one, but which is not to deny that there are close connections between the two factors. Thus, for example, a frog cannot see objects not because of the structure of its eyes alone, but because of the way its whole brain is configured in a one dimensional linear fashion. As a result, its brain can only register stimuli of a very specific kind; otherwise it sees nothing of its immediate environment.

Some mindless animals register very specific and precise single features of their environment, as for example, bees that respond to colours of flowers or spiders to types of prey, but they are stereotypically fixated to those alone, and cannot shift their gaze, as it were, to anything else. Donald remarks that such animals have visual systems that "are designed to carry out certain specialized operations with great efficiency,"[40] but are otherwise completely inflexible. However, as he goes on to point out, it is different with animals of a higher evolutionary calibre:

> [Insects, fishes, reptiles and frogs] cannot move beyond this and adapt to novel situations. A frog may perceive its mate; it cannot create categories for the other players in its environment. It is inflexible. With birds and mammals the picture changes. They are flexible knowing creatures with

40 Merlin Donald, *A Mind So Rare*, op. cit, 124.

> a wide view. This introduces new criteria for awareness, flexibility and adaptability.[41]

These are, of course, also criteria of mind. We have already seen how ravens, as demonstrated by Heinrich's experiments, can see things that few other birds can, namely, how one thing, a lump of meat, is connected to a piece of string such that pulling on the latter will bring the former into reach. A rat would be able to see this as well but with some difficulty and a chimpanzee with great ease.

Chimpanzees stand at the apex of animal perceptual capacities, just below humans. Much of the animal kingdom can be arranged in a hierarchy of grades of perception of increasing complexity and abstraction, which, of course, is closely correlated with levels of intelligence. Thus as Donald explains:

> Apes can discriminate hand signs that are too complex or subtle for dogs, dogs can read aspects of behaviour missed completely by rats. Events can be arranged in a hierarchy of complexity; the simplest events are those that are closest to the level of object perception. A hand sign is an object in motion, and the perception of a hand sign or visual emblem as a unified event is well within the capacity of an ape.[42]

But, as Donald goes on to state, "even this level of event perception does not suffice for language", which points to the fact that human perception is of a still higher order, as it were, on another dimension of mind. As we shall soon see, it involves the apprehension of symbolic and other such meaningful relations between things that no animal can perceive.

According to Donald, "the perception of events is the ultimate objective of the perceptual process, at least in reasonably complex animals."[43] And he goes on to add that "apes are particularly good at visual-event perception."[44] Ever since Köhler began these investigations at the time of the First World War, a tremendous amount of research has been performed on the perceptual capacities of apes, especially chimpanzees and bonobos their pigmy cousins. A lot has been learned, but still not anywhere near enough about how and why they cannot compete with humans in perception. Clearly, symbolic skills are

---

41 Ibid, 124.

42 Merlin Donald, *Origin of the Modern Mind: Three Stages in the Evolution of Culture and Cognition* (Cambridge, Mass.: Harvard University Press, 1991), 253.

43 Ibid, 53.

44 Ibid, 154.

involved in human perception of which apes are incapable. To what extent these derive solely from language is a moot point, since they are also to be found in the non-linguistic arts such as picture making and seeing, as well as in music. Apes are, of course, incapable of such artistic creation or apprehension for they have no aesthetic sense, and this, as we shall argue, marks a crucial difference between them and humans.

Much of the work on visual perception of the Gestalt school relies on seeing pictures, drawings and diagrams from the one static point of view, as is also the case with many visual illusions. Hence, Gestalt theories of form, figure-ground and so on are much more appropriate to human rather than to animal vision. Experimental work to substantiate all this relies on subjective linguistics reports from human subjects on what is being seen and how aspects can change, rather than on behavioural indicators. Hence, much of what the Gestaltists propound as universal features of perception only hold for certain modes of human perception.

According to James Gibson, the figure-ground phenomenon is derived from seeing outline drawings on paper: "the paper surface appears to become 'background' and to recede while the enclosed form seems to take on 'figural' qualities and to stand out."[45] On the basis of this account, which he supports by reference to Kurt Koffka's work of 1935, *The Principles of Gestalt Psychology*, Gibson comes to the following conclusion:

> A corollary of this definition is that the figure-ground phenomenon has been derived from a perception of outline-form, not from a study of all forms or of all perceptions. The universality of the phenomenon as ordinarily described is therefore questionable. It is one of the most convincing tenets of Gestalt theory in its battle with elementarism but whether it will serve as the fundamental basis for a complete theory of perception is not so certain.[46]

As Gibson also points out, "the original laws of visual organization were formalized by Wertheimer (1921), on the basis of observations with dots, lines and outlines, and their implications were developed by Koffka (1935)."[47] What relevance such Gestalt laws can have for animal perception is far from clear. Thus,

45 J.J. Gibson, "What is Form?" *Psychological Review,* 1951, 58, 403–13, in Edward Reed and Rebecca Jones, eds. *Reasons for Realism: Selected Essays of J.J. Gibson*, (Hillsdale, NJ: Lawrence Erlbaum,1982), 307.

46 Ibid, 307.

47 Ibid, 304.

elsewhere he concludes that "these considerations have led me to suspect that object perception and its complement 'hole perception' are not based on figure-ground perceptions as we have supposed but are on the contrary simpler and more direct."[48]

Gibson himself develops what he calls an ecological theory of perception, which though closely related to the Gestalt theory, some authors in fact refer to it as such, yet is different from it in that it is primarily designed to account for animal perception. Hence, its emphasis on "ecology" and the relation of the animal to its environment in determining what and how an animal perceives. As we shall see, it is far less suitable for specifically human forms of perception, except, of course, in so far as humans are animals and can also perceive in the same way. To relate an animal's perceptions to its environmental niche, Gibson coins the term "affordances", which he explains as follows:

> I have suggested that the environment consists basically of substances, surfaces, places, objects and events. These perceivables have a special kind of meaning that I call affordances; animals and children learn to see what they afford *for them*. The learning is tacit not explicit, i.e. most of it is not put into words. They are perceived by looking around and getting around (ambient and ambulatory vision as described in *The Ecological Approach*, 1979).[49]

He goes on to note that "these parts of the world have not been recognized by psychologists and philosophers as what animals perceive."[50] Affordances are "ecological, in the sense that they are properties of the environment relative to an animal."[51] They are not what psychologists and philosophers assume perceptions is based on, they are not sense-data or sensations or "phenomenal qualities of subjective experience (tertiary qualities, dynamic and physiognomic properties, etc.)" or "simply the physical properties of things now conceived by physical sciences."[52] The study of all the various kinds of affordances that are either innately given or acquired through experience by animals and young children has been one of the main preoccupations of James Gibson and his wife Eleanor Gibson. She has particularly concentrated on the affordances

48 J.J. Gibson, "Information contained in Light", in Reed and Jones, op. cit, 58.
49 J.J. Gibson, "Direct Perception and Indirect Apprehension", in Reed and Jones, op. cit, 292.
50 Ibid, 292.
51 J.J. Gibson, "Affordances", in Reed and Jones, op. cit, 404.
52 Ibid, 404.

perceived by babies and young children, such as revealed in her visual-cliff experiments.

Gibson is quite clear that the perception of affordances is only given or acquired by what he calls higher animals, namely those with minds. Lower animals, those that are mindless, do not perceive, they merely register and respond to stimuli impinging on their sense organs. According to Gibson, "perception is essentially an act of attention", which clearly makes it a mental act.[53] Gibson is very insistent that the perceiver does not "contribute anything to the act of perception" but merely "picks up" the "information" that the ambient array of light makes available. This has put him at odds with most other theorists of perception, above all with Richard Gregory, as we shall see later. It has also led to accusations that he had adopted a naïve realist philosophy of perception. But this is far from the case, and is due to a deliberate misunderstanding of his purely psychological approach, as we shall seek to show.

However, it is true that Gibson eschews any representative theory of perception, or any that relies on unconscious hypotheses or expectancies or anything else of this inferential nature. Perceiving is very different from the mere receiving of stimuli and interpreting them, it is an act:

> Acts are not responses to stimuli and precepts are not responses to stimuli. An observer is not "bombarded" by stimuli. He extracts invariants from a flux of stimulation. Affordances, and the stimulus information to specify affordances, are neither subjective nor objective but transcend this dichotomy. The actor/perceiver and the environment are complementary.[54]

Hence, lower animals that merely register stimuli do not perceive, and no mind can be ascribed to them. This provides one clear criterion based on perception whether an animal has a mind or not, which can be applied under laboratory conditions and tested through experimental controls; but it can also be utilized "in the world", as it were, by means of ethological observation of the animal in its natural habitat.

According to Gibson, lower animals do not have senses in the proper sense, merely physiological organs for registering outside energies impinging on them; they have photoreceptors and not eyes, strictly speaking:

53 J.J. Gibson, "On Theories for Visual Space Perception", in Reed and Jones, op. cit, 89.
54 J.J. Gibson, "Affordances", in Reed and Jones, op. cit, 411.

> We are accustomed to say that the stimulus for an eye is light and in truth it has long been known that the stimulus for the kind of photoreceptor found in lower animals is the light falling on it. The effective limits of wavelength and intensity may be measured. Accordingly it is easy to assume that the stimulus for the eye of a higher animal is the light falling on a mosaic of photoreceptors. But this analogy is misleading, for the characteristic activity of an eye is not that of a simple photoreceptor.[55]

Gibson goes on to ask what it is that stimulates what he calls "an eye proper": "to what is it sensitive, considered as an organ instead of a mosaic of cells?"[56] His answer is his concept of "ambient light", for mere light as such "is much too simple an answer". This move opens up a new approach to vision and perception in general.

Gibson's whole approach to perception is philosophically very ambitious. He is intent on denying the basis for the kinds of indirect or representative theories of perception that have been current since Kepler and Descartes, as we shall show in Chapter 6. He aims, instead, to provide a theory of direct perception somewhat reminiscent of Aristotle, and thereby to remove the ground from under psycho-physical parallelism sensation theories (Helmholtz, Müller, et al.), from Behaviourist theories of stimulus and response (B.F. Skinner), and from Cognitivist theories of information processing and the detection of cues (David Marr, et al. at MIT). There are swingeing critiques of all of these in his work. Instead, as his friend Gunnar Johansson puts it, "the main function of ecological optics is to develop a model for how reflected light carries visually decodable information about an organism's environment."[57] But what "decodable information" means and how "ecological optics" differs from scientific optics are issues that would have to be considered at length and in detail in any extensive study of this approach, which we are unable to undertake here. However, what is clearly apparent even on a cursory view is that it must not be confused with any simple-minded naïve realism such as philosophers love to refute.

Gibson's ecological optics and the rest of his ecological theory of perception is very useful at marking the differences between the mere stimulus response of the lower animals and perception proper of the higher ones. As we have said, this is a clear criterion in establishing the difference between mindless creatures and those that have minds. It is not so useful in establishing the next

55 J.J. Gibson, "Ecological Optics", in Reed and Jones, op. cit, 61.
56 Ibid, 61.
57 Gunnar Johansson, "A Letter to Gibson", in Reed and Jones, op. cit, 77.

emergent step, namely, from animal minds to human minds. This is because his ecological theory of perception, one that relates an animal to its environment, is only relevant to humans in so far as they, too, share animal capacities, but it tells one little about how human perception proper differs from animal perception. Human perception is fundamentally different in that it involves much more than mere "affordances" and extends to the meanings of things and relations that no animal can grasp. For the human mind has capacities that are way beyond animal minds. It is true that the human mind is not distinct from and independent of the animal mind, but that in many ways it encapsulates and so incorporates the animal mind, but it also operates on another dimension, as we shall proceed to show in the next chapter.

Gibson partly obfuscates this difference by extending his theory of affordances to the perception of objects that have uniquely human characteristics and cultural meanings of which animals can have no awareness. Thus Gibson lumps together affordances such as a "tree branch (affording arboreal support for a primate)" and other such purely natural potential animal uses with objects serving uniquely human functions and utilities, such as a hammer, needle, spear or any other tools.[58] Clearly, these are not affordances in the same sense, for a hammer no more affords hammering nails or any other of its countless human uses than a car affords driving or its uses. Once we enter the human world of meaning and culture, the whole concept of "affordances" designed for an ecological context ceases to apply.

In his early work Gibson was well aware that different theories of perception are required for the fundamental differences separating the perceptions of animals (including humans qua animals) and humans in their proper roles as meaning seeking and understanding creatures. He shows this in his critique of Köhler's attempts at a unified theory of form perception, for he remarks that "the effort to determine what happens in the brain when one perceives form-in-general will prove fruitless". Instead he proposes a differential set of theories for form perception:

> At least three separate levels of theory will be required: first, a theory of how we perceive the surfaces of objects – a theory of slant-shape or, in other words, of shape-constancy; second, a theory of how we perceive representations, pictures, displays and diagrams; and third, a theory of how we apprehend symbols. There is no reason to suppose that the physiological concomitants of all these experiences will be the same; in fact, since pictures and symbols presuppose objects, their physiological

58 J.J. Gibson, "Affordances", in Reed and Jones, op. cit, 405.

> explanations will probably have to be found at increasing levels of complexity.[59]

Gibson is clearly critical of Kőhler's efforts at discerning the same kinds of electrical circuits in the brain whenever forms of any kind are perceived.

In his own later work Gibson has devoted most of his effort at establishing a theory of how we perceive the surfaces of objects. This is, indeed, his ecological theory of perception with its concept of affordances which applies so well to animal perception. Of the other two types of theories of form perception, which are uniquely human, he has much less to say. Much of what he does say about human forms of perception is contained in his occasional papers on the perception of pictures where he is engaged in aesthetic controversies with the philosopher Nelson Goodman and the art historians Ernst Gombrich and Rudolf Arnheim. He quite correctly counters the purely conventionalist accounts these thinkers are offering of what it is to see a picture, and refutes their idea that it has to be interpreted or "read" like a written script. Pictures are seen and not deciphered; representations of the pictorial kind directly elicit visual experiences of things and not symbolic readings of meanings. Unfortunately, he does not spell out what bearing this has on the perception of real things in the visible world. Had he done so, he would have realized that seeing pictures tells us a great deal about how human perception differs from animal perception.

Gibson's notes on aesthetic perception – one cannot call them a theory since they are still so undeveloped – certainly provide vital insights as to what distinguishes human from animal perception, but obviously they do not go far enough. To go further in establishing the separation between the two modes of perceiving we need to turn to rival theories that focus more directly on human perception. Such are the closely related New Look theories developed by Jerome Bruner and Leo Postman in America and the cognate approaches of Donald Broadbent and Richard Gregory in England. These theorists collaborated closely with each other and opposed the Gibsonians, just as the latter were critical of them in return.

It is not to our purpose in this contest to argue out the pros and cons of these rival approaches or to adjudicate between them or support one side as against the other, though we will soon show that the differences are not as large as they might appear. Our main aim is merely to assess what each has to offer in describing animal or human perception and thereby accounting for the fundamental differences between them. We shall contend that neither provides a completely comprehensive theory of all perception, but that both are

59 J.J. Gibson, "What is form?", in Reed and Jones, op. cit. p. 314.

necessary and complementary with each other in covering different types of perceiving. This is a little like the opposition of the Behaviourists and Gestaltists in theorizing about animal behaviour where the former are more relevant to the lower species and the latter to the higher species.

There is no doubt, however, that there are marked theoretical differences between these opposed approaches and it is not our intent to try to reconcile them, even if such a thing were possible. Nevertheless, the apparent philosophical contradictions that separate them, and that figure in the very heated polemics between their exponents, need to be countered and moderated for the differences are far more relative than absolute. And this is all that we aim to achieve.

Gibson was adamant that perception was direct and unmediated, the eyes can pick-up all that was necessary to see from the environment itself, from the "ambient array of light" and from "invariances" in the ever changing flux of visual sensations. There was no need for any inferences, representations, unconscious hypotheses or models. The New Lookers argued just the converse, they took off from the work of Hermann von Helmholtz and maintained that perception involved "unconscious inferences" and, as Gregory puts it, "human perception is but indirectly related to objects, being inferred from fragmentary and often hardly relevant data signalled by the eyes, so requiring inferences from knowledge of the world to make sense of the sensory signals."[60]

Gregory makes a point of emphasizing *human* perception as against the perception of animals, and restricts the relevance of Gibson's approach to the latter:

> The "affordance" notion might be seen as an extension of the ethologists' concept of innate "releasers", which trigger innate behaviour, such as robins responding aggressively to a red patch. This fits Gibson's "ecological optics"; but how new objects, such as telephones, are recognized without acquired knowledge is far from clear.[61]

The knowledge in question is such as only human beings can acquire. Hence, there is a sense in which animals cannot see a telephone, though they obviously can see the object which is a telephone. This brings out something of the difference between the two modes of vision. This difference is also brought out by a study of visual illusions. Gregory notes critically that "to maintain that

60 Richard L. Gregory, "Knowledge in Perception and Illusion", *Philosophical Transactions of the Royal Society* (1997), 352: 1121–8.

61 Ibid.

perception is direct, without need of inference or knowledge, Gibson generally denied the phenomenon of illusion."[62] Strictly speaking this is not true, Gibson did not deny the reality of illusions, but he did deny it any importance in a theory of perception. He did so because most illusions are easily resolved by recourse to movements of the head or the whole body; the kind of static and fixated perception illusions require is very rare and mostly does not apply to animals; and because humans who do see them generally know that they are illusions or tricks of perception, for example, that the alternate figures they see are lines or colours on a flat surface. But it is precisely this that makes them so relevant to artistic perception, such as seeing paintings and even more so drawings.

For Gregory and the other New Lookers, such illusions are the staples of their trade, for it is illusions that bring out the contribution that the mind or brain has to make to perception. The difference between what is there to be seen and what is actually seen is produced by the knowledge and unconscious inferences that the mind or brain has to supply. Gregory's main argument runs as follows:

> Following von Helmholtz's lead we may say that knowledge is necessary for vision because the retinal images are inherently ambiguous (for example, for size, shape and distance of objects), and because many properties that are vital for behaviour cannot be signalled by the eyes, such as hardness and weight, hot or cold, edible or poisonous. For von Helmholtz, ambiguities are usually resolved, and non-visual object properties inferred, from knowledge by unconscious inductive inference from what is signalled and from knowledge of the object world.[63]

The qualities that Gregory refers to are precisely such as Gibson terms "affordances" and which animals perceive. Gregory's view that what we see is determined by the image on the retina is also something that Gibson refutes. What this indicates is that Gregory has simply decided to dispense with Gibson's work.

However, it is possible to retain Gibson's findings and yet make Gregory's point by arguing that knowledge and unconscious inferences make a difference to human perception, as the example of seeing the telephone demonstrates. Over and over again the New Look experiments demonstrated that not

62 Ibid.
63 Ibid.

only knowledge of objects, but even valuations, ideals and attitudes can make a difference to human perception. As Margaret Boden reports:

> They saw perceptions (and concept learning) as being guided by conscious and unconscious *expectancies, hypotheses,* or *models*, and *tested* by more or less reliable *cues.* In short, everyday cognition was thought of as an informal version of scientific inferences or reasoning, though with social/personal factors playing a larger role than they do in science … But they're mostly learned, not innate. And they depend on a person's values as well as their beliefs.[64]

As Chris Frith, a former student of Gregory puts it: "My perception is a prediction of what ought to be out there in the world. And this prediction is constantly tested by action."[65] As we shall soon see, the question of who or what agent makes these predictions, hypotheses and test is an issue, for it conjures up the image of a homunculus in the brain, which creates problems for any such view of unconscious inferences. But for the moment we can let this pass.

Whether or not such expectancies, hypotheses, models etc., relying only on slender cues, constitute the whole of perception, they certainly can make a difference to human perception. Whether this also applies to animals is a moot point. But that they do work in humans is demonstrated by the kinds of illusions that humans are fooled by but which animals do not perceive, such as most of the ones that Gregory discusses: Muller-Lyer, Hering, Ponzo, the Penrose impossible triangle, the Kanizsa triangle, hollow face illusion, size-weight illusions, Magritte's mirror, Mach's corner and a host of others. Many of these are ambiguous Gestalt-like figures which can be seen in alternate ways according to preference. What they reveal is the kind of knowledge and pre-established experience that shapes human vision. Thus, for example, the Muller-Lyer illusion is very strong for people who live in conventional rectilinear rooms and houses and straight streets, namely city folk, but very weak for rural peasants, such as Zulus, who are everywhere surrounded by round objects, round dwellings and round exterior tracks and paths. This proves that habitual experiences and knowledge do make a difference to human seeing. How big a difference it makes is a matter of psychological investigation. Similar conclusions can be drawn from many other illusions.

64 See Margaret Boden, *Mind as Machine*, vol. 1, op. cit, 298.

65 Chris Frith, *Making up the Mind: How the Brain creates our Mental World,* (Oxford: Blackwell, 2007), 132.

The big issue to be resolved is the problem of determining what differences basic human knowledge, in the form of language and culture, makes to human perception in general as compared to animal perception where such knowledge is absent. This is crucial in determining how human mind differs from animal mind in general. This issue is at the heart of the dispute regarding perception between Gibson and Gregory. We can accept Gibson's view that animal perception is one of "affordances" which are purely "objective" ecological features which have nothing to do with unconscious inferences or background knowledge, which in humans is based on language and culture. Yet, at the same time, grant Gregory his main point that in humans such knowledge determines their mode of perceiving. There could be no human vision without language and culture acting to provide the necessary background knowledge setting. There certainly could be no human artistic vision without it. Seeing a painting depends on it, which does not mean, as so many art critics, such as Ernst Gombrich, Rudolf Arnheim and most extreme of all the philosopher Nelson Goodman have maintained, that visual artistic representations are purely conventional cyphers or arbitrary symbolic devices. Gibson is surely right in insisting that we see directly what a picture represents, we do not merely "read" it or decode it.

Another way in which art demonstrates the role of knowledge and the experienced mind in human perception is through the ability of artists to train their perceptual powers to see things differently from the way they are usually and "naturally" seen. The experienced painter can see the landscape unfolding before his gaze as a series of flat colour patches and converging perspectival lines. In other words, the artist can perceive it as a visual field, which an untrained eye cannot do, for people are very strongly habituated to see distance and depth and all the other veridical visual features that objects in fact possess; including their real colours, as against the colours in a given illumination that, say, an Impressionist painter can see. It is in this sense that everything we see can also be seen differently, either because of unusual conditions which deceive veridical perception and create an illusion or because we have trained ourselves to see differently. Mostly, however, we see what there is to be seen for we have been so habituated; namely, in the vast majority of cases we see veridically and not in an illusory way, though this is always potentially possible. Suggestions playing on imagination can also determine what is seen, as Polonius under Hamlet's instigation, sees alternatively a camel, weasel and whale in a nondescript cloud.

According to Gregory, visual illusions which he calls "cognitive illusions", are "due to misapplied knowledge employed by the brain to interpret or read sensory signals."[66] Gregory and others in the Helmholtz tradition avail themselves of a whole host of terms such as "read" and "interpret" to convey what supposedly the brain is doing; many of these are highly intellectual activities, such as infer, hypothesize, judge, model, test, validate and many more. Thus Frith expresses himself in such terms as follows:

> My brain discovers what is out there in the world by constructing models of that world. These models are not arbitrary. They are adjusted to give the best possible predictions of my sensations as I act upon the world. But I am not aware of this complex mechanism.[67]

Apart from all the well-known philosophical problems created by Frith's contention that we can never directly know the world, only models of the world in the brain, there is the added difficulty of determining who or what does this discovering and knowing: is it the "I" who acts upon the world, which would seem to be the body, or is it the brain that "discovers" and "constructs" and "adjusts" and so on? If the latter, then surely it must be a homunculus in the brain, since neurologically considered the brain is nothing but neural processes. But if it is all a "complex mechanism" of which I am not aware, then how can such a mechanism carry out the activities of a homunculus? One standard way of resolving these difficulties is to argue, as the Cognitivist psychologists do, that it is all "computations" in the brain or information processing. But in that case, there is no "I" and no "sensations" either. We shall return to this "solution" in Chapter 7 and show that it solves nothing. In the meantime we must attempt another way of resolving these problems.

If we are to avoid any suggestion of a homunculus in the brain, yet still retain Helmholtzian talk of "unconscious" inferences, hypotheses, models, etc. then we should resort to a Neo-Freudian vocabulary. Experimentalists like Gregory and Frith would be totally averse to Freud, far more so than the New Lookers like Bruner and Postman, as Boden reports:

> Bruner and Postman had posited "perceptual defence" in "normal" subjects as well as neurotics, to explain why people fail to perceive emotionally threatening stimuli. Their subjects would register at some unconscious level (measured by galvanic skin response) but conscious

66 R.L. Gregory, "Knowledge in Perception and Illusion", *Philosophical Transactions of the Royal Society,* 352 (1997), 1122.

67 Chris Frith, *Making up the Mind*, op. cit, 132.

> perception was delayed or prevented … (prompting an explosion of work on "subliminal" perceptions).[68]

Such work on subliminal perception as well as the whole idea of unconscious operations of the mind in perception prompts the view that there is a perceptual unconscious parallel to that of an appetitive-emotive unconscious promoted by Freud. Rather than any homunculus or mechanism in the brain, we should refer to the unconscious operations and activities of the mind. Perhaps there are many close parallels in the working of the perceptual and Freudian unconscious, so that, for example, the Ames illusion can be explained as arising out of a kind of "defence" mechanism according to which the unconscious mind finds it easier to accommodate the perception of people of different sizes rather than distorted rooms.

The idea that the human mind has an unconscious as well as conscious side to it is the fundamental fact that serves to distinguish it from animal mind where no such distinction is to be drawn. It obviously depends on language and culture, as psychoanalysts, such as not only Freud himself but also Jacques Lacan and many others, have sought to establish. For experimental psychologists, exploring the various facets of the unconscious has, indeed, become of late a basic task, as we shall see in the next chapter. For perceptual experimental psychologists, plumbing the perceptual unconscious has not as yet been made as explicit as it ought to be. In a way, Gregory showed the way, for if Freud declared that dreams are the royal road to the Freudian unconscious, Gregory might be taken as saying that illusions are the royal road to the perceptual unconscious.

There is no doubt that there is an unconscious dimension to human perception. But whether that is relevant to animal perception of affordances, such as Gibson studies, is what is in question. In so far as humans also perceive affordances in an animal way, they do so without reference to the unconscious factors that Gregory identifies. Reacting to a flying missile by ducking is obviously exercising that kind of perception to which Gregory's strictures do not apply. On the other hand, seeing a telephone is not perceiving an affordance in Gibson's sense.

It seems, then, that both kinds of theories are necessary to account for the full range of perception from animal to human. It is true that expressed in philosophic terms the two types of theories stand in a contradictory relation to each other, the one being a kind of naïve realism and the other

68 Margaret Boden, *Mind as Machine*, vol. 1, op. cit, 304.

representationalism. But seen in purely empirical scientific terms, apart from any philosophical interpretations, the experimental evidence for the two theories is a matter of degree: in some experimental situations the one works better and in others the other.

This difference is largely offset by the fact that each theory relies on very different experimental conditions and amasses its evidence accordingly. Gibson relies on as realistic perceptual settings as possible where the observer can move freely and take in the whole scene or context in which the specific object to be seen is situated. Gregory relies on highly restricted settings where the visual information is tightly controlled and the object is perceived from the one immobile perspectival orientation. In fact, Gregory relies for his experimental data on visual illusions where much less that the whole object of perception is present. Little wonder that they reach such opposed conclusions.

But perhaps both theories are correct in relation to their experimental conditions. When confronted with illusions, it is obviously the case that it is the brain or unconscious mind that has to interpret the few cues or visual indicators and do so by means of something approximating to "hypotheses", such as expectancies based on experience and learning from previous perceptions. On the other hand, when a huge wealth of visual data is available through head and ambulatory movement in relation to a situated object in a complete scene, the role of the brain is very different, for hypotheses no longer enter into it. Hence, to establish the relation between visual information and the brain and to assess which contributes more in different experimental settings will be the endeavour of further experimental work, and the conclusions arrived at are subject to empirical findings. No a priori philosophical issues are relevant in settling such differences between psychological theories.

As we have already indicated, the crucial area of perception where the differences between animal and human mind show up most clearly is that of aesthetic perception, such as the seeing of pictures. What a human can see in a picture is far more and very different in kind to what an animal can see, that is, if the animal can see anything at all. However, we know that many species of animals, and not just apes, but birds as well, can be trained to see various things in pictures. But what such animals see is not what humans see, even ones who have had no artistic education previously. And this is true not just of pictures but all other types of art, such as music where animals cannot appreciate such basic features as rhythm. The emergence of the human mind, since the very earliest separation of hominins from their nearest ape relatives, can be traced by reference to the origin of the arts. To do so we must turn to the palaeoanthropologists.

3 Donald and Deacon on Culture and Language

Art and the aesthetic sensibility in general is certainly a key place to look for what differentiates the human from the animal mind. This idea was first presented by Friedrich Schiller in his celebrated work *On the Aesthetic Education of Mankind*, published in 1794.[69] Basing himself on Kant's third critique, *Critique of Judgement*, Schiller went on to elaborate a view of mankind as creatures who engage in play in the aesthetic sense: "man only plays when in the full meaning of the word he is a man, and he is only completely a man when he plays."[70] Aesthetic play distinguishes the human not only from the animal, but also from what he inappropriately called the savage: "it is the same in all races who have escaped the slavery of the animal state: a delight in *appearance*, a disposition toward ornament and play."[71] Schiller's views were extremely influential and passed over into those of his friend, the philologist Wilhelm von Humboldt, as well as to Hegel. In England they exercised a profound effect on Coleridge and the Romantic poets. They were equally influential in America on the Transcendentalists and eventually on the Pragmatist philosopher Dewey. It is to Dewey that we "owe the important insight that it is primarily in the aesthetic dimension of experience that we encounter consummated human meaning", as Mark Johnson puts it.[72] Huizinga, the author on whom we first called to expound this aesthetic side of Homo Ludens, is also a late exponent of this tradition.

Of course, the human aesthetic sensibility and its propensity to artistic creativity do not emerge out of nowhere, it has its evolutionary preconditions in animals. Thus, for example, the impulse to engage in play is found in many species of mammals, especially among the young where it serves the purpose of learning by mock practice that which as mature adults they will have to do in earnest, as it were. Play serves the same ends in human children; but that is where the resemblance ends, for what children enact in games are the symbolic encounters and roles they will perform when they grow up. Animals do not engage in symbolic play, nor do they have symbolic objects such as dolls or toy cars. Animals also do not mime adult activities or mimic each other. All

69 Friedrich Schiller, *On the Aesthetic Education of Mankind,* trans. and ed. Elizabeth M. Wilkinson and L.A Willoughby, (Oxford: Clarendon Press, 1967).

70 Letter xv, op. cit.

71 Ibid.

72 Mark Johnson, *The Meaning of the Body: Aesthetics of Human Understanding* (Chicago: Chicago University Press, 2007), x.

these are the progenitors of art. But, as we know, countless millennia of play-acting separate child's play from *Hamlet*.

The author who has taken art as the index of humanity the furthest is Donald. Like Gibson, he, too, focuses on seeing pictures and drawings as a way of differentiating what animals and humans can perceive. He has also taken issue with Goodman's purely conventionalist account of seeing pictures. He goes on to ask some crucial questions:

> As Goodman [*The Language of Art* (1968)] has pointed out, the pictorial image is essentially symbolic; it is a referential device for the user of the image. But where, in the representational architecture of the oral-mythical mind, is the referential system for visual-pictorial images?[73]

The oral-mythical mind is the consummate human mind, and in that mind there are no such referential systems for decoding pictorial images. Such images are seen and not read or interpreted, or if they are interpreted then this is itself a way of seeing. Donald points out that some animals, such as the higher apes, can to a certain limited extent also see pictures, which is also indicative of their non-conventional nature:

> Some pictorial images can be accurately perceived by apes: a gorilla can respond appropriately to certain quite subtle pictorial images; these pictorial images, on one level, must possess episodic reference. Thus far we know of no animals other than apes that can interpret drawn images, that is, see them in terms of their referents; however, this question probably has not been explored as thoroughly as it might be.[74]

Donald is correct in his assessment, it is only through exploration of what various kinds of animals can or cannot see in pictures that the full extent of the differences between human and animal perception will be revealed. But as we shall see, Donald does not take full account of pictures in the making of the oral-mythic mind.

The differences in seeing pictures is a much more difficult issue to settle than making pictures, for there is little doubt that animals have not the least capacity for any kind of pictorial representation, such as even young children evince. At best, chimpanzees can be trained to splash paint on canvas, but that is about as far as their artistic talents go. To explain the reason for this inability

73 Merlin Donald, *The Origins of the Modern Mind,* op. cit, 283.
74 Ibid, 283.

we must once more turn to Donald whose theory of human evolution provides a very plausible account of what specific capacities and aptitudes separated the earliest hominins from their hominid cousins, the chimpanzees. According to his account, an evolutionary shift took place even prior to language in what he calls "culture", from an animal stage of episodic culture to a more human stage of mimetic culture, which preceded the fully human or "modern mind" stage of oral-mythic culture. His theory was first set out in his book *Origins of the Modern Mind* (1991) and later extended and adumbrated in numerous essays. Its application to art was developed in a later essay of 2006, "Art and Cognitive Evolution":

> I have argued that art is an inevitable by-product of mimesis – a primordial and truly human, cognitive adaptation that occurred very early in hominid prehistory and became the signature feature of the human mind. Mimesis had enormous cognitive consequences on the group level, resulting in a characteristically human form of communicative culture that later increased its influence with the emergence of language.[75]

Thus, according to Donald, it is not so much language, a late arrival on the evolutionary stage, but mimesis that constitutes an original pre-linguistic start to human mind, one that separates it irrevocably from animal mind. But whether art is "an inevitable by-product of mimesis" is a further matter with which we shall take issue later. He holds that "mimesis describes a cluster of capacities that were made possible by a single neuro-cognitive adaptation. They go together historically because they share certain key neural components."[76] The four central abilities that arise out of mimesis are mime, imitation, gesture and the rehearsal of skills. Animals are lacking in all these capacities, though there are some rudimentary indications of them in apes, which suggests that they evolved from these beginnings in early hominid species.

Donald holds that the evolutionary emergence of this cognitive faculty was coupled to "a redistribution of frontal-cortical influence during the early stage of the evolution of the species *Homo*, when the prefrontal and parts of the premotor cortex expanded enormously in relative size and connectivity."[77] This generated a new kind of self-aware mind which is peculiarly human:

---

75 Merlin Donald, "Art and Cognitive Evolution", ed. Mark Turner, *The Artful Mind and the Riddle of Human Creativity* (New York: Oxford University Press, 2006), 14.

76 Ibid.

77 Ibid, 15.

> The expansion of the prefrontal cortex was crucial in improving conscious self-regulation and metacognition. This created a new metacognitive field, a greatly expanded and differentiated working memory, in which hominids could observe themselves as actors and rehearse and refine whatever they were doing. This also gave them some ability to reflect on the cognitive process itself, and the option of deliberately reflecting on, and shaping, their own actions.[78]

Animals have no such reflective and reflexive powers; it is as if they concentrate completely on the objective external world and are incapable of attending to their own inner world of subjectivity. Only humans can do that.

All this inevitably raises the question of which came first, the expanded brain or these cognitive capacities? This is, of course, a chicken-and-egg question. Since a relation of emergent coupling obtains between them, it is not possible to say that the one precedes the other, for they are both aspects of the one system. There is an indissoluble bond between them, yet they are not identical for the one cannot be reduced to the other. But, as in all such cases of emergence, we can adopt different causal perspectives on this relation: from some points of view and scientific purposes it is possible to say that the new cognitive capacities cause the brain to grow in size; just as from other perspectives and in other scientific contexts we say that the enlarged brain causes new cognitive powers. We shall consider this reversibility of cause-effect relations in emergent contexts more fully and explicitly in Chapter 5, where we shall also explain how emergence involves an indissoluble non-identity between two entities.

When we look at the matter from an evolutionary point of view, the causes of the expanded brain size and its new internal structural rearrangements are to be found in the evolutionary pressures resulting from social and cultural developments in early hominin species. It is society and culture that shape the brain, as well as vice-versa. This is even more apparent once language starts to develop. If one takes an evolutionary perspective on this matter, one can explain the size, form and structure of the human brain by reference to the demands on it made by the new social and cultural practice of linguistic communication. This is, indeed, the argument propounded by Terrence Deacon:

> The evolutionary miracle is the human brain. And what makes this extraordinary is not just that a flesh and blood computer is capable of producing a phenomenon as remarkable as a human mind, but that the

78 Ibid, 16.

> changes in this organ responsible for the miracle were a direct consequences of the use of words. And I don't mean this in a figurative sense, I mean that the major structural and functional innovations that make the human brain capable of unprecedented mental feats evolved in response to the use of something as abstract and virtual as the power of words. Or, to put this miracle in simple terms, I suggest that an idea changed the brain.[79]

"An idea changed the brain" – Deacon explains what he means by this enigmatic phrase, which he insists he does "not mean in a figurative sense". It is not a piece of poetry but a down to earth scientific truth. But it does take him nearly a hundred pages of print to explain the idea behind it. He realizes that it "inverts our common sense notion of causality that physical changes required physical causes."[80] But, actually, this is only a very doctrinaire scientistic assumption that any serious consideration of emergence can disabuse. In an emergence coupling, cause and effect can easily be inverted. In evolutionary emergence this happens continually. Thus, as Deacon explains, "the first use of symbolic reference by some distant ancestors changes how natural selection processes have affected hominid brain evolution ever since. So in a very real sense I mean that the physical changes that make us human are the incarnations, so to speak, of the process of using words."[81]

To make this case, Deacon appeals to a variant of the Darwinian theory of natural selection called "Baldwinian evolution", after the American psychologist Mark Baldwin who first proposed it in 1895. According to this theory: "by temporarily adjusting behaviours or physiological responses during its lifespan, in response to novel conditions, an animal could produce irreversible changes in the adaptive context of future generations."[82] Deacon applies this approach to what he calls "the co-evolution of language and the brain", the subtitle of his book. He argues that "once symbolic communication became even slightly elaborated in early hominid societies, its unique use for representational functions and open-ended flexibility would have led to its use for innumerable purposes with equally powerful reproductive consequences."[83] These consequences played themselves out in human bodies in evolutionary changes that produced what he calls "the architecture of modern human

79 Terrence W. Deacon, *The Symbolic Species,* op. cit, 322.
80 Ibid, 322.
81 Ibid, 322.
82 Ibid, 322–3.
83 Ibid, 349.

brains, and it is the key to understanding the subsequent evolution of an array of unprecedented adaptations for language."[84] Like Baron Munchhausen, language and the brain pull each other up to ever greater evolutionary heights by pulling on each other's bootstraps. Undoubtedly, there must have been some such evolutionary process, but what it actually was like is now purely speculative since all relevant evidence from missing species is lacking; and, as Deacon himself admits, "evolution is a historical process, it is like entropy in the sense that it cannot be analysed backwards…"[85]

Speculations regarding the evolutionary origin of language and how this relates to that of the modern mind, as Donald calls it, have abounded. They change drastically every generation or so, as new technological instruments of detection are applied; as new approaches deriving from other areas of science come into play; and as new discoveries are made which completely throw out all previous evolutionary continuities because wholly new intermediate species are found, or our view of the known ones has to be revised. In the last decade or so incredible new discoveries have put the whole field of palaeoanthropology into turmoil and put in question many of our previous assumptions about the origin of our own species and its close relatives now extinct. The latest evidence shows that Neanderthals shared the human propensity for symbolic art and that they made outline pictures on cave walls in Europe at least 65,000 years ago, which is 20,000 years before our species arrived there. This seems to indicate that a propensity to art might have already been present in an ancestor of both Neanderthals and homo sapiens as early as half a million years ago. Such facts, if substantiated, open up speculative conclusions of various kinds, including ones like those Donald hazards on the score of human prehistory.

Donald's approach to the problem of the origin of the human mind is the converse of that of Deacon and all other palaeoanthropologists who focus on language as the key component. According to Donald, language is a latecomer to the evolutionary game. Donald makes the bold and unprecedented move that culture takes precedence over language. Instead of arguing, as others do, that culture derives from language, he maintains, to the contrary, that language derives from culture:

> The great divide in human evolution was not language but the formation of cognitive communities in the first place. Symbolic cognition could not spontaneously self-generate until those communities were a reality. This

84 Ibid, 346–50.
85 Ibid, 351.

> reverses the standard order of succession, placing cultural evolution first and language second… This makes the mechanism of language highly dependent on culture. This may be said of all symbolic processing.[86]

Accordingly, there must have been a pre-linguistic culture among early hominins, for without that evolutionary background language could not have arisen. He calls this archaic culture mimetic.

Mimetic culture is what we have already encountered in Donald's account of the origins of art. As he puts it:

> In short, art is the expressive culmination of the most ancient domain of the human mind, as manifested in rituals, public actions, and gestures that characterize any human society. It is woven into the deepest layer of meaning that can be called uniquely human… Viewed in an evolutionary context, art originated in the earliest stages of hominid evolution, the so-called Mimetic phase. Newer forms have been scaffolded onto older ones, and as human beings have evolved complex languages and technologies, artists have developed new forms that contain within them all the elements of evolutionary history.[87]

Art thus acts as a kind of Grand Canyon of the evolutionary history of the human mind in which archaeologists can discover ever deeper strata of humanoid species going right down to the original rock foundations, the animal mind of a creature like the chimpanzees. The main aim is to explore the intermediate layers, the so-called missing gap of evolutionary pre-history of the mind. Donald believes he can span that abyss by unearthing a so-called mimetic culture, to which task he devotes a large part of his first book, *Origins of the Modern Mind*.[88] As he puts it in his second book *A Mind so Rare*:

> What we are trying to do is establish a set of logical intermediate stages, or pre-adaptations that could have bridged the huge evolutionary distances between the minds of modern humans and those of our distant ancestors. Mimesis is my best guess at the nature of the first step… Mimetic expression can be shown to function independently of language in children and certain people with language disabilities. People who lack language because they were born deaf and never learned to sign can

86 Merlin Donald, *A Mind So Rare*, op. cit, 254.
87 Merlin Donald, "Art and Cognitive Evolution", op. cit, 19.
88 Merlin Donald, *The Origins of the Modern Mind,* op. cit.

> nevertheless manage the purely mimetic aspects of human culture with ease.[89]

There is no doubt that the evolution of mimetic abilities such as Donald describes was a crucial component of the origin of the human mind. But doubt still remains as to whether this constituted a stage prior to any language whatever or whether it was the accompaniment of an early stage of language commonly known as a protolanguage.

Donald's arguments from the capacities for people who are handicapped, such as those born deaf and dumb, who can nevertheless function in society is not altogether convincing about the course that evolution took. That people with genetically "modern" human brains can learn and adapt to a human society in the ways that animals, even the most intelligent chimpanzees, cannot do is no proof that there must have been a speechless mimetic stage between animal culture and human culture which is based on language. The capacity of handicapped individuals to function with the support of normal people in a social context does not mean that such people could constitute a society of their own or that they reflect some earlier cultural stage in which genetically human beings did not yet exist. A kind of "Planet of the Apes" scenario cannot be constituted with deaf and dumb people. So the question still remains of whether an early form of language must have been there from the very start.

We are now in a better position to frame the fundamental opposition between two anthropological approaches to the emergence of human mind from animal mind, that of Donald versus that of Deacon, which we referred to earlier. Deacon takes the more conventional approach and holds that the evolution of language is the key to the emergence of mind and the growth and restructuration of the brain. Donald, as we have shown, adopts the heterodox view that it is not language but what he calls culture that is crucial to the emergence of what he calls the modern mind. There is much at stake in this confrontation, which unfortunately the two authors do not actually engage in explicitly. In their extant publications there is barely a reference of the one in the other: Donald makes a few passing mentions of Deacon's neurological ideas; Deacon makes no reference to Donald at all, despite his commendation of Donald's first book. Since they do not engage directly with each other, the debate between them has to be constructed as it does not exist in print.

According to Deacon the key to understanding language and the human mind is symbolic reference. He puts it unequivocally that "theories of language and mind that fail to address this issue head on, or suggest that it needs no

89 Merlin Donald, *A Mind So Rare*, op. cit, 268.

explanation, ultimately assume what they set out to explain."[90] Obviously Donald's theory and some others fall into this category. All such theories cannot account for what "makes human mind different from nonhuman minds."[91] For Deacon, it is symbolic reference, not syntax as for Chomsky, nor culture as for Donald, which is fundamental to the evolution of both mind and brain. The evolution of the brain was a consequence and not a cause of the evolution of language.

According to Deacon, the evolution of language began with simple languages or protolanguages: "though simple languages exist in no society found today, they almost certainly existed at some point in our prehistory."[92] However, once the symbolic process began, once "despite their cognitive limitations, our ancestors found a way to create and reproduce a simple system of symbols", then the evolution of the brain followed suit, as "the brains that originally struggle to support simple languages were replaced by brains better suited to this awkward adaptation."[93] This "inversion of cause and effect has enormous consequences … [for] our unique mentality must also be understood in these terms."[94] And "the implications for brain evolution are also profound", since the evolution of language "created selection pressures to reshape our lineage's ape brains to fit this new function."[95] This fits in well with our main thesis that cause and effect relations in emergent phenomena can always be inverted for different purposes in different scientific contexts, as we shall go on to argue in Chapter 5.

Having thus briefly outlined Deacon's main thesis of where the essential differences between human and animal mind is to be found, we turn to Donald's contrary point of view. According to Donald, the key transition from animal mind to human is that from the episodic culture of apes to the mimetic culture of species of early homo. Language is, apparently, not involved at all since mimetic culture is completely pre-linguistic. Episodic culture spans the whole gamut of ape abilities, including the potential of chimpanzees, such as the celebrated Kanzi, to use symbols, as Donald states: "his symbol use was more spontaneous than that found in any previous trained ape…"[96] Thus, according to Donald, Kanzi and presumably other chimpanzees as well are capable of

90 Terrence W. Deacon, *The Symbolic Species*, op. cit, 43.
91 Ibid, 44.
92 Ibid, 44.
93 Ibid, 45.
94 Ibid, 45.
95 Ibid, 45.
96 Merlin Donald, *Origins of the Modern Mind*, op. cit, 136.

grasping the trick of "symbolic reference" that Deacon claims is the distinguishing mark of the human mind, yet they still fall far short:

> Despite the brilliant efforts of modern primate researchers, apes still use symbols for traditional primate agenda. If symbols alone had been enough, many apes would presumably be much further down the human road by now, moving rapidly toward collective representational systems, like the denizens of *Planet of the Apes*. But they are not. While individuals have made great advances, collectively they have never been inclined to construct their own symbolic culture, not even on the small scale of small working or family groups… This negates any notion that symbols have any immediate transformative power to generate culture.[97]

But it also negates something of Donald's own arguments in his earlier book that the performance of individual apes, under human tuition, is any indication of what apes can achieve collectively or what might be the next stage in their evolutionary development.

Donald's main argument as to what separates apes and humans might, nevertheless, be correct, namely, that there might be a stage prior to language which he calls mimetic, and that apes are incapable of reaching it:

> But perhaps the most important conclusion to be drawn from the neuropsychological literature is that human intelligence without language has properties that set it apart from ape intelligence, just as Darwin predicted. Among the uniquely human capacities found in the absence of language are a capacity for spontaneous gesture and mime, which can be returned to after language loss; toolmaking and praxis in general; emotional expression and social intelligence, including an ability to comprehend complex events and remember roles, customs and appropriate behaviour.[98]

According to Donald, it is not language which is the source of culture, but culture which is the source of language. Hence, language cannot exist where the above cultural capacities are lacking. As he goes on to argue:

> We cannot dismiss the possibility that language itself, especially some of its most esoteric semantic and grammatical features, might be just

97 Merlin Donald, *A Mind So Rare*, op. cit, 204.
98 Merlin Donald, *Origin of the Modern Mind,* op. cit, 93.

> another product of culture, drawn by our expanded level-3 conscious capacity.[99]

Consequently, Donald denies any animal access to language, since no animal is capable of the kind of consciousness necessary to give it the relevant cultural capacity. This holds even for such an intelligent animal as Kanzi, trained in the use of symbols by his keeper Savage-Rumbaugh, because, as he states, "it will become clear that Kanzi remains several steps removed from human linguistic ability", but he does so on other grounds than any such as stipulated by Deacon. Clearly they disagree on what is essential and fundamental to human language.

According to Donald, apes, even the cleverest of them, are incapable of language because they are still stuck in the evolutionary stage of episodic culture, that of animal mind. They are capable of event perception, such as that of seeing "a passing car … a kick, or a threatening grimace, or the lifting of a spear, or a hand sign"[100]; "yet this level of event perception obviously does not suffice for language."[101] We are reminded here of J.J. Gibson's theory of the perception of "affordances" which also does not suffice for language and cannot account for human perceptual abilities. Episodic mind, that of animals, is limited:

> …the limitations of episodic culture are in the realm of representation. Animals excel at situational analysis and recall but cannot re-present a situation to reflect on it, either individually or collectively… The cognitive evolution of human culture is, on one level, largely the story of the development of various semantic representational systems.[102]

Donald goes on to quote David Premack who "argued that language should not necessarily be seen as *the* human adaptation. Consciousness, pedagogy, social attribution, and aesthetics are also uniquely human."[103] Clearly, this runs contrary to all of Deacon's later work which is focused on languages.

Deacon, in common with other palaeolinguists, propounds the thesis that prior to our fully developed language there must have been an earlier stage in the evolution of language. This is usually designated as a protolanguage, a grammarless speech like that of a two year old child. This is frequently conceived of

99 Merlin Donald, *A Mind So Rare*, op. cit, 204.
100 Merlin Donald, *Origins of the Modern Mind*, op. cit, 153.
101 Ibid, 154.
102 Ibid, 160.
103 Ibid, 161.

as a distinct evolutionary stage in human development. Chomsky and his followers, such as Steven Pinker, would deny that this is language in its proper sense since that requires a further evolutionary departure in the origin of Universal Grammar through a language acquisition module in the brain.[104] Thus they would most probably deny David Reich's assertion that Neanderthals "would, most probably, have had language, culture and sophisticated behaviour."[105] Deacon would undoubtedly agree with this.

Donald's attitude to the evolution of language is very different from that of the above thinkers and has itself gone through considerable variation. At an earlier stage he was more in accord with the common view, shared by Deacon and many others, that there was a protolanguage as a distinct evolutionary stage, but in his late work he has resiled from this view quite sharply. His initial position closer to the received view is set out as follows:

> The earliest human languages would have gone through a "protolinguistic" phase, and the form of the protolanguage might well have resembled Bickerton's protolinguistic, grammarless speech acts, but it seems to have arrived much later than he suggested, originating during the past 300,000 years, rather than a million years earlier, and *only after* the principle of voluntary self-cued memory retrieval had been formally established in the human brain. Grammatical invention appears to be just another product of a general capacity for lexical invention but there is some neurophysiological evidence supporting Bickerton's view that grammatical invention was the "second phase" of language emergence, requiring different brain adaptation from lexical invention.[106]

However, in his later work he goes against any such idea that grammatical invention required any "different brain adaptation from lexical invention." He is particularly averse to Chomskian notions of Universal Grammar or any such linguistic acquisition module in the brain. He has called this the phlogiston of the late twentieth century. Instead he has strongly favoured a socio-cultural view of the whole of linguistic development, as he puts it:

104 See Steven Pinker, *The Language Instinct: The New Science of Language and the Mind,* (New York: Penguin Books, 1994).

105 Interview with Robin McKie, *The Observer,* April 8, 2018.

106 Merlin Donald, "Human Cognitive Evolution: What We Were, What We Are Becoming", *Social Research*, vol. 60, no. 1 (Spring 1993), 157.

> The integration of morphological addresses into a larger descriptive system is an inherently social activity, and one is tempted to predict that this process could not be confined within the isolated brain: that is, one should not expect to find the "language acquisition device" that Chomsky predicted entirely inside the individual brain. Rather, the emergence of language depends on a community of brains in interaction. There is not yet any viable computational model of this process, and neural network models have not yet reached the point where anything so complex could be simulated.[107]

Whether such "network models" will ever reach that point is a matter we leave till Chapter 7.

According to Donald, the sole strictly biological precondition for language was the adaptation that generated mimetic culture, that which separated the hominin species from their nearest hominid relatives, the chimpanzees. As he puts it:

> Mimetic skill was necessary for the later evolution of language because the evolution of mimetic skill made it possible to tamper with the morphology of action… To invent a primitive lexicon, even in the halting, grammarless manner of a two-year-old, one must be capable of a considerable degree of mimetic invention and refinement.[108]

This view of the origin of language is firmly asserted in a late paper of 2017 where he states that "language is, in this sense, not a feature of the brain, per se… It is the child of an interactive cultural imagination, that is, of groups of brains in collusion."[109] In other words, language is purely a social product.

Thus there is no special language module in the brain; the evolution of language did not have to wait for any such thing to arise. It was an extremely long term development of over 3 million years which depended on two preconditions: "the first was the emergence of a general supra-modal capacity to rehearse and refine skills, and the second was the emergence of material culture as a major force shaping the direction of human cognitive evolution."[110] It

107 Merlin Donald, "Preconditions for the evolution of protolanguage", in Michael Corballis and Stephen E.G. Lea, *The Descent of Man: Psychological Perspectives on Human Evolution*, (Oxford: Oxford University Press, 1999), 150.

108 Ibid, 148.

109 Merlin Donald, "Key Cognitive Preconditions for the Evolution of Language", *Psychonomic Bulletin and Review*, vol. 24, no. 1, February 2017, 204.

110 Ibid, 204.

all kicked off "in primate motor learning capacity early in the Pleistocene epoch that enabled the deliberate modification of procedural memories."[111] It led to what Donald calls "externalizing memory" brought about by means of material culture. The idea that material objects of human manufacture externalize memory is a conceptual innovation on Donald's part of a more radical nature than he might be perhaps aware, and it will cause him difficulties later in relation to his account of human culture. But at this stage it is not necessary to question it. What is crucial in his theory of language and of mind in general is that "the direction of evolution could shift gradually away from evolving fully equipped self-contained individuals, toward interactive and social capacities that would ensure the successful operation of the social network in a distributed cognitive system".[112]

However, it is odd, in view of Donald's emphasis on mimetic culture as a precondition for the evolution of language, that he has paid so little attention to the one form of material culture that is fully symbolic and cognate with language – the making of pictures. It points to a general bias that Donald shares with almost all other palaeoanthropologists in favour of language as the uniquely essential factor in the constitution of what he calls the "modern" mind. Thus he calls the second cultural stage in the formation of this mind linguistic-mythic and defines it with almost no reference to symbolic objects or imagery. When he does take cognisance of graphic symbols or pictures it is only as the precursors to writing, which he sees as the basis for literacy and theory, the third of his stages in the development of mind.

This bias towards language as the defining characteristic of the human mind is very widespread and is fundamentally inherent in Western thinking from the very start. We must tackle it first before we can proceed any further with Donald's account of the evolution of the mind, which is the most complex, complete and far-reaching theory that we currently possess. However, it suffers from some shortcomings which we will attempt to address, beginning with language.

4 An Appreciation and Critique of Donald

It has been a fundamental and age-old assumption, at least in the West, that what distinguishes Man from beast is the possession of language. Man has been called the speaking animal, or as Aristotle put it, the *zoon logon echon*.

111 Ibid, 205.
112 Ibid, 205.

Logos, the Word, was the prime concept of order in the flux of the world at least since Heraclitus. This privileging of language became a basic bias of Western philosophy and later of science as well.

It was the same in the complementary Hebraic tradition that Christianity took over, which also maintained that in the beginning was the Word. According to the Bible, God created the world by means of His Word: "And God said, Let there be light: and there was light". Adam, made in God's image, did something similar when he took possession of the creatures by naming them: "and whatever Adam called every living creature, that was the name thereof. And Adam gave names to all cattle, and the fowl of the air, and to every beast in the field." Theologians and poets have been speculating ever since as to what was this original Adamic language, supposedly the most perfect language of all where word and thing fit perfectly. Philosophers in the Jewish tradition, such as Philo of Alexandria, made the bold Platonic move of identifying God's Word with the Greek logos, thereby linking the two traditions of language. The early Christians followed suit; St John, the presumed author of the fourth Gospel, began it with the memorable line "In the beginning was the Logos"; thereby announcing the new Logos as the Word of God in pronouncing a renewed creation with the coming of Christ.

Modern philosophy and science from Descartes onwards have gone along similar lines in seeing language as the essential difference between Man and beast or mind and machine. Anthropologists have not escaped this fixation on language, considering it alone to be the defining characteristic of the human mind. Thus much effort in experimental psychology has been expended in demonstrating that animals cannot master language-like communication. Even though it has been shown by ethologists, such as David Premack and Sue Savage-Rumbough, that chimpanzees can be taught to communicate by means of gestures and tokens, yet this is disallowed as proof of language mastery since they cannot go on to learn the trick of grammar, to combine elementary symbols into sentences that convey propositional meaning. Chomsky and his followers have been particularly intent on turning grammar into a shibboleth for the absolute separation of human and animal minds. This is one of their motives in postulating a Universal Grammar language acquisition module in the human brain. And even though critics, such as Deacon, have denounced this on sound evolutionary grounds, yet they, too, have adhered to the primacy of language and postulated instead the idea of a protolanguage as the initial evolutionary departure of humanity.

It is to his great credit that Donald escapes from this trap of language by postulating instead his pre-linguistic mimetic stage of evolutionary development which separates the original hominin culture from the episodic culture of the animals, including our closest relatives the chimpanzees. This has been an extraordinary departure in evolutionary theorizing in that it turned palaeoanthropology away from its habitual preoccupation with stone tools and other such practical material remains and turned it back to culture, that is, to cognitive issues to do with the development of mind. However, Donald has not altogether broken free of the linguistic bias for he returns to language as the sole criterion of his next stage after the mimetic, that of the linguistic and mythic. In doing so he has tended to overlook the kinds of symbolic material remains that attest to advanced culture and mind, namely, the signs, graphic arts and symbolic objects of prehistoric humans and even earlier hominin species such as the Neanderthals which we shall presently consider.

However, before we do so, we must consider another reason for the prevalent preference in popular, philosophic and evolutionary thinking for language as that which alone makes for the crucial difference that constitutes the human mind. It is an obvious observation that children begin to speak almost before they learn to do anything else. Even when infants are too weak and immature to stand and walk upright and can barely handle objects, they are already capable of uttering sounds and communicating. Much of the effort in child rearing is in teaching them to speak in ever more complex ways; utilizing peculiarly human methods of teaching by means of instruction which we shall consider in the next chapter. These pedagogic techniques are successful in almost all cases except for autistic children, or perhaps geniuses, such as the young Einstein, who according to his mother did not speak till he was four years of age.

It is very tempting and hard to resist transferring this sequence of learning in children to the earliest humans, that is, of assuming that evolution goes through analogous stages to those of child maturation. This is the old fallacy that phylogeny recapitulates ontogeny first enunciated by Ernst Haeckel in the late nineteenth century and never quite banished ever since. That it is mistaken is quite evident from the simple fact that babies cannot do very much except use their voice – to cry, wail, babble and eventually speak – whereas the ancestral hominins could already manufacture tools and kindle fires perhaps even before they could utter intelligible sounds; though evidence concerning the origins of vocal capacities is extremely equivocal, since apart from DNA there is little in skeletal remains to go by.

Evidence for the evolution of language is scant and mainly inferential, but evidence for the evolution of art is direct and by now almost overwhelming. This need not necessarily lead to the conclusion that art came before language, as Donald is inclined to imply, but rather that art, in the full and proper sense, is at least as old as language. We now have clear evidence that graphic symbols of all kinds and even pictorial representations go back to the Neanderthals of at least 64,000 years ago, about 20,000 years earlier than any such human remains. What kind of language these Neanderthals possessed is still largely unknown. But they were already artists of some distinction, as the recently discovered paintings from the caves at La Pasiega, Maltravieso and Ardales in different widely scattered parts of Spain unequivocally attest. Dirk Hoffmann and his other co-discoverers are certain that these are the work of Neanderthals; and Hoffmann is quoted as saying that "Neanderthals and early humans were cognitively indistinguishable."[113] Such depictions of animals as were found in these caves come relatively late in Neanderthal history, but even as early as 180,000 years ago there is evidence that they were already capable of symbolic material expression, such as the stone circles found deep within the Brunique cave in France.

When Donald wrote his major works little was as yet known about Neanderthal symbolism and nothing about their art; however, there was already a wealth of knowledge about human prehistoric art. Yet Donald makes almost nothing of this in his account of the linguistic-mythic culture, the first stage of fully human development and the clear manifestation of human mind. If he does refer to pictorial representation, it is only as the preparatory prelude of his next stage of human development, that of literacy and theory; graphic symbols and depictions only prepare the ground for the later much more culturally significant origins of writing, according to Donald. In a paper of 1993 he puts it as follows:

> A partial list of devices mastered by humans along the way to full symbolic literacy includes (in rough historical order) iconography, maps, emblems, totems, pictorial representations, pictographs, sequence-markers like knotted cords or prayer beads, various types of tokens, currencies, property markers, ideographic writing systems, counting systems, mathematical notations, schematic and geometric diagrams, lists, syllabaries and alphabets, scrolls, books … computing languages, and a variety of modern multimedia storage devices that employ virtually all of the above.[114]

113 Report by Kiona N. Smith, *Ars Technica*, 23 Feb. 2018.

114 Merlin Donald, "Human Cognitive Evolution: What We Were, What We Are Becoming", *Social Research*, vol. 60, no. 1, (Spring 1993),160–73.

It is not clear what he means by iconography in this context, but he cannot be referring to art since he places it next to maps in his order of historical succession and so presumably earlier than pictorial representations which are placed later in his list. Maps, as we know, were a late historical development which does not predate the early civilizations. Similar listings in more or less the same order in his later works do not alter this impression that he has not taken count of prehistoric art, which consequently does not figure in his account of linguistic-mythic culture.

Where Donald does consider pictorial art he holds it to pertain to the stage prior to language, the purely mimetic stage. As he states, "I have argued that art is an inevitable by-product of mimesis – a primordial and truly human, cognitive adaptation that occurred very early in hominin prehistory and became the signature feature of a human mind."[115] On this view, pictorial art and graphic symbolism came prior to language and is not a coeval symbolic form that arose coextensive with language. But if the latter is in fact the case, then art is certainly not an "inevitable by-product of mimesis". It is true, as he puts it elsewhere, that "mimetic capacity is one of the basic building blocks of language, and must have preceded its evolution in hominid prehistory."[116] But this is equally true of symbolic pictorial art, for which mimesis was also only the preparatory evolutionary stage.

It seems that Donald is mistaken in treating art together with simple mimetic activities, such as primitive imitative performances and expressive gestures, as the one "mimetic dimension of culture … [which] extends beyond art, to encompass other non-verbal aspects of cultural life, including such things as public spectacle, athletic events, body language, and the non-verbal transmission of skills."[117] Pictorial art might be non-verbal, but it is certainly symbolic though in a purely visual way where meaning is seen. In cognitive terms, it is way beyond the kinds of bodily skills and performances, such as "athletic events, body language and the non-verbal transmission of skills", with which Donald brackets it together. Cognitively considered, art is fully on par with language, and just as crucial for Donald's linguistic-mythic stage, for without it religion and the higher cultural activities inherent in myths are inconceivable, as we shall presently show.

115 Merlin Donald, "Art and Cognitive Evolution", op. cit, 14.

116 Merlin Donald, "The roots of art and religion in ancient material culture", in Colin Renfrew and Iain Morley, eds. *Becoming Human, Innovation in Prehistoric Material and Spiritual Culture,* (Cambridge: Cambridge University Press, 2009), 96.

117 Ibid, 96.

As a corrective to Donald, we might designate this specifically human stage of evolution as linguistic-graphic-mythic; and venture the evolutionary hypothesis, for which there is overwhelming evidence, that language was not alone in constituting the fully human mind, but that art, too, in the form of pictorial imagery and other graphic symbolism, was equally indispensable. Human meaning and communication evolved in parallel along two lines: through vocal utterances as speech to constitute language, and through graphic markings and symbolic objects to form art and material culture in general. Both evolved together in close symbiosis, the one supporting the other and the two ratcheting each other up to ever higher stages. What these stages were and how this took place, we can now reconstruct purely conjecturally, since almost nothing of the development of language is now known; though we might infer it partially from the development of art, which in graphic material form has survived, and with which language was closely linked throughout the whole evolutionary process.

Thus we might surmise that myth – such as the narratives of origin which Australian Aborigines assign to Dreamtime – could not have arisen and developed except through the conjoined confluence of both language and art. Language alone can only rely on memory to pass on stories from one mind to another and from one generation to its successors. But memory is very variable, labile, and distorting; extraordinary techniques of mnemonic art had to be invented much later to enable texts to be reasonably accurately passes on by oral means alone, as for example the Vedic hymns in India. The effort necessary to learn off by heart extended stories is prodigious; and prior to writing it is frequently done by distributing parts of a complex myth to different clans or moieties, as is the case among Australian Aborigines.

If for no other reason but to serve as *aides mémoire*, symbolic objects such as graphic signs, sculptures and paintings, play an essential role in the rituals and ceremonies whereby myths are enacted and learned, and thereby conserved and maintained relatively unchanged over the generations. How this takes place varies from culture to culture. Among Australian Aborigines such objects feature in corroborees and are specially made for the occasion and often discarded soon after; though we know from surviving cave art and other glyphs that the techniques of making more permanent art and the practices associated with it were current until very recent times. We can assume that prehistoric cave art in Europe was utilized for tens of thousands of years in some such way as enduring records in the transmission of what might be called primitive religion, and that myths were faithfully conserved by these means. This would mean that what Donald calls "external memory" is nothing new, but as art goes back to the origin of language itself, and is, perhaps, coextensive with what he calls “internal memory”. Whether these are, in fact, two forms of *memory* we leave for later consideration.

Language and art are both means of communication which function in different ways for correlative purposes. Language works by voice and ear and generally serves for the rapid transmissions essential for practical purposes, as in giving orders, making demands, asking and answering questions, and many other such speech acts. Art works through hand and eye by the slow and laborious processes of marking – by means of scraping, carving, applying pigments and so on – which once made can be very rapidly perceived at a glance and their meaning grasped. Marking in all its forms serves to communicate ideas that are intended to endure within the one group, or to convey messages from one group to another. Cultural values and religious beliefs are relayed and conserved in this way.

Speaking and marking, language and art, convey meaning in quite different ways. It is a fundamental semiotic error to conceive of the one as if it was like the other: to take art as language or language as art. The former misconception is common among aestheticians and art critics, as we have already noted in our critique of the conventionalist reading of art by Goodman, Gombrich and Arnheim. According to these thinkers, art is like a text, it is to be decoded and read according to rules of syntax and semantics. What they fail to realize is that meaning in pictorial art inheres in perception not in conception.

The opposite error of thinking of language as like art is to be found among both poets and philosophers. The old Latin saw *ut pictura poesis* – derived from Simonides of Kos and quoted in Horace’s *Ars Poetica* – has played havoc in aesthetic theory throughout the ages, especially in the contemporary period of Modernist poetry. At all times it has had its defenders as well as critics; among the latter Lessing was one of the most notable. We side with Lessing in arguing that it has led to confusion in our understanding of the arts. A similar confused idea was utilized by Wittgenstein in his *Tractatus Logico-Philosophicus* where he presented a theory of linguistic meaning as picturing, taking the *Satz*, a propositional sentence, as a logical depiction of a *Tatsache*, a fact. Much philosophical ink has been spilled in both elucidating and refuting this idea.

Given that there are two such opposed but parallel systems of meaning making that together govern human culture and constitute the human mind, it is somewhat odd that this has been but rarely recognized by palaeoanthropologists. One possible exception is André Leroi-Gourhan, and as an extended quotation reveals, he came close to realizing the inherent interrelation between language and art:

> Throughout animal evolution, especially that of the higher species, the motor nervous system of the face and that of the hand are closely

> coordinated – a coordination indispensable for feeding. In the case of man, this coordination attains to a high degree, that of verbal expression; the facial muscles are so organized that both sign language and articulate language become possible. The links between language and the hand remain close, and down to our day gestures unconsciously accompany words. But the gesture does not raise to the height of language until it becomes drawing and writing: then thought can avail itself of two vectors, the facial organs and the hand; it rediscovers, in a sublimated form, the initial biological balance. Man certainly acquired language long before the Upper Palaeolithic period, but the development of tools most likely went hand in hand with the increasing capacity for abstract thinking and the beginning of graphic skills.[118]

Whatever the merits of Leroi-Gourhan's physiological account, which might be questionable, it is certainly the case that tool making capacity is also the beginning of the graphic arts, starting with the most elementary scratching and markings. Later during the Upper Palaeolithic both graphic abstract symbols and pictorial images are present together, as Leroi-Gourhan points out:

> From the aesthetic and philosophical point of view, the existence of a system of symbolic representations in the Upper Palaeolithic is highly instructive…This system of abbreviated representation is not a secondary phenomenon of Palaeolithic art: the symbolic signs make their appearance at the outset, and the earliest were still in use during the last period.[119]

Leroi-Gourhan's book provides a comprehensive listing of the huge variety of such semi-abstract and fully abstract signs and symbols, whose meanings are now largely inscrutable; though he makes a determined effort to decipher them according to what he takes to be the "Palaeolithic religion of Man". As to what this is about, he admits himself baffled:" I found myself in the end confronted with a system of unexpected complexity – the skeleton of a religious thought, as impervious to my understanding, as a comparative iconography of sixty cathedrals would be to a Martian".[120]

However, it is in such symbolic signs that we find the very first propensity towards writing, which marks the next major stage in Donald's evolutionary

118 André Leroi-Gourhan, *Treasures of Prehistoric Art,* trans. Norbert Guterman, (New York: Harry N. Abrams, 1966), 43.

119 Ibid, 80.

120 Ibid, 144.

sequence of the human mind, that of literacy and theory construction. Writing is a compound system for the integration of speech and graphic signs that first arose in the early civilizations. It constitutes a higher stage in the coalescence of language and art in that fixed signs or pictograms come to represent words or ideas or other linguistic meanings. In the first place, it appears as ideographic script, such as the Egyptian hieroglyphs. Its advantage over spoken language is that it can be made permanent, like art; its advantage over art is that it can convey propositional sentences and thereby convey whole narratives, like language.

Alphabetic writing was the next stage in the confluence between language and art. Invented by the Semitic people of the ancient Near East and perfected by the Greeks, the alphabet could link totally abstract signs to single phonemic sounds in any language whatever. This freed writing from being tied to the words and ideas of any specific language or culture and enabled it to be transferred from any one culture and language to any other; this eventually made universal communication across the whole globe possible. Few cultures or languages have been able to do without alphabetic script; the only major exception has been China and its close neighbour Japan, where Chinese characters arrived together with Chinese culture but not the language itself. But it has proved to be next to impossible to universalize those characters so as to make them function independently of Chinese culture; they are firmly rooted in their original cultural context. Not so with the alphabet, for its letters could be varied and adapted to suit different languages and cultures across the whole globe.

To what extent the alphabet was essential in the great transition towards the higher civilizations known as the Axial Age (roughly from 700 to 300 BC) is a much debated issue dealt with at length in our previous works.[121] That it was not absolutely indispensable is demonstrated by the history of China. However, China has always been to some degree handicapped in not having an easy to learn script, and writing was confined to a narrow elite stratum of scholar-gentlemen or *shih*. The alphabet promotes a much higher degree of literacy and a much greater range of texts; in that limited sense it democratizes culture and enables books to circulate more widely.

This is relevant to Donald's questionable contention that literacy together with theory constitute a new stage in human cognitive development, one

121 See Harry Redner, *A New Science of Representation: Towards a Unified Theory of Representation in Science, Politics and Art,* (Boulder, Co.: Westview Press, 1994); and *Beyond Civilization: Society, Culture and the Individual in the Age of Globalization*, (New Brunswick, NJ: Transaction Publishers, 2013).

which is as transformative of the human mind as the two previous evolutionary changes which separate the human from the animal mind, namely, mimesis and language. Donald's view is pre-eminently focussed on the Axial Age developments in Greece, where the introduction of the alphabet and more widespread literacy undoubtedly made possible the rise of theoretic thought in the form of metaphysical philosophy and early science. Donald has little to say about the simultaneous but independent Axial Age revolutions in Israel, India or China and, to a more limited extent, Persia. His focus on Greece means that he has to grant that in the short period between the still mythic age of Homer and the theoretic age of Aristotle a fundamental transformation of the human mind occurred, brought about not by evolution but by culture alone, a seemingly most unconvincing assumption, as Donald himself suspects:

> We are equipped with basically the same brain we had 50,000 years ago, and it might be argued that the shift to external memory was purely cultural, and therefore not as fundamental a change as the two previous ones. However, using the *same criteria* used to evaluate earlier major cognitive changes, recent changes constitute strong evidence for a third major breakthrough in our cognitive evolution. Like the two previous ones, the physical medium as well as the functional architecture of human memory have changed, and new kinds of representation have become possible. External symbols have radically changed the medium of storage. In this they constitute a real hardware change in memory, albeit a change in technological rather than in biological hardware.[122]

We might allow that what occurred in Greece during the Axial Age was, indeed, exceptional, it has rightly been called the "Greek miracle", but was it all that different from the Axial Age cultural changes that took place elsewhere? And was it that much more momentous than comparable transformations that occurred both before and after the Axial Age, such as the Neolithic Revolution, the rise of civilizations, or the onset of the age of Modernity in Western Europe from the sixteenth century onwards? In what follows we shall argue as against Donald, that the Neolithic Revolution and perhaps also Modernity, as the prelude to the present stage of technological globalization, have even better claims to be considered fundamentally new stages in human culture, and in that sense transformative of the human mind.

122 Merlin Donald, "Human Cognitive Evolution: What we Were, What we Are Becoming", op. cit, 161.

However, there is an even more fundamental objection against Donald's contention of a third major cognitive transformation of the human mind, the theoretic stage; it is that the criteria he has used to evaluate earlier evolutionary changes do not apply to such purely historical developments as those previously referred to. Evolution and history move along different time lines. They are, in fact, incommensurate with each other and there are no common criteria or standards along which they can be compared. Donald holds that memory is one such standard; but, as we shall argue in what follows, he has invoked a purely metaphoric conception of "external memory" which is highly questionable in a scientific context. This is, of course, also a philosophical issue and not a purely empirical one, and it shows once again that philosophy has a crucial role to play in science; just as science, in this case palaeoanthropology, is indispensable for philosophy of mind.

But apart from the philosophic case for the incommensurability of evolution and history, it can be shown on purely historical grounds that the move to theory, first achieved in Greek philosophy, was not as crucial a transformation in culture as the earlier one of the Neolithic Revolution; or perhaps even that which is being brought about by our present technological age of globalization. Starting in the Middle East around 13000 years ago and repeating itself independently around the world in other continents involving different varieties of crops and animals, the Neolithic Revolution has a much better claim to be considered a major turning point in the history of mankind than Greek philosophy. This original historical departure transformed scattered bands of hunter gatherers into settled populations of farmer and herders, which in turn rapidly gave rise to cities. Until very recently up to 95 per cent of the human race on the planet remained as peasants not all that much removed from the Neolithic state. The cultural manifestations of this revolution were also remarkable in that it gave rise to organized religion based on monuments and temple architecture; and in due course also to a priesthood, whereas previously there had only been shamans or witch doctors.

The Neolithic Revolution was a much more decisive break through in development than anything that came after for it fundamentally altered Man's relation to Nature. The distinguished palaeoarchaeologist Ofer Bar-Yosef considers it "the most crucial revolution of humankind after 2.5 million years of cultural evolution."[123]. The Neolithic Revolution led to Man taking control of Nature and thereby entering into History. Hunter-gatherers are in thrall to Nature in that they can only hunt or gather what their environment provides. They have

123 Ofer Bar-Yosef, "PPNB Interaction Sphere", Review Feature, *Cambridge Archaeological Journal,* 11:1 (2001), 117.

very little capacity to change the natural environment for their own benefit; perhaps just a little, as the Australian Aborigines demonstrated in using a technique of fire setting to create open grazing lands for the kangaroos that they hunted. By contrast, the Neolithic farmers and herders transformed Nature completely by their own actions to their own advantage by selectively cultivating new crops and breeding domesticated species of animals.

This need not be seen in Marxist terms as a purely material development, as many of the early palaeoanthropologists, such as V. Gordon Childe, were prone to see it. It was equally revolutionary in transforming the mentality of the people involved and in that limited sense also the human mind. In fact, there is some indication that it began as a cultural change before it assumed its transformative material outcomes. According to the latest finds at Gőbekli Tepe in Southern Turkey, dated to around 9600 BC, there was a temple-complex there before any villages and land cultivation. According to Klaus Schmidt, one of the archaeologists on site, "the construction of a massive temple by a group of foragers is evidence that organized religion could have come *before* the rise of agriculture and other aspects of civilization".[124] Together with temple architecture there was also a change in symbolic pictorial representations pointing to a new stage in human social consciousness.

These findings are fully in accord with Jacques Cauvin's thesis that a cognitive revolution preceded an economic one in the constitution of the Neolithic stage of human development.[125] Cauvin calls it expressly a "revolution of symbols." It is to this fundamental change in mentality that we owe organized religion based on the divine personifications we call gods. This differs from the shamanistic religions of the hunter-gatherers based on spirits and changes of state from one animal form to another. According to Cauvin, the first divinities were an earth-fertility Goddess and her bull son-consort. It began as a "change in collective psychology which must have preceded and engendered all the others in the matter of the process of neolithization".[126] There is an obvious symbolic relation between agriculture, domestication and the woman-bull cult in that it represents virile Man taking possession of the earth and its products. This marked the start of a Neolithic way of life which persisted with but few adaptations for most people till almost now. Later historical developments, even such crucial ones as the invention of writing and the rise of philosophy in Greece and elsewhere, concerned only small elites in a few civilizations.

124 Quoted in Harry Redner, *Beyond Civilization* op. cit, p. 34.

125 Jacques Cauvin, *The Birth of the Gods and the Origins of Agriculture,* trans. Trevor Watkins (Cambridge: Cambridge University Press, 2000).

126 Ibid, 23.

This was so till the present stage of globalization when the effects of science and technology are ubiquitous and affect all people.

Donald pays little heed to the Neolithic Revolution which he scarcely mentions, and neither do any of the palaeoanthropologists who discuss his work, with the partial exception of Colin Renfrew. Renfrew does not question Donald's periodization in the evolution of the human mind but charges him with omitting a crucial intermediate stage between the linguistic-mythic and the literate-theoretic: "while noting the origins of visuographic invention in the pictorial representation of the Upper Palaeolithic cave art, his attention then shifts directly to early writing systems in Mesopotamia, and the only systems of External Symbolic Storage to which he gives careful consideration are writing systems."[127] Renfrew goes on to propose an intermediate stage between Donald's two major stages, and locates it in the Neolithic Revolution: "it may be a valid approximation to suggest that the third transition here may often be equated with the so-called 'Neolithic revolution' of the Old World."[128]

Renfrew is surely right in highlighting the importance of the Neolithic Revolution, and the reasons he gives for its significance are also the correct ones: it introduces "External Symbolic Storage employing symbolic material culture, characteristic of early agrarian societies with permanent settlements, monuments and valuables."[129] However, what he fails to note is that this development of an unprecedented new mode of production, the very first truly human one, and its accompanying cultural manifestations – such as Renfrew notes "in the field of religion, where the distinction made between deities" first arises – casts in the shade such limited advances as the invention of writing coming with the early civilizations or philosophy with the Greeks. Once the Neolithic changes occurred, it almost seems as if all that followed was more or less inevitable. Once reliable means of subsistence were developed, there were no longer any limits to the size of human population. Hence, as populations grew, cities became an inescapable consequence; and where ecological conditions were favourable, to intensive agriculture, as in river valleys, so, too, were civilizations. Writing then emerged as a simple practical necessity in order to keep accounts of transactions in produce and other goods; and from there it was extended to take stock of everything else in a society. First wisdom literature and then philosophy were the next steps in the growth of literacy, which at first

127 Colin Renfrew, "Mind and Matter: Cognitive Archaeology and External Symbolic Storage", in Colin Renfrew and Chris Scarre, eds. *Cognition and Material Culture: The Archaeology of Symbols*, (Cambridge: McDonald Institute Monographs, 198), 2.

128 Ibid, 4.

129 Ibid, 4.

and for a long time were hardly transformative of the human condition or the mind. It is only when philosophy becomes modern science and technology, as in our time, that a new stage in human existence might be in the offing.

Renfrew does not take issue with Donald's periodization of the stages of culture and mind, merely adding as a qualification to it an intermediate stage. And neither does he take issue with Donald's concept of "external memory" on which Donald bases the importance of writing as constituting another kind of memory in contrast to what he calls "internal memory". Renfrew follows Donald in speaking of "External Symbolic Storage", which he also tries to link to memory:

> It should be born in mind, in the present context, that monuments are built for remembrance. They are often memorials. It is the role of a memorial to serve the memory, often the collective memory... All this is very relevant to our central theme of 'external symbolic storage'.[130]

Thus Renfrew does not question Donald's concept of "external memory", he merely backdates it from writing to the monuments characteristic of the Neolithic age. But one might go on to ask, if monuments qualify as "external memory", why not the much earlier cave art; and if that qualifies, why not the very first pictorial outlines and graphic symbols, which, as recent discoveries have shown are already in evidence among the Neanderthals? Are not all these graphic stages forms of "external symbolic storage"? If so, then "external memory" might go back to the very origins of humanity and might be almost as old as "internal memory".

If there is something odd with this line of reasoning in taking "external memory" to be primordial, this is because there is something odd with Donald's concept of "external memory". What is questionable about it is pointed to by the philosopher E.J. Lowe, another contributor to the volume in which Renfrew's article features, when he notes its similarities and obvious derivation from computer jargon:

> ...we should be extremely wary of Donald's heady talk of "the emergence of visual symbolism and external memory as major factors in cognitive architecture" (Donald 1991, 17). The notion of "external memory" and "cognitive architecture" that are being invoked here are ones drawn from

130 Op. cit, 5.

> computer science and the related computational conception of the mind.[131]

He points out that the computer notion of "information storage" is the obvious model for Donald's "external symbolic storage". According to Lowe, this is a mere metaphor which should not be taken too seriously: "precisely because a page of a book, like the hard disk of a computer, 'store information' in a sense which is radically different from the sense in which human beings possess knowledge or beliefs…"[132] Furthermore, human groups can possess knowledge or beliefs in common, and in that sense collective knowledge or beliefs, in a way that they cannot be said to possess collective memories. Memory is an individualized psychological concept, like perception or thought or feeling, which does not readily transfer to groups, unlike knowledge and belief which are far more interpersonal, and can be shared by many people at once. Mathematical knowledge or religious beliefs have little do to with memory; and though group memorials or practices can be called "collective memories" this term can hardly be applied to geometric theorems or religious dogmas.

However, according to Donald, once they are set down in writing geometric theorems and religious dogmas become "external memory", which he also calls "collective memory". It is a nice poetic conceit to regard all the books of mankind as its "collective memory"; and perhaps to add to this store of memory all the customs, traditions, rites and everything else of culture passed down from generation to generation. But this is an expanded concept of "memory" which is far removed from memory as this is understood in its psychological sense, where it pertains to what individuals are able to recall. Why is it that Donald feels the need to extend the concept of "memory" way past its logically accepted bounds in coining the expression "external memory"?

Partly because he is inspired by the computer metaphor, to which he is otherwise utterly opposed as a model for the mind, but much more so because he needs a contrasting term to what he calls "internal memory", which is what is ordinarily meant by memory. Donald is quite correct and theoretically inspired in proposing that the evolution of voluntary memory recall is the key to the distinctively human capacities of mind. Donald makes this clear in a late and very important paper in which he states his case at the very start:

> Human memory is special in many of its superficial manifestations, but, from an evolutionary standpoint, its major distinguishing feature is the

131 E.J. Lowe, "Personal Experience and Belief", in Colin Renfrew and Chris Scarre, op. cit, 95.
132 Ibid. p. 94.

> ease with which its contents are made accessible to consciousness. The evolution of voluntary conscious access to memory may account for many, if not most, of the distinguishing characteristics of human memory. The unique capacity for accessing our memory banks provided a platform on which the spiralling co-evolution of human cognition and culture could be constructed.[133]

By contrast, as he goes on to state, "animal memory lacks any capacity for *self-initiated* recall… Passive episodic memories – that is, detailed involuntary memories of specific events in their past lives – are found in many animal species … [but only] humans have an option to initiate a search-and-retrieval process, whereas nonhuman animals do not."[134]

Having thus established autobiographical memory or "internal memory" as the crucial evolutionary transition underlying the first mimetic stage in human development, Donald feels impelled to postulate something of analogous magnitude, an "external memory", in order to be able to assert that there is a development that occurred much later – though this time purely cultural and not biological – that underlies a transition of equal importance, that from language and mythical culture to literacy and theoretic culture. It is this preconceived need to postulate two such stages in human evolution and to regard writing and theory as of equal standing to language and myth that has led Donald to postulate two forms of memory, internal and external. On this view, just as the origins of language are to be sought in the evolution of "internal" or "autobiographical" memory, so the origins of writing are to be sought in the development of something he calls "external memory". But once it is apparent what role it is meant to serve, external memory becomes very questionable; for the two stages on which Donald's theory is based, language and myth versus writing and theory, are not parallel as he suggests. And, furthermore, memory cannot serve as a common criterion on which they can be compared.

In the 1998 paper in the Renfrew edited volume Donald makes it amply apparent that, whatever its precursors in earlier graphic signs, it is writing that is the preeminent carrier of external memory. "The transition from preliterate to symbolically literate societies … has been marked by a radically new development: the externalization of memory storage. External memory (as opposed to internal or 'biological memory') involves completely new memory media with

133 Merlin Donald, "Evolutionary origins of autobiographical memory: a retrieval hypothesis", in Dorthe Berntsen and David C. Rubin, *Understanding Autobiographical Memory: Theories and Approaches*, (Cambridge: Cambridge University Press), 269.

134 Ibid, 275.

properties that are fundamentally different from those of biological memory."[135] Thus, external memory is compared to internal memory along a range of common parameters, which Donald sets out in a table with two columns, the one headed engrams for internal memory the other exograms for external memory, in a terminology borrowed from Lashley. The two types of memory are compared along nine parameters, starting with media, "fixed physiological media versus virtually unlimited physical media", and continuing with format, permanence, capacity, retrieval paths, perceptual access, organizational modality, working area of memory and accuracy of retrieval. Thus, we are given to understand that books are memory storage devices that serve analogous functions to brains. Donald explicitly invokes the computer analogy to explain what he means by this table of comparisons, as he puts it: "if we are speaking of computers, we would have no difficulty accepting that a system that would use storage properties listed [under] external memory would have radically different capacities from a system limited to those of internal or biological memory."[136] This might be so if these were two types of computer systems, but not when the one is a biological system whose workings are largely unknown, so that even to speak of engrams is a questionable theoretical supposition, and the other is a completely known and transparent human symbolic artefact.

But Donald requires books and brains to be placed on a common footing, supposedly that of memory, for this alone enables him to claim that the cognitive revolution brought about by "external memory" as script is as great and significant as the biological evolution brought about by voluntarily accessible memory in early species of hominins that gave rise to mimetic culture, the major departure from animal to human mind. If the concept of "external memory" is shown not to be memory at all but simply knowledge storage, this would make Donald's case far less persuasive. As we have previously argued, the Neolithic Revolution has a far better claim to such a role than literacy or even theory. But our fundamental philosophical contention is that such far removed processes are incommensurable, that biological evolution cannot be compared to cultural revolution for there is no common basis or standard according to which they can be compared and that memory does not provide this.

However, it is possible that in the far distant future a fundamental transformation of human nature will take place, such that we will no longer remain human beings in an individual sense, and that this metamorphosis will be comparable to our original emergence as individual minds. In a very recent

135 Merlin Donald," Hominid Enculturation and Cognitive Evolution", in Colin Renfrew and Chris Scarre, *Cognition and Material Culture*, op. cit, 15.

136 Ibid, 15.

paper Donald has in fact raised this possibility in speaking of a fourth major transition that is at hand:

> Cognitive evolution seems to be moving from the personal to the social level, indeed, to a level of hybridized social networking with smart machines. This situation constitutes a very basic change in our traditional cognitive governance … just as basic a change as the previous ten-thousand year transition period (the third transition mediated by a series of exographic revolutions)…[137]

Donald's fourth transition in the evolution of the human mind, if it ever took place, would correspond to a sixth stage of emergence in our terms, which means that the title of this work would have to be amended to *Sextessence of Dust*. However, at this point of time it seems a too speculative lunge into the unforeseeable future. One might be sceptical about such predictions preferring instead to exercise *epochḗ* in the ancient sense of suspending judgement. Moreover, it is not clear whether what is being promised is a transcendence of the human mind or its annihilation. Donald himself wavers on this point:

> Moreover, smart technology might well bypass the previous constraint imposed by the natural limits of the brain, leaving humans out of the loop… If smart technology becomes much more autonomous and human beings lose what limited control they have enjoyed over their cognitive and social governance, the locus of control will move to the machine domain.[138]

Once more, to be or not to be, that is the question, this time not for the individual alone but for our species as a whole. Will we knowingly embark on a course of development the end of which might well be our collective suicide?

Whatever one makes of Donald's last quest in forecasting the future of the human mind, there is no doubt that his previous expeditions into its past have yielded some of the most remarkable insights that the science of palaeoanthropology has so far afforded. He must be given credit for redirecting the whole course of this science away from its standard preoccupation with

137 Merlin Donald, "Self-Programming and Self-Domestication of the Human Species: Are We Approaching a Fourth Transition?" in Anders Klostergaard Petersen, Gilhus, I.S., Martin, L.A., Jensen, J.S. and Sorensen, J., eds. *Evolution, Cognition and the History of Religion: a New Synthesis*, (Leiden: Brill, 2019), 170.

138 Ibid, 171.

bones and stones towards culture and cognition. Unlike the exponents of the new cognitive sciences, he is fully aware that cognition cannot be separated from culture, namely, that the human mind cannot be conceived of apart from a social cultural context and that this holds true not merely for ontogenesis but also for phylogenesis as well. And alternatively, culture cannot be conceived of apart from cognition, namely, that stages of cultural development are also stages in the evolution of the mind.

Even though he is not as preoccupied with material remains as most palaeoanthropologists, and perhaps less so with pictures than he might be, Donald does not resile from all the other material or social aspects of human evolution. He asserts unequivocally that "a satisfactory theory of human cognitive evolution should begin with this realization: cognition did not evolve in isolation of these other dimensions [viz. anatomy, diet, mating behaviour etc.]."[139] In his own comprehensive theory "the central thrust was both social and cognitive. Human beings evolved a cognitive survival strategy that gradually made the species better able to perform cognitive work in groups…"[140] In developing such a multi-factored theory – for as he puts it, "there is no variable that can be said to constitute the irreducible cognitive 'atom' of a uniquely human mental substance"[141] – Donald draws on his extensive knowledge and expertise in a remarkable range of sciences and humanities. Perhaps it is this breadth of mind more than anything else that makes his approach so outstanding.

Having completed our account of the origin of the human mind and how it differs from animal mind, we can now proceed to study its inner workings Thus we shift gear from having been largely intent on anthropology and evolution to being mainly concerned with psychology and pedagogy. In this exploration of the human psyche our initial guide is Freud, the founder of the aptly named science of psychoanalysis, but our analyses of the psyche will also go beyond Freud to other reaches of the mind with which he was less concerned.

---

139 Merlin Donald, "Mimesis Theory Re-examined, Twenty Years After the Fact" in Gary Hatfield and Holly Pittman, eds. *Evolution of Mind, Brain and Culture*, (Philadelphia: Penn Museum International Research Conferences, 2013), 170.

140 Ibid, 172–3.

141 Ibid, 173.

CHAPTER 4

On the Architectonics of Mind

1 Freud and the New Unconscious

Shortly before his death in 1938, Freud summed up the major problem facing any science and philosophy of mind:

> We know two things about what we call our psyche (or mental life): firstly, its bodily organ and scene of action, the brain (or nervous system) and, on the other hand, our acts of consciousness, which are immediate data, and cannot be any further explained by any sort of description. Everything which lies in between is unknown to us, and the data do not include any direct relation between these two terminal points of our knowledge. If it existed, it would at most afford an exact localization of the process of consciousness and would give us no further help toward understanding them.[1]

As John Horgan, who quotes this statement, points out, Freud was extremely well qualified to make it. Before he turned to psychoanalysis, Freud spent the first ten years of his medical career as what is now called a neuroscientist. According to Horgan, "he published more than three hundred papers and five books on neurobiology, including a monograph on aphasia and other conditions resulting from neural trauma."[2] As for his knowledge of the data of consciousness, his subsequent work as a psychoanalyst bears ample proof.

Freud is correct, the data of consciousness is primarily what we know of ourselves. Secondarily, we can correlate that with observation of what transpires in our bodies, especially in the brain and nervous system. These two forms of knowledge underlie the two major areas of science and philosophy. The data of consciousness is the subject matter of philosophical phenomenology and all those psychologies that focus on conscious experiences, above all those that study thought, perception, emotion and memory. As for knowledge of the body, that is a rich field of numerous medical and biological sciences of which neurology is the prominent instance. The problem Freud raises as to the relation between these two opposed sciences has been a major source

1 Sigmund Freud, quoted in John Horgan, *The Undiscovered Mind*, op. cit, 46.
2 Ibid, 45.

of contention and disputation for many centuries. It still constitutes a focus for research at present, as the attempt to establish a "Naturalistic Phenomenology" currently under way in Paris bears witness. We shall return to it in Chapter 6.

Like Husserl before him, Freud implicitly contends that no such "Naturalistic Phenomenology" is possible, because "the data do not include any direct relation between these two terminal points of our knowledge, [and] if it existed, it would at most afford us an exact location of the processes of consciousness and would give us no further help toward understanding them". Indeed, since Freud's day tremendous advances have been made in determining the brain location of the processes of consciousness. Thanks to our new technological instruments of detection and recording, we can now observe the operations of the brain both overall in areas of gross activity as well as in specific neural detail, sometimes localized down to single neurons. Nevertheless, Freud is right; we still do not understand how these brain processes give rise to conscious experience. The view, which Freud and Husserl seem to share, that the two phenomena are fundamentally opposed to each other, points to the original provenance of their approaches in the philosophy of Descartes. An unbridgeable gap between thought and extension, that is, between mind and matter, is what Descartes first established; and this idea has continued throughout the history of all subsequent Dualistic approaches to which both Freud and Husserl, in their quite different ways, ultimately subscribe.

In his dichotomy of the two types of data to which we have direct access, consciousness and neurological readings, Freud has left out a crucial *tertium quid* – the observation of behaviour. We are aware of and can also observe in various ways our own behaviour, but most crucially that of other people. It is through their behaviour that we come to know their minds, as we learn to infer or intuit from that their intentions, desires, feelings, beliefs, thoughts and so on. Whether this is a quasi-inductive inference or some much more instinctive and empathetic communion between one mind and another is not an issue we need to debate at this point. This holds both for relations between humans and humans and animals. In all cases communication takes place through the medium of behaviour, whether this be gestures and expressions or speech, as is mostly prevalent among humans. By means of their verbal utterances we can gain an insight into the content of other peoples' consciousness, which is not possible with animals.

One does not have to be a Behaviourist to grasp the crucial importance of behaviour for the mind. Merleau-Ponty, who argues strenuously against Behaviourism, has entitled his book *The Structure of Behaviour*; though the French term *comportment* he uses is somewhat wider in meaning than the English "behaviour".[3] It constitutes the central third term in his work by means of which he is able to break down the binary opposition of mind and body traditional in philosophy since Descartes and still evidenced in Freud's duality of consciousness and physiology. Instead of a two term binary opposition, he introduced a third relational factor that he insists belongs wholly neither to the mental side nor the physical, but instead serves to mediate between them. He puts this expressly at the very start of his book in his critical stand against certain unnamed philosophers and psychologists:

> Thus among contemporary thinkers in France, there exists side by side a philosophy, on the one hand, which makes of nature an objective unity constituted vis-ā-vis consciousness, and, on the other, sciences which treat organism and consciousness as two orders of reality and, in their reciprocal relation, as "effect" and as "cause".[4]

Merleau-Ponty might have had Gestalt psychology in mind in the first category, and he would have included Freud's psychology and possibly also Sartre's philosophy of the *en soi* and the *pour soi* in the second.

Instead of either of these opposed approaches involving a two term relation, he introduces his own three term one of mind, body and behaviour:

> We will come to these questions by starting "from below" and by an analysis of the notion of behaviour. This notion seems important to us because, taken in itself, it is neutral with respect to the classical distinction between "mental" and "physiological" and thus can give us the opportunity of defining them anew.[5]

By starting his analysis of the mind-body relation with this "neutral" term behaviour, he believes he can escape the impasse in which Freud landed himself, having to juxtapose consciousness against physiology with no way of bridging this classical Cartesian gap. Merleau-Ponty thinks he has found a way of approaching the problem of consciousness without reifying it into a substantial reality: "by going through behaviour, however, one gains at least in being able to introduce consciousness, not as a psychological reality as a cause, but as structure."[6] What he means by "structure", which is synonymous with his other

3 Maurice Merleau-Ponty, *The Structure of Behaviour*, op. cit.
4 Ibid, 4.
5 Ibid, 4.
6 Ibid, 5.

term “form”, we have already partly examined in our discussion in Chapter 1. However, his assertion that consciousness is “not a psychological reality as a cause” is very much in keeping with our later account of mental phenomena in Chapter 5 where such emergent entities are taken neither as causes nor as effects of physiological processes. Whether Merleau-Ponty’s term “structure” is the right word for this depends on what it is taken to mean, which is a difficult issue of interpretation that we cannot pursue any further here.

In respect of the human mind and its self-consciousness, behaviour means something quite other than what it means in relation to animal mind, as we have already established in the previous chapter. Primarily it means action, a concept which is by no means reserved only for humans for some animals also act, but which gains especial saliency in relation to the human mind because of the symbolic import it carries. Human acts are meaningful in the way that no animal behaviour can be and to which animal acts can only vaguely approximate. This difference is, of course, largely due to the symbolic systems inherent in language and culture, which are themselves realized through special symbolic expressive means, such as speech acts, symbolic gestures, signs, tools and pictorial objects etc. that play such a crucial role in cultural life. By these means, human beings not only communicate with each other but come to know each other’s minds and the contents of each other’s consciousness. Hence, without behaviour in this sense, consciousness would be inaccessible, and really could not exist. For humans, mind and symbolic behaviour are interdependent notions that mutually imply each other because they are logically intertwined.

In the previous chapter we examined at some length the kind of behaviour that is paradigmatically human largely by reference to the views of Donald. In general terms, it is what might be called play behaviour, that which is characteristic of *Homo Ludens* for which Hamlet is our metaphoric symbol. This is the kind of behaviour that makes creativity, imagination, the arts and every kind of aesthetic appreciation possible. In language it is what underlies figures of speech or tropes that since Aristotle have been studied under the general rubrics of rhetoric and poetics but which extend far beyond this narrow domain in accounting for how new meanings are created. Tropes also play a large role in psychoanalytic interpretation in the processes that Freud called symbolization, condensation and displacement.

It thus goes without saying that in his clinical practice Freud was extremely well aware of symbolic behaviour since he took careful note of what his patients did and what they said, for it was the primary material for psychoanalytic interpretation. Aberrant behaviour, such as obsessive compulsive disorders and parataxes of speech, were obvious symptoms of unconscious processes. But beyond that, it was listening to the patient's speech that constituted the main focus of the psychoanalytic technique. The interpreting of dreams was particularly important since, as Freud declared, dreams are the royal road to the unconscious. Thus there is an odd discrepancy between what Freud pronounced as a philosopher in the Dualist tradition and what he took for granted about the role of behaviour in disclosing the mind in his practice as therapist.

As we saw from his pronouncement quoted at the start, Freud did not discount or neglect the body, but he regarded it only as a neurologist might, as merely the material correlate of consciousness. Something similar is true of Husserl who took careful note of the body but tended to discount behaviour. Merleau-Ponty sought to remedy this omission. In the phenomenological tradition the body has remained a very fruitful field of investigation, especially at the hands of Merleau-Ponty's now numerous followers, such as Varela, Thompson and Mark Johnson. Yet it is still the case that their approach does not take full account of bodily behaviour and the means of communication that it affords: initially among animals as mere signals and cues, later among humans as gestures and articulated speech. In the latter case, this brings in the whole dimension of language and culture which is crucial for the constitution of the human mind. It also affords a way of connecting the disparate data of consciousness and neurology, for communicative behaviour, above all speech, is governed both by conscious acts of mind in so far as it is mental and at the same time by neurological processes in so far as it is physical. By means of it, we begin to glimpse a way of spanning the huge divide that separates these two dimensions.

The human mind is not, however, just two dimensional, as we showed animal mind to be in the previous chapter. We owe it largely to Freud that a third dimension of the mind has been opened up and explored – that of the unconscious. Freud, of course, did not discover or – as some critics would like to put it – "invent" the unconscious. It had already featured before him among philosophers and psychologists of the nineteenth century, as in Edward von Hartmann's work of 1869, *Philosophy of the Unconscious*. Yet it was Freud who instituted it as a fundamental feature of the human mind and constituted an elaborate architectonic schema of mental activities below conscious awareness in the preconscious, subconscious and deep unconscious levels. This was a remarkable achievement regardless of how much of Freud's theory one happens to accept: his view for example that the unconscious is constituted solely by repression or the specific repressed content or the drives or censorship mechanisms. As we shall soon see, there are also other ways of viewing the unconscious and what it contains.

The importance of the unconscious in the architectonic of the human mind is undeniable. It is one of the crucial factors, in fact, a whole dimension, distinguishing it from animal mind. The behaviour of an animal is too simple for us to infer a hidden but meaningful background to it; and its lack of speech makes it impossible for it to convey mental states of awareness that can be interpreted as manifesting an unaware or hidden side. The animal's lack of self-consciousness means that it cannot have an unconscious. But it is different with humans, starting with very young children, as Freud's analysis of his grandson's verbal behaviour – in playing a "*fort-da*" game to symbolically enact the separation from his mother – reveals.[7] No animal could possibly evince that kind of symbolic behaviour. Anna Freud and her great rival Melanie Klein both explored the unconscious in children at various ages, which they rightly took to reveal the stages in the development of mind. The unconscious, of course, is not present at birth, any more than is the mind in general. It is not a receptacle waiting to be filled. It has to emerge ontogenetically through socialization and enculturalization like everything else in the human mind. It's origin is closely aligned with the acquisition of language, but whether it is tied to language in the way that Jacques Lacan mandates, when he states that the unconscious is structured like a language, is another matter.

There is much more to the unconscious, as we shall presently see, than the Freudian unconscious of the Id and Superego and the whole dynamics of repressed desires, fears and traumatic memories or the ways these manifest themselves in consciousness and behaviour. Apart from Behaviourism, of course, which denies mind or consciousness altogether, every other major psychology brings with it its own conception of the unconscious. The most recent of all such movements has arisen calling itself The New Unconscious. Obviously, we cannot discuss all of them here, nevertheless, a few crucial features regarding these various versions of the unconscious will be explored next in what follows.

Most of these researchers now contend that before we can study consciousness we must first understand its unconscious basis. For in some sense consciousness emerges out of the unconscious, such that every conscious act has its unconscious preconditions. But this can also be inverted, as every automatically carried out activity demonstrates, since what a person has come to perform automatically – and in that sense without conscious controls and even in many cases any awareness – had to be consciously learned and practiced in the first place. The pianist who as an adult can play scales in his sleep, as it were,

7 Sigmund Freud, *Beyond the Pleasure Principle*, trans. James Strachey (New York: Norton, 1961), 9.

had to persistently and assiduously acquire that capacity as a child. Even earlier in life, many motor and perceptual capacities had to be consciously acquired before they came to be performed automatically and at times unconsciously. Thus the conscious-unconscious coupling constitutes an interlocked and mutually determining architectonic structure of the human mind.

A simple example of the workings of this system in perception is the recognition and identification of faces, which despite some animal precursor recognition abilities is a pre-eminently human accomplishment that begins in infancy. We can all distinguish between male and female faces, but how this is done is largely unknown to us because it is achieved unconsciously. It must have been learned at some early stage in life; and when and how this took place can be studied experimentally by tests carried out on babies and young children. A much more developed skill, that only very few can master, is being able to tell apart the sex of birds even when they are chicks and not distinguishable to the untrained eye. So called chicken-sexers have obviously learned this unconscious perceptual feat at some point in their lives. How they do it is as much a mystery to them as it is to unskilled observers. It clearly involves unconscious perceptual processes that are the subject of experimental study in the psychology of perception. There are many other such perceptual feats that are largely achieved unconsciously.

Experimental work in Gestalt and other such psychologies has demonstrated that much in perception that we believe is purely conscious – and, therefore, presumably clearly given to the senses – is really the outcome of contextual factors of which the perceiver is largely unaware, as James Crutchfield reports:

> The human perception of colour in the small region of space, for example, can depend on the colour composition of the entire scene, not just on the spectral response to spatially-localized retinal detectors. Similarly, the perception of shape can be enhanced by global topological properties, such as whether or not curves are open or closed.[8]

There are numerous such examples where unconscious factors play a crucial role in what is perceived.

But even more remarkable than such unconscious features in perception is the fact that perception itself can be to a lesser or greater degree unconscious. This is obvious from many common experiences, one such is what is called the

8 James P. Crutchfield, "Is Anything Ever New? Considering Emergence", in Mark A. Bedau and Paul Humphreys, eds. *Emergence*, op. cit, 208.

cocktail-party effect, suddenly hearing one's name or some such salient detail emerging from the background hubbub in a crowded room made up of numerous conversations and music to which one had not been paying any attention and can remember hearing nothing else at all. Nevertheless, these conversations must have been unconsciously monitored for such an item of interest to have been picked up so unerringly.

Unconscious perception is a fact. The term is almost anathema to Cartesians, but is nevertheless a reality that has been experimentally established. It comes in many guises that are being investigated by perceptual psychologists. The most obvious is subliminal perception, where something is presented, usually as an image on a screen, too fleetingly to be consciously perceived, and yet its effects on the perceiver can be clearly detected. There has been much experimentation carried out along these lines. The most startling of all forms of unconscious perception is the recently discovered phenomenon of blindsight, where a brain-damaged patient who is otherwise blind and has no consciousness of vision or sensory data is, nevertheless, able to intuitively register the presence of objects and perform appropriate actions in relation to them. This has led some psychologists and philosophers to the unwarranted conclusion that all perception could be unconscious and that consciousness is strictly speaking unnecessary for the human mind. It is, perhaps, only an accidental accessory to the unconscious mechanisms, a kind of epiphenomenon that does not do real work and has no function to perform.

Such extreme conclusions are being mooted by psychologists in the newly arisen school calling itself The New Unconscious.[9] In numerous studies that they have undertaken, they have sought to show that anything which is done consciously can also be done unconsciously. Some very startling findings have been unearthed, particularly so as regards unconsciously performed action and social behaviour, on the basis of which Jack Glasser and John Kihlstrom conclude as follows:

> Bargh and his colleagues … have brought their research to bear to make the compelling argument that all major mental processes, including motivation, can operate automatically. More recently, Bargh, Gollwitzer, Lee-Chai, Barndoller and Trötschel (2001) have demonstrated that nonconscious goal pursuits possess properties similar to those deemed fundamental to conscious motivation, specifically vigorous action toward goal satisfaction, persistence and resumption after disruptions. Although

9 See Ran R. Hassin, James S. Uleman, John A. Bargh, eds. *The New Unconscious* (New York: Oxford University Press, 2005).

> to date less comprehensive than the research on unconscious cognition and affect, their work strongly indicates that goals and behaviour can also be activated automatically and will be pursued unconsciously.[10]

John Bargh affirms this in stating that "action tendencies can be activated and put into motivation without the need for the individual's conscious intervention: even complex social behaviour can be unfolded without an act of will or awareness of its sources."[11]

This is, indeed, true for specific actions in exceptional circumstances performed by people in unusual states. It is not true for all actions, for from the premise that some things can be done unconsciously, the conclusion that all can does not follow. One cannot lead a normal human life unconsciously; spurious philosophical arguments for zombies apart. As Donald has argued, there are types of complex social behaviour that can only be performed consciously. Engaging in a contentious intellectual argument among a number of interlocutors can only be followed and participated in by taking careful note of what each person is saying, and how and why each point raised bears on the whole argument. Almost by definition, this must be conscious. And even more crucially, as Donald points out, nearly everything we learn must be learned consciously, there is almost no such thing as unconscious learning; even though once learned that which is mastered can become automatic through constant repetition and so become unconscious. [12] This is so in the case of the pianist who can play scales in his sleep.

There are many psychologists and philosophers who deny the reality or the necessity of consciousness altogether. These are the people Donald calls Hardliners, and according to him "the *capo di tutti capi* of the Hardliner school is Daniel Dennett" who argues that "the illusion that we are consciously in charge is itself an after-the-fact perception, a by-product of brain processes that we cannot consciously control."[13] In other words, consciousness is an elaborate "Darwinian illusion" that the brain creates in order to deceive itself for evolutionary reasons.[14] Dennett goes much further than any of the New

---

10 Jack Glaser and John F. Kihlstrom, "Compensatory, Automaticity: Unconscious Volition is not an Oxymoron", in Ran R. Hassin et al., *The New Unconscious*, op. cit, 73.

11 John A. Bargh, "Bypassing the Will: Toward Demythologizing the Nonconscious Control of Social Behaviour", in Ran R. Hassin et al., *The New Unconscious*, op. cit, 54.

12 Merlin Donald, *A Mind So Rare*, op. cit, Chapter 1.

13 Ibid, 39.

14 Ibid, 40.

Unconsciousness psychologists, "he is actually denying the biological reality of the self: selves, he says, hence self-consciousness, are cultural inventions."[15]

Bargh is not a Hardliner, he does not deny the reality, indeed, the necessity of consciousness, and he quotes Donald in order to establish what its role might be and why it evolved in the first place. He puts it that "in a very real sense, then, the purpose of consciousness – why it evolved – may be for the assemblage of nonconscious skills."[16] But once having completed this preparatory task, it is no longer needed: "intriguingly, then, one of the primary objectives of conscious processing may be to eliminate the need for itself in the future by making learned skills as automatic as possible… the evolved purpose of consciousness turns out to be the creation of ever more complex unconscious processes."[17] Bargh has clearly not grasped Donald's main point that such "complex skills" simply cannot be practiced unconsciously, as the example of the conversation demonstrates.

The argument that the exponents of the New Unconsciousness deploy, generally by implication, is fallacious. From the premise that anything can be done unconsciously, it does not follow that everything can be done unconsciously; any more than from anything can be an illusion does it follow that everything can be an illusion. And just as we can tell that something is illusory because most other things are real, so, too, the fact that some things can be done unconsciously is made possible by the fact that most things are done consciously.

Some exponents of the New Unconscious go even much further than Bargh is prepared to do and argue, almost like Dennett, that our sense of consciousness, of will, of intention, indeed, our sense of Self are all illusory. Thus Daniel Wegner contends as follows:

> Theories of controlled processes often imply that the person (or some other inner agent such as "consciousness" or "the will" or "the self") is a legitimate cause of the person's observed thought or behaviour. This supposition undermines the possibility of a scientific theory of psychology by creating an explanatory entity that cannot itself be explained.[18]

Wegner and most of his colleagues in the New Unconscious psychology movement seem to be in agreement with Daniel Dennett when he asks rhetorically:

15 Ibid, 40.
16 John A. Bargh, "Bypassing the Will", op. cit, 53.
17 Ibid, 53.
18 Daniel M. Wegner, "Who is the controller of Controlled Processes?", in Ran R. Hassin et al., *The New Unconscious*, op. cit, 19.

"what is consciousness for, if perfectly unconscious, indeed, subjectless, information processing is in principle capable of achieving all the ends for which conscious minds were supposed to exist?"[19] Wegner, in agreement, also holds that "controlled processes can indeed be understood as mechanistic processes – for example as in cybernetic and dynamical processes posited in control theories."[20] Most of Wegner's effort in his article is devoted to explaining what it is that produces the illusion of a controller, such as will, intention, consciousness or what he calls a "feeling of authorship". He purports to show that "our experience of conscious will is normally a construction … this construction yields a feeling of authorship of the action."[21]

As our argument concerning emergence shows, to propound a scientific psychology does not require relegating consciousness, will, intention and Ego, indeed the major part of our mental life, to the illusory realm of constructed "feelings of authorship." Nor does it mandate a mechanistic approach such as that of Cybernetics or its successors such as Cognitivist information processing, as we shall argue in Chapter 7. But, this is precisely what the New Unconscious movement expounds, as James Uleman's introductory proclamation makes evident:

> So what is new about the new unconscious? It is still basically cognitive, firmly embedded in cognitive science and historically beholden to the computer as a metaphor. The computer metaphor legitimizes complex theories about unobservable processes while apparently avoiding the sins of anthropomorphizing and using homunculi as causes.[22]

By relying on the "computer metaphor", that is, an information processing model of mental functioning, the New Unconscious psychologists have reverted to a reductivist account in which emergence plays no role.

Indeed, the information processing account provides the illusion that a neurological reduction of unconscious processes is already available or that it will be so very soon. Thus Bargh quotes the conclusion of Posner and Di Girolama "that the information-processing and the neurophysiological levels of analysis have achieved a level of mutual support greater than previously

19 D.C. Dennett and D.R. Hofstadter, eds. *The Mind's I*, (New York: Basic Books, 1981), 13.
20 Daniel M. Wegner, "Who is the Controller of a Controlled Process", op. cit, 22.
21 Ibid, 22.
22 James S. Uleman, "Introduction: Becoming Aware of the New Unconscious", in Hassin et. al. *The New Unconscious*, op. cit, 6.

imagined."[23] Bargh jumps from this to the more extreme conclusion that "conscious intention and behavioural (motor) systems are fundamentally dissociated in the brain", which amounts to saying that intention and action have nothing to do with each other. This could only be true on the assumption that intention, will and consciousness are illusory or purely epiphenomenal with no causal consequences, as Wegner, and many others argue. Bargh, however, confuses this with the view that "the evidence shows that much if not most of the working of the motor systems that guide action are opaque to consciousness", which is, indeed, true. But this is not to say, as Bargh goes on, that "such a dissociation of nonconscious social behaviour and consciousness is now emerging as a basic structure of the human brain."

What we must now conclude from all this work is that the unconscious is much wider and more ubiquitous than even Freud supposed it to be. For beyond the Freudian unconscious there are also numerous other forms of the unconscious in a number of its modalities. In the previous chapter we made a study of one of these, the perceptual unconscious. We surmised that behind human acts of perceiving there are unconscious inferences, expectancies, hypotheses, models, valuations and interpretations which perceptual psychologists from Helmholtz onwards have studied. Bruner and Gregory have been prominent modern researchers in this field of the perceptual unconscious.

Their favoured objects of investigation have been visual illusions for these reveal what it is beyond the nature of the stimulus or even the Gibsonian affordances that influences what is perceived. These determining factors are unconscious but mind governed. They are not a direct outcome of the brain, though this is often how experimenters who have not thought through the philosophical issues speak, using expressions such as "the brain interprets or hypothesizes or builds models", etc. But the brain only carries out neural processes, it does not act in a meaningful way, only the mind can do that; this is also the case where the unconscious rather than the conscious mind is involved. Though it is, of course, true in general that these unconscious activities of the mind emerge from brain processes in ways we understand no better than the conscious activities despite all the on-going neurological researches.

We can generalize these insights to all the acts and faculties of mind. Not only is there a Freudian unconscious and a perceptual unconscious, but also an unconscious underpinning for every faculty of the mind. This is even true for acts of free will, such as rational decisions and choices. As Crick is quoted remarking, "what you are aware of is a decision, but you're not aware of what

23 John Bargh, "Bypassing the Will", op. cit, 54; see M.I. Posner and G. Di Girolamo, "Cognitive Neuroscience", *Psychological Bulletin* 126, 873–89.

makes you do the decision."[24] What makes one decide for one thing or another or to follow a certain rational course or another equally possible one or none at all always involve unconscious motivations. As we shall show later in this chapter, this does not invalidate free will, except in the purely metaphysical sense of the will that is utterly uncaused and undetermined by anything else. Such total "freedom" does not exist.

As Freud and others have argued, the human mind has an unconscious dimension that is even more capacious and complex than the conscious one. For example, our unconscious memories are vastly larger than what we can consciously bring to mind. Such an unconscious is the distinctive feature of the human mind as a whole and distinguishes it from animal minds. Animals have no unconscious or if they do it is a purely incipient one, only to be found among the higher apes. It is possible that such animals dream, for example, but without any language capacity they can give us no inkling of the content of such dreams. In any case, the content of human dreams is closely bound up with linguistic meaning, as Freud demonstrated. Similar considerations hold for many types of visual illusions, such as those in which visual judgement is exercised, and also for Gestalt perceptions of figures with alternating aspects which can only be verbally reported.

The role of the unconscious in informing the conscious in all its aspects is now widely recognized, not only among psychoanalysts and psychologists, but also among neurologists. Gerald Edelman and Giulio Tononi devote a chapter of their book on consciousness to the unconscious. They write that "unconscious aspects of mental activity, such as motor and cognitive routines, and so-called unconscious memories, intentions, and expectations play a fundamental role in shaping and directing our conscious experiences."[25] Later in Chapter 6 we shall see how this applies even to such experiences as those of free will. Edelman and Tononi put forward a neurological theory of "possible neurophysiological substrates of those scores of processes that continually go on in the brain, which, although unconscious, can nevertheless influence or be influenced by consciousness."[26] The interaction between the conscious and the unconscious is a two way process, either one can play a determining role for the other, either one can be cause of the other. Consciousness is crucial since learning can only take place in consciousness, there is no unconscious learning. However, once learned, all kinds of routines can rapidly be performed

---

24 Quoted in John Horgan, *The Undiscovered Mind,* 267.

25 Gerald Edelman and Giulio Tononi, *A Universe of Consciousness: How Matter Becomes Imagination*, (Basic Books, New York, 2000), p. 176.

26 Ibid, 177.

unconsciously in all kinds of ways. Highly elaborate skills become automatic and in that sense unconscious once they have been perfected and can be done effortlessly without any supervisory attention, as is clear from the ability to drive a car while listening to the radio as well as carrying on a conversation.

How all this is neurologically accomplished by the brain is what Edelman and Tononi are intent on explaining. Their theory requires identifying the thalomocortical system as giving rise to what they call a "dynamic core" where a subset of neural groups contributes directly to conscious experience. This is their "dynamic core" hypothesis which cannot be gone into in this context; if accepted, it serves to provide a general neurological basis for how conscious and unconscious processes interact. They are careful to specify that it is only a very general explanation for such relations of the mind; as they declare in all modesty:

> In order not to indulge in speculative neurology, we resist the temptation to suggest possible neural mechanisms for aspects of unconscious cognition that, while having obvious psychological significance, are far removed from neurophysiological understanding. We refer, for instance, to the role of unconscious contexts, such as unconscious expectations and intentions, in shaping conscious experience; to the effects on consciousness of surprise or violation of expectation, to the conscious and unconscious relation of attention; and to the substrates and mechanisms of the Freudian unconscious.[27]

Freud would have been gratified by that statement for he had to give up the attempt to explain neurologically the workings of the unconscious despite his long early career as a medical researcher and extensive publications in physiology. However, he never gave up the hope that someday such studies of the unconscious would become possible and yield results. It would have saddened him to discover that a century after most of his own work in psychoanalysis was completed this is still not possible, except in the generality of Edelman's and Tononi's treatment.

Nevertheless, it must be acknowledged that great progress has been made in the understanding of mind and brain, and the relation between them. But there is still a long way to go before such understanding will be scientifically adequate, which some scientists measure in centuries or millennia and others hold that it will never be completely realized. But if anything like adequate understanding is nowhere in sight, and nowhere near that available in other

27 Ibid, 178.

areas of medical science – for this is perhaps the ultimate problem of science – there is much that is known. In the next section we proceed to say something about what is known, beginning with a general outline of the structural architecture of the human mind.

2 Vertical and Horizontal Structures of Mind

Our previous discussion of the unconscious-conscious relation permits us to sketch out a preliminary version of the overall architectonics of the human mind. This can be presented in graphic terms as consisting of a vertical structure and a horizontal structure. In its vertical structure it invokes the Freudian model of the various layered unconscious levels between neurology and consciousness previously adumbrated. In its horizontal structure it involves the vestiges of its previous evolutionary stages that remain encapsulated in its present state, such as we discussed in the previous chapter by reference to Donald and other neurologists. Both the vertical and horizontal structures are emergent entities in both the synchronic and diachronic senses of the term, as these are defined by Paul Humphreys and discussed in the next chapter.[28]

In its vertical structure, the mind/brain system consists of three levels working synchronically in unison: a conscious stratum of mental acts and behavioural expressions; an unconscious stratum of the Freudian and non-Freudian types previously outlined; and, finally, a non-conscious stratum of brain neurology coupled to a body and to both physical and socio-cultural environments external to the body. This human mind/brain system thus contrasts to the two dimensional and one dimensional systems that are present in animals. As we shall see in Chapter 6, speaking of mind/brain systems lends itself to interpretation in Spinoza's metaphysical terms of a substance that has different attributes, such as, according to Spinoza, thought and extension, but which we can interpret much more broadly than that.

However, the mind/brain system can also be treated horizontally in terms of its evolutionary development. Seen in these terms, the human and animal systems are not at all opposed to each other for the former in some sense contains the latter as a surviving vestige. The one dimensional and the two dimensional animal mind/brain systems are still partially encapsulated in the three dimensional human one. They are hierarchically nested within it and can operate in a quasi-autonomous way outside the control of the human mind/brain system. In fact, there is very little of the body's autonomic functions or its instinctual

28 Paul Humphreys, *Emergence: A Philosophical Account*, op. cit.

and automatic behaviour that is or can be brought under conscious control. As Donald points out, "this includes many basic reflexes, such as coughing, sneezing, vomiting, balancing, blinking when the eye is touched, and many complex instinctual systems, such as those involved in hunger, thirst, eating behaviour, and eye movement."[29] The reason for this has obviously to do with evolutionary survival value and is explained by the fact that our vastly developed mind/brain system emerged from the two older systems, vestiges of which it still retains.

The Darwinian term "vestiges" is the key to the relation between the one, two and three dimensional mind/brain systems. Donald uses it to "refer broadly to any remnant of the past that endures, whether it is functional or not. The word thus conveys a sense of evolutionary continuity."[30] And he issues the necessary admonitory caveat to any study of the human mind:

> Vestigial structures are an overwhelming reality in the study of the nervous system, and it is therefore central to our argument. We have inherited a deep mental structure that is a direct reflection of the evolutionary history of our species. This fact must be taken seriously by any discipline that pretends to study the human mind.[31]

Indeed, as we shall see in Chapters 6 and 7, there are many disciplines that do, indeed, pretend to study the human mind but do not take this fact seriously. As Donald stresses, "humans have primate brains, and it does not make sense to postulate a complete rewrite of primate cognition in our case", in other words, we still retain much of previous animal minds. However, "the regions of the primate brain that expanded most dramatically in humans were those associated with executive function … these vastly expanded areas are important for supervisory, governing and metacognitive functions."[32] These are the functions which, according to Donald, are crucial to decision making, self-regulation, consciousness and self-consciousness, to which we shall turn later.

However, at this point it is necessary to stress that our conception of the three-dimensional mind-brain system linked to the idea of Darwinian vestiges does not mean having to fall in with Paul MacLean's conception of the triune brain.[33] According to MacLean, the human mind-brain is a composite of three

---

29 Merlin Donald, *A Mind So Rare*, op. cit, 107.

30 Ibid, 107.

31 Ibid, 106.

32 Ibid, 112.

33 Paul D. MacLean, *A Triune Concept of Brain and Behaviour: the Triune Brain in Evolution,* (Toronto: Toronto University Press, 1973).

distinct components localized in different parts of the brain: a reptilian part responsible for instincts and basic drives located in the basal ganglia; a palaeomammalian part responsible for emotions, habits and memories located in the limbic system; and a neomammalian part responsible for abstract thought, imagination, decision and language located in the neocortex. This once popular picture of a little crocodile, a little rat and monkey in the brain has since become discredited. The work of Damasio, for example, has shown that thought, decision and emotion cannot be separated and neatly compartmentalised in different parts of the brain. The brain operates as a unified integrated system in which various parts co-ordinate their activities, and not like three separate components that evolution has somehow bolted on to each other. We cannot pursue this neurological theme any further, but it is also perhaps necessary to note that it has a bearing on the Freudian vertical distinction of the conscious, pre-conscious and unconscious; there is no little savage lurking in the Id, as Freud at times seems to imagine.

The main way we differ from Freud is not only in respect of the vastly enlarged capacity of the unconscious and what it is capable of accomplishing (the so-called Freudian unconscious is only a small component of the total unconscious, which greatly exceeds the capacity of consciousness), but, furthermore, unlike the Freudian conception of the unconscious, the activities and processes of the unconscious are not necessarily governed by repression of primitive, animalistic drives. Freud is often inclined to view the unconscious as a kind of homunculus, like a little savage in the brain, and therefore far less logical or rational or restrained or disciplined than is the conscious mind. In Freud the Id is contrasted with the Ego, as is the Pleasure principle to the Reality principle. The unconscious is seen as totally bereft of logic, reason or reality, as it manifests itself, for example, in dreams. But this is not the only kind of unconscious activity.

We have shown previously by reference to the New Unconscious movement that the unconscious can be just as logical, rational, realistic, restrained and disciplined as is our conscious Self. The manifest proof of the rationality of the unconscious is the fact that difficult problems that one cannot consciously solve often solve themselves unconsciously. Answers to intellectual conundrums in the sciences or mathematics or any other endeavour with which a thinker has struggled consciously in vain for a long time can suddenly appear when least expected. There are numerous reports of a solution to a problem coming unbidden from the unconscious after someone has been grappling unsuccessfully with it and given up in desperation. Hence the advice given to people facing a great difficulty or dilemma – sleep on it; meaning, let your unconscious come to grips with it.

The unconscious can be not only rational but also creative, for many artists report that their best ideas come to them from the unconscious, and so, too, do inventors. The unconscious can be more restrained and disciplined than the conscious Self for very often people have unconscious forebodings and premonitions when they are about to make a rash decision or are tempted by deceitfully inviting prospects. None of this, of course, supports the New Unconscious theorists in their contention that everything can be accomplished unconsciously and that consciousness is otiose.

However, the unconscious must not be identified with the non-conscious, that is with brain processes or even with information processing as Cognitivist psychologists and neurologists are prone to do. The unconscious lies on the mental side of the mind/brain dichotomy and there is still, therefore, an explanatory gap separating it from what happens in the material side of the brain itself. We know that there is a relation of emergence between the two and explaining how this emergence takes place is the function of neuroscience. Whether neuroscience will ever provide a full and comprehensive explanation is a matter for future science that we cannot now predict. It is possible that the explanatory gap will be gradually closed as the mental and material sides are brought asymptotically closer and closer together without ever touching, like the finger of God and that of Adam in Michelangelo's fresco in the Vatican. On the other hand, it is possible that a large explanatory gap will always remain because some insurmountable barrier of complexity or chaotic turbulence is encountered in the brain. We cannot now predict what future research will discover; there are no *in principle* answers to such problems. Hence, it is futile to argue whether in principle this problem is or is not completely solvable. To assume in advance that it will be solved is an unwarranted expression of scientific hubris. We shall return to the issue of emergence and what it entails for science in the next chapter.

It is undoubtedly the case that tremendous scientific progress has been made in closing the explanatory gap during the last half century, and we can confidently expect this to continue into the foreseeable future given the enormous financial resources and scientific brain-power being devoted to this project. Some of the key advances already accomplished a quarter century ago are set out by Raymond Tallis and Howard Robinson:

> First of all, there have been important advances in experimental method: single cell recordings… novel methods of imaging the internal structure of neurons and the connections between them; cell culture techniques permitting growth and development of nervous systems to be studied in great detail; and sophisticated computational and statistical analysis

> data permitting better mathematical modelling of large and small scale events in the nervous system and its functional connections. Second, there have been important conceptual advances. Hubel and Wiesel [show that] nervous systems are so structured as to ensure that certain events of special importance to the organism have an increased likelihood of triggering activity in the relevant places. Alongside this, there has been growing awareness of the plasticity of the nervous system and of the extent to which the hard-wiring of the wetware of the brain is itself modulated by the brain's own experience.[34]

Since that time much more has been achieved and even more promised. In America the Obama administration invested 100 million dollars in its Brain Initiative and the European Union has granted its Human Brain Project one billion dollars. Numerous institutions are involved in these researches, for as well as universities there are also privately funded foundations such as the Allen Institute for Brain Science established in Seattle in 2012. Research on insects has already come close to providing a complete mapping of the brains of flies. Research on mammals, such as mice, is just beginning and is nowhere near that degree of detail, but the ambition is to attain similar results in the future. The aim is "to develop electron micrographs that show every neuron and every connection in that part of a mouse brain" that is active when it sees a picture and presses a lever.[35]

That degree of detail is still a long way off, perhaps 20 to 30 years away, as one hopeful young scientist involved in that research, Clay Reid, estimates.[36] But even if that were accomplished it would still be a long way away from explaining all that goes on in its brain when a mouse sees and acts. And still further removed by many degrees of difficulty of what goes on when a human does this. Discovering what Reid calls "the code in which information is moved around that part of the brain" might turn out to be a hopeless quest, as indeed other brain researchers, such as Gerald Edelman, would contend, as we shall see in Chapter 5. And Torsten Wiesel's comment on the kind of research that his student Reid is undertaking at the Allen Institute is that it will "provide a lot of useful information [but] it won't necessarily create breakthroughs in our understanding of how the brain works."[37]

---

34 Raymond Tallis and Howard Robinson, eds. *The Pursuit of Mind,* (Manchester: Carcanet Press, 1991), 86.

35 James Gorman, "The Brain's Inner Language", *The New York Times,* February 24, 2014.

36 Ibid.

37 Ibid.

This kind of research into the brains of mice and men raises once again the issue we dealt with in the last chapter, the difference between the animal and human mind. What distinguishes the three dimensional mind/brain system and the two dimensional one? It is imperative to avoid an obvious source of misconception, the idea that the human mind is simply the animal mind with an extra dimension, namely, that something has to be added to the animal mind to turn it into the human mind. Various candidates have been proposed for this added factor. It began with Descartes assuming that it was thought. Later it was believed to be consciousness that had to be added to sheer behaviour to turn animals into humans, for otherwise, it was held, the two kinds of bodies are much the same. Another differentiating factor which has persisted from Descartes to Chomsky is the idea that language is all that animals lack. Chomsky held that the acquisition of language is governed by a special module in the brain that makes humans genetically different. Chomsky has called it a Language Acquisition Device; and others, such as his follower Fodor, have gone on to try to identify more such genetically endowed modules that the human brain uniquely possesses. This modularity thesis has generated a minor industry leading to the theory of "massive modularity", that even Fodor has denounced as "modular madness", and which we shall discuss in Chapter 7.

Unfortunately for all such views, there are no such added ingredients that turn animal minds into human minds. There are no specific factors that make the crucial difference. The human mind is a completely different mind/brain system than animal mind. These differ in every respect, since most of what a human mind can do is unavailable to animals, as we discussed in the previous chapter. This holds true for all the so-called faculties of mind, not only for action and perception, but also for emotion and sensation. A human being not only carries out activities in a way that is incomprehensible to an animal and sees things of which an animal can have no glimmering, but a human also feels and senses things in a way of which no animals is capable. There are many kinds of emotions that are specifically human, such as pride or shame; and those that are common to animals and humans assume a peculiarly human aspect, as, for example, fear when one fears failure or a threat to one's dignity. Sensations can also be experienced differently by humans, such as pains which the proverbial Spartans or Stoics learn to bear with equanimity. Even intense unbearable pains can be allayed in humans by acupuncture treatment, but not so for animals. There are even sensations that only humans can experience and not animals, one such is the tickling sensation that makes one skittish and titter or laugh and feel squirmy. It is significant that one cannot tickle oneself and that this sensation can only be aroused by unthreatening others who are familiar to the subject being tickled. It is, as it were, a social sensation.

However, in apparent contradiction to all this is the fact that humans can also behave, perceive, feel and sense just like animals do. This happens both under normal and under exceptional circumstances. Human beings have analogous reflexes to those of animals; they respond automatically and instinctively to a wide range of stimuli; they are subject to basic animal emotions, such as fear of predators; they feel the same kinds of needs and desires, such as hunger and thirst, and so on. Under extreme conditions of privation or torture human beings can be reduced to something approximating an animal state. Under intense hunger they resort to cannibalism just as some animals do. In the Nazi concentration camps, in an almost experimental way administered by sadistic doctors, human beings were turned into walking corpses who only retained bare shreds of humanity, and it is not even possible to call them animals for they had become so debilitated.

To explain such animalistic phenomena we must turn from the vertical axis of the structure of the human mind to the horizontal axis, that which concerns the evolutionary stages of human development. It is a fundamental truth of evolutionary science that human brains and minds still retain features of their prior stages of development, namely, those deriving from their animal precursors, both apes and earlier species of hominins. In other words, the human three dimension mind/brain system still retains vestiges of the animal one dimensional and two dimensional systems. The latter are encapsulated or nested within the former like Russian dolls, except that they do not exactly replicate each other on smaller scales. For in evolution nothing is ever completely lost, but neither is it retained *in toto*, instead it is partially preserved and adapted to new functions in a different systemic complex. So it is with the animal mind/brain systems within the human one. One might call this a kind of evolutionary dialectic where prior stages are *aufgehoben* in Hegel's sense, both cancelled and preserved, in later developments.

This evolutionary nature of the horizontal structure is almost a *cliché* among anatomists and neurologists. Thus the neurologist Antonio Damasio remarks on this encapsulation principle from the anatomical point of view:

> When we survey the list of regulatory reactions that ensure our homeostasis we glean a curious construction plan. It consists of having parts of simpler reactions incorporate as components of more elaborate ones, a nesting of the simple within the complex. *Some* of the machinery of the immune system and of metabolic regulation is incorporated in the

> machinery of pain and pleasure behaviour. *Some* of the latter is incorporated in the machinery of drives and motivations… some of the machinery from the prior levels… is incorporated in the machinery of emotions proper.[38]

Thus, though nothing in evolution is ever completely lost, nothing ever remains completely the same for parts of it are readapted to new uses. As Damasio puts it, "each reaction consists of tinkered rearrangements of bits and pieces of simpler processes below." [39]Thus it is not the whole of the animal mind/brain system that is retained in the human one but only rearranged components of it. This is another reason why the triune brain model does not apply.

Another neurologist, Donald, applies an analogous principle of the encapsulation of earlier stages in later ones during the course of the evolution of the human mind:

> From the start I have made the simplifying assumption that each cognitive adaptation in human evolutionary history has been retained as a fully functional vestige. The simplest working hypothesis, by far, is that when we acquired the apparatus required for mime and speech, in that order, we retained the knowledge structures, and the cultural consequences, of previous adaptations… I have also used a structural criterion for establishing previous cognitive transitions: they were accompanied by basic architectural change, meaning that a different superordinate organization was imposed on cognition.[40]

Utilizing this criterion, Donald is able to argue that "there was a new overall architecture to human cognition" as compared to animal cognition, or that of our more animalistic evolutionary precursors. On this basis, Donald claims to be able to trace back the evolutionary sequence whereby the human mind emerged from animal mind; it is the series of what he calls cultures that we discussed previously. But this is also relevant to our present discussion of encapsulation:

> On the principle of the conservation of previous gains, episodic culture would have been surrounded by, and largely preserved, within the larger context of mimetic culture. *Erectus,* and modern humans, retained the

---

38 Antonio Damasio, *Looking for Spinoza*, op. cit, 37–8.

39 Ibid, 38.

40 Merlin Donald, *The Origins of the Modern Mind*, op. cit, 268.

> basic cognitive features that characterize episodic culture; there is no reason to believe that any higher cognitive skills were lost in the transition. For the most part, considering that the adaptations were entirely within the constraints established in the primate line, the transition to mimetic culture involved adding to the cognitive architecture already in place.[41]

It is clear from this that the principle of conservation does not just apply to the individual mind/brain system, but also to the whole socio-cultural system on which the mind is dependent. It is as much relevant to what Hegel calls "objective" mind or its socio-cultural aspect, *Geist,* as to subjective mind or individual psychology.[42] The connection between objective and subjective mind is what Donald's work brings out in terms of contemporary sciences of which Hegel had no inkling, of course, but which, nevertheless, relate obliquely to his philosophy and retrospectively shed light on it.

The study of the evolutionary transition from the mindless system of the lower animals to the two dimensional mind/brain system of the higher animals and eventually on to the three dimensional human one is now the subject matter of numerous sciences which barely existed in Hegel's time. Neuroanatomy, neuroscience, the sciences of both animal and human behaviour, the psychology of behaviour, learning capacities and many more mental processes have been studied with a view to establishing the forms of transition that occurred in the course of evolution. As we have already seen, there is a major barrier standing in the way of establishing continuities between the animal and human minds, the fact that the intermediate hominin species between apes and homo sapiens are missing. This is the missing link that cannot be completely recovered, but only surmised with a great degree of uncertainty. This, as we shall soon see, has very serious consequences for our ability to explain the emergence of human mind.

This phylogenetic gap has a kind of complement on the ontogenetic side. Of course, there are no gaps in the development of a human being from embryo to adult. However, our capacity to scientifically study that continuous development is seriously constrained by ethical considerations, necessarily so. There are few ethical restrictions on the breeding, rearing and killing of animals or subjecting them to conditions that would be considered inhumane if practiced on humans. Hence the possibilities of studying the ontogenetic development

41 Ibid, 197.

42 G.W.F. Hegel, *Philosophy of Mind: Part Three of the Encyclopaedia of Philosophical Sciences*, trans. A.V. Miller (Oxford: The Clarendon Press, 1971), Section 2.

of the human mind are far more restricted than those for animal minds. If it were ethically feasible to do to children now what Dr Mengele did to Jewish children in Auschwitz, then undoubtedly much faster progress would be made in explaining the ontogenetic emergence of human mind. But this must never be permitted if we wish to remain human morally speaking. Hence, we must remain dependent on fortuitous, though invariably tragic, accidents to reveal how it is that the brain forms and how the mental faculties develop, and in general how the mind/brain system constitutes itself in individual human beings.

Despite these irremediable shortcomings, both phylogenetic and ontogenetic emergence are now substantially, though highly imperfectly, understood. These two modes of diachronic emergence shed considerable light on what we can discover concerning synchronic emergence. They also make it easier to understand what this involves, which is necessary because this is the mode of emergence that philosophers and scientists find it most difficult to accept. The bugbear they are haunted by is the thought that consciousness can emerge from the firing of neurons, a sheer impossibility to their Cartesian minds. And in a sense this is true, for consciousness does not directly emerge from the firing of single neurons, only very indirectly via intermediate processes involving patterns of neural organization that are not directly observed, like the firing of single neurons, but can only be hypothetically surmised by means of holistic models of the working of the brain as a whole.

If we called such models "virtual" emergence, then we might put it that mental entities do not directly emerge from the firing of single neurons but only through the mediation of numerous types of "virtual" emergence. But such "virtual" emergence is not real emergence, it is only a way of expressing the hypothetical nature of neurological models that both neurologists and psychologists necessarily employ. Determining what is real and what is merely "virtual" in the brain is no easy matter. Are the "maps", "hypotheses", "images", "expectancies", and so on that psychologists from Edward Tolman onwards have postulated, real or merely virtual? This is really a debate about the nature of theoretical entities that are not observable but have to be postulated to explain what is observed to be taking place. What takes place in the brain, the connections and interrelationships beyond the observed firing of individual neurons, is so complex that it cannot be completely observed. Hence, "virtual" models are inevitable in neurology and "virtual" emergence is a concept we have to invoke to explain real emergence, such as that of the contents of our minds.

Most philosophers and scientists seem willing to grant the reality of diachronic emergence, both phylogenetic and ontogenetic, but far less so synchronic emergence. What they cannot grasp is that synchronic emergence is the cognate mode of the two diachronic ones. If it is granted that animal mind emerges in the course of evolution, in accordance with all the purely physical developments in animal bodies, above all, the growth of the brain, then the idea that an animal mind is instantiated at the same time as that brain comes into operation should not seem weird. And the same kind of reasoning goes for ontogenetic development at the animal level. It is only when synchronic emergence is seen as the logical outcome of diachronic emergence also at the human stage that difficulties proliferate. But this is mostly because of the seemingly fundamental discrepancy between consciousness and neural firing. This discrepancy is, of course, much more pronounced in Cartesian philosophies of mind than it is in Aristotelian ones, as we shall go on to show in Chapter 6. Hence, the main problem seems to be one of having to change one's philosophy of mind before synchronic emergence can be accepted.

Synchronic emergence is usually studied in two alternative ways, either following a top down trajectory or a bottom up one. One can call these downward causation and upward causation, provided one does not take emergence to be an inherently causal relation, but one to which a causal direction can be ascribed either way for practical, heuristic or scientific convenience, as we shall argue in Chapter 5. Following the top down route, it is the human subject who initiates the observation by performing a physical or mental act or by reporting on a mental experience. The scientist then tries to correlate this with what can be observed going on in the brain and body, and that might be taken as the physical effect of that mental cause in what is called downward causation. Such observation is enabled by high-technology instruments, such as fMRI, MEG, EEG and others, each of which has its own inbuilt limitation. Generally, the main limiting factor is that the observation of states of the brain over-all, such as encephalographic readings of general electrical activity or chemical action potentials, cannot be resolved into its structural details. On the other hand, highly specific recordings of details, down to the firing of single neurons, can only be done with very few of the millions and possibly billions of neurons that are jointly involved in the myriads of ongoing activities. These two types of observations can rarely be brought together and integrated.

The same sorts of problems arise on the alternative route of the bottom up trajectory in upward causation protocols, when it is the researcher who initiates the causal process by stimulating the brain in some way, either physically, chemically or electrically, and the subject then reports what effect this had produced in the conscious mind. The stimulation can be conveyed by acting through the senses or directly on the brain itself. If the stimulation is through the senses, such as shining lights or making sounds or touching certain points on the body, then it becomes difficult to follow its course once it passes the sensory nerves and enters the central brain areas, and impossible to trace it any further beyond certain sensory locations in the brain. On the other hand, stimulating nodal points in the brain directly and noting the mental effects that follow only provides the initial casual factors and tells us nothing about all the other activities that had to take place in the brain in the course of the synchronic emergency of the mental act or experience.

The dilemma of neuroscience is that it either produces detailed knowledge with no sense of any over-all system as in single cell recording; or it produces knowledge of over-all very general activity, such as electrical or blood circulation changes, with no way of determining the constitutive factors or details. It is a dilemma that cannot be overcome by further observational research alone. It can only be mediated by means of theories and models, at present mainly computer models. Without such theoretical guidance all observations are merely fortuitous or accidental correlations.

But, unfortunately, as we have learned over the past century or so since brain research began in earnest, theories and models come and go; like other rapidly changing approaches in science, they almost alternate with the seasons. Thus, Margaret Boden writes: "And in the computational psychology of the 1960s-1970s GOFAI [Good Old Fashioned Artificial Intelligence] was the preferred approach. Grossberg attracted scepticism accordingly. Today [namely in 2006], he's lauded as one of 'a few hardy researchers' who persevered with neural-network models."[43] She is referring to Steven Grossman, a neuroscientist who began at MIT, but was dismissed at least partly on account of the fact that the neural-network approach was in disfavour there, while the GOFAI method ruled supreme for both theoretical and funding reasons. Boden devotes a whole chapter to the struggle between GOFAI and Connectionism, the approach that favoured Parallel Distributed Processing. It began on the personal level as a feud between Marvin Minsky at MIT and Frank Rosenblatt at Cornell. Minsky won the first round, but Connectionism came back with a vengeance in the late 1980s largely at the hands of J.L. McClelland and D.E. Rumelhart. Boden documents this battle royal between the two preferred types of models which went on for half a century or longer.

Late in the game, a third type of model entered the fray, that of dynamical systems theory. As Boden reports, "a few neuroscientists and psychologists were starting to explain brain and behaviour in terms of dynamical attractors."[44]

---

43 Margaret Boden, *Mind as Machine*, vol. II, op. cit, 1159.

44 Ibid, 1308.

But, as she goes on to point out, there is a huge variety of approaches and theories even under this one umbrella model:

> Dynamical systems go under various names. They're called circular causal systems, reaction-diffusion systems, neural networks, cellular automata, NK networks and CTRNs (Continuous-Time Recurrent Networks). And they even include gases. At the most fundamental level all these are similar (though the defining rules are different).[45]

A theoretical model that applies equally to gases as to brains and minds leaves one in some doubt as to its explanatory relevance or scientific utility. And Boden seems to share that doubt, for she writes that "dynamical systems are still more theoretically opaque than either GOFAI or orthodox connectionism."[46] As she goes on to point out, "cognitive scientists who manage to get interesting behaviour out of a dynamical system typically don't know just why it occurred ... Dynamicists rarely know just what it was about *this* dynamical system which made it capable of generating *that* behaviour."[47]

Here we encounter another fundamental dilemma in understanding and explaining the mind/brain system: the models we build, especially when they are computer models, are themselves almost as complex as the thing they are intended to model. This also holds true of the mechanical counterparts or machines that mimic the brain system: if they generate interesting behaviour, there is no way of knowing how it has been accomplished. This is the kind of dilemma of modelling the mind by mechanical means to which John von Neumann adverted as early as the Hixon conference of 1948 when he stated that "it is futile to look for a precise logical concept, that is, for a precise verbal description, or "visual analogy". It is possible that *the connection pattern of the visual brain itself* is the simplest logical expression of this principle."[48] Boden spells this out as follows:

> Certainly, one can (today) simulate a vast amount of psychological and neurological data in a single computer model. But one runs the risk of having an all-singing, all-dancing model that performs well (i.e.

45 Ibid, 1308.
46 Ibid, 1308.
47 Ibid, 1332.
48 Quoted in ibid, 421.

> that matches the experimental data) but which no one can actually *understand*.[49]

As we shall see, this raises all kinds of difficult conceptual problems as to what it means to understand a model of the brain we have built or a device that can equal some of the brain's functions. Can such a device be considered a machine if its operations can no longer be surveyed or in that sense understood? We must postpone such difficult issues till later in Chapter 7.

The holy grail of the sciences of mind is to establish theoretical linkages between unconscious mental processes and neurological models of the workings of the brain on the large scale level of holistic structures. It is in such linkages that a scientific explanation of the emergence of mind is to be sought. At present such accounts that span the mind-brain divide are only tentative and apply, if at all, in the vaguest generality. But as science progresses more detailed and specific and better supported theoretical linkages will no doubt be devised. Whether a full and comprehensive theory will ever be available is unpredictable, as is so much else about future science. Nevertheless, even such theories spanning the divide as are now current – partial though they be and almost certain to be superseded – point to the fact that the emergence of mind is not an incomprehensible mystery. It can be scientifically tackled and at least tentatively explained.

Perception is the one significant area where work on such linkages between mind and brain has gone on for longest and has perhaps been taken further than in any other field. Psychologists have always sought for the physiological and neurological correlates of the phenomena of perception and the unconscious perceptual processes on which they depend. This was already a methodological aim for the Gestaltists. From the very start of this movement with Max Wertheimer, they were committed to eliciting the underlying and largely unconscious laws governing the Gestalt phenomena of conscious perception. But these laws, they held, derived from fundamental organizational processes in the neurology of the brain. They focused, above all, on how the brain processed the sometimes meagre data that the stimulus object provided through the physiology of the sense organs, primarily the eyes. That is the main reason they were so intent on so-called multistable or alternating aspects phenomena, for in these the stimulus stays the same while the perception keeps on changing. The question that puzzled them, as it has done generations of later researchers, is what is taking place in the brain to explain such alterations in the phenomenological experience of perception.

---

49 Ibid, 421.

Among the Gestaltists, none was more intent on addressing this issue than Wolfgang Köhler. He devoted all his efforts during his last years in America to it. Unfortunately, he went about it in the wrong way because at that time, in the early years after the Second World War, very little was as yet known of the neurological operations of the brain. Also, he was mistaken in seeking a too direct and parallel correlation between the Gestalt phenomena and their laws and what supposedly had to take place in the brain to account for them. In short, he sought for electrical circuits in the brain that would correspond to the Gestalt laws of closure in the mind; and, of course, he found none. Psychophysical parallelism is not the right approach, as we shall ague in Chapter 6.

How he got himself into this fix is recounted by Michael Stadler and Peter Kruse:[50]

> Köhler referred to the existence of physical systems in which complex structures originate spontaneously from the system's own inner dynamics... Consequently, Köhler applied the model of physical fields, striving independently to a balance of forces, directly to the way the visual system functions.[51]

In other words, Köhler assumed that self-constituted structures comparable to soap bubbles arose in the brain to account for Gestalt features in perception. Given the knowledge of his time, this was reasonable, since "the only principle of self-organization known in his time [was] the tendency towards final thermodynamic equilibrium [but this] did not hold for the increase of order in the development of biological systems", as Stadler and Kruse argue. Consequently, though suggestive, his approach proved "hardly tenable and almost provocatively contradictory to the findings of neuroanatomy and physiology". According to Kőhler's "field-theoretic model of perception, the brain is not seen as a complex network of many different interacting neurons, but as a homogenous conductor (in an extreme sense comparable to a container full of water)."[52] To postulate water in the brain is a prescription for failure in this kind of research.

According to Stadler and Kruse, "it is easy to suppose that if the knowledge about the creation of order far from equilibrium obtained in the last two decades had been available to Köhler, he would have been led to a completely

50 Michael Stadler and Peter Kruze, "Cognitive Systems as Self-Organizing Systems", in W. Krohn, G. Küppers and H. Novotny, *Self-Organization,* op. cit.

51 Ibid, 183.

52 Ibid, 183.

different, rather network-oriented development of his application of the principle of self-organization to perception."[53] They are referring to the numerous theories of self-organization arising from the work of Ilya Prigogine in chemistry, Umberto Maturana in biology, and Herman Haken in synergetics, which "considered multistability an outstanding example of phase transition and symmetry breaking and therefore of self-organization in cognitive systems."[54] According to this view, these theories of self-organization linked to the Gestalt theory of perception take us a step closer to being able to link the brain with the mind:

> The theory of self-organization may be considered as a necessary methodological expansion for modelling inner-systemic interactions on the microscopic level, and gestalt theory may serve as a corresponding fund of descriptions of macroscopic behaviour. Therefore, as a linkage between neurophysiological and psychological aspects of cognition, it may help to solve the problem of reduction without reductionism.[55]

It is a consummation devoutly to be wished, but whether following this route will actually attain the holy grail is to be doubted.

All that Stadler and Kruse can point to is that "a promising example for the development of a linkage between microscopic and macroscopic modelling may be seen in the perceptual phenomenon of multistability. Indeed, there are already first steps to stimulate multistable perception on the basis of self-organizing neural networks."[56] They point to ongoing research work along these lines and quote from a number of articles.[57] But whatever success there appeared to be in the 1970s and 80s in modelling multistability, that is, aspect-changes, that is still a far cry from their conclusion that "following self-organization theory one could assume that every percept is principally multistable. There are only differences in the time the system needs to decide and in the duration of the elicited alternatives."[58] Namely, that self-organization theory is sufficient to explain the Gestalt phenomena of perception.

53 Ibid, 184.
54 Ibid, 186.
55 Ibid. 184.
56 Ibid, 185.
57 P.K. Kienker, T.J. Sejnowski, G.E. Hinton and L.E. Schumacher, "Separating Figure from Ground with a Parallel Network", *Perception* 15 (1986), 197–216; T. Poston and J. Stewart, "Nonlinear Modelling of Multistable Perception", *Behavioural Sciences*, 23 (1978), 318–34.
58 Michael Stadler and Peter Kruze, op. cit, 187.

This idea is taken up by Klaus Mainzer who develops it at book length. He states it as follows:

> The complex system approach offers the possibility for modelling neural interactions of brain processes on the microscopic scale and the emergence of cognitive structures on the macroscopic scale. Thus, it seems to be possible to bridge the gap between neurology of the brain and the cognitive science of the mind, which traditionally has been considered as an insoluble problem.[59]

It might or might not be possible to bridge this gap, at present we do not know, but it is highly unlikely that the complex systems approach on its own can do so. According to Mainzer's account, the key to this approach is that "the overlapping area of brain and cognitive sciences is modelled by the emergence of macroscopic properties from microscopic neural interactions during phase transitions in complex neural systems."[60] But phase transitions in complex neural systems remain on the neural side of the divide, and that still leaves a gap in explaining how cognitive states in the mind emerge. Nevertheless, Mainzer is certainly right in assuming that the gap is now much narrower than it was under "the old-fashioned formulation of an inadequate and obsolete metaphysics that assumes some interacting substances like colliding balls in mechanics."[61] He does grant that his approach is not the complete answer, that "it may fail in the long run".[62] However, he believes it to be the best research program that we have at present. Other neurologists, such as Edelman and Damasio, would disagree with that, for they pursue their own quite different approaches.

Nevertheless, the general position which they all seem to be conveying regarding the mind-brain relation does seem to be converging towards a view of emergence as an indissoluble non-identity. This is what Mainzer enunciates quite clearly and explicitly:

> Concerning the traditional mind-body problem, the complex systems approach shows that cognitive activity is neither completely independent and different from brain activity nor simply identical, nor an

59 Klaus Mainzer, *Thinking in Complexity: the Complex Dynamics of Matter, Mind and Mankind*, (Berlin: Springer, 1997), 150.

60 Ibid, 157.

61 Ibid, 157.

62 Ibid, 168.

> epiphenomenon. Thoughts and feelings are assumed to be both products and producers of neural processes, without being identical to them.[63]

Not only thoughts and feelings, but also all other mental activities such as intentions, volitions and actions are both "products and producers of neural processes, without being identical to them". In the next chapter, where we take up this idea, we shall find very similar formulations enunciated by both Edelman and Damasio. What we are to make of it philosophically is what we shall attempt to elucidate.

But before we come to that, we must explore in greater depth what it is that makes the human mind so unique and separates it by such a vast gulf from animal mind, the topic we have already discussed in the previous chapter. Basically this means establishing the mental capacities that humans have, but animals lack. There is much we have already said on this score regarding action, perception and emotion in line with the views of Donald, Gibson, Damasio, Edelman and others. In particular we have stressed the play and art or aesthetic features of human capacities, those for which Hamlet is our symbol. However, much more needs also to be said along the more traditional lines that Hamlet himself invokes in his peroration on Man starting with "how noble in reason" and ending with "how express and admirable in action". Reason and free will were traditionally taken as the main distinguishing features that separate the human from the animal. To this duo we shall add two more that Hamlet elsewhere adverts to, conscience or self-consciousness and Ego, to make up a quartet that characterizes the human mind at its best.

3 On the Quartet of Cardinal Qualities

The four cardinal qualities, reason, self-consciousness, Ego or subjectivity and free will, constitute the highest defining characteristics of the human mind. But before we can proceed to discuss each of them separately, we must first consider their common evolutionary basis. For like all other features of the human mind, they emerge from predispositions which are already in evidence among animals, especially the chimpanzees, our nearest relatives. Nevertheless, the form these take in humans is beyond anything that can be encountered among the animals.

The key predisposition for the quartet of cardinal qualities is to be found in the unique features of human learning. Learning is, of course, a common

63 Ibid, 156.

feature of all animal minds without which there could be no mind; as sheer behaviourist conditioning, it is to be even found in the lower animals with no minds. The higher animals exhibit higher forms of learning, but none of them is capable of the specifically human form.

Chris Frith calls this uniquely human capacity "learning by instruction" and distinguishes it from the higher animal mode of learning by observation. As he explains:

> There is some evidence that this special kind of learning through instruction may be uniquely human. While apes can learn by observation, there is little evidence for deliberate instruction, of the use and recognition of ostensive signals that instruction will be forthcoming. Learning through observation can certainly lead to the spread of knowledge through a group creating a form of culture, but this mechanism will be far less efficient than one based on deliberate instruction.[64]

Learning by instruction relies on deliberate teaching strategies, where "the mother explicitly demonstrates her knowledge and ensures that the infant is in a receptive state for acquiring this knowledge"; it is thus fundamentally different from the young animal observing its mother and acting likewise in a nascent form of mimesis. The latter can give rise to the evanescent culture of ape troupes, whereas the former serves the stable inter-generational transmission of knowledge and practice which is the mainstay of human culture, and therefore the social and psychological basis for the human mind. Hence, it is more than just a more efficient mechanism for the transfer of knowledge, but a mode of acquiring knowledge of a totally different kind.

Frith outlines the specific features of the teaching scenario which the human mother has to secure first in order to enable her infant to enter into the teaching situation in which knowledge will be passed on, in the first place the Adamic knowledge of the names of things:

> The mother first establishes eye contact with the infant, the mother then looks at and points at an object and the mother names the object. The first signal in the process, the mother looking at the infant, is not just a means of attracting the infant's attention, but is also an *ostensive* gesture. An ostensive gesture indicates that the signal that follows will be a

64 Chris D. Frith, "Social Cognition", in Colin Renfrew, Chris Frith and Lambros Malafouris, eds. *The Sapient Mind: Archaeology Meets Neuroscience*, (Oxford: Oxford University Press, 2009), 160.

> deliberate communication about something of relevance to the receiver… Infants show special sensitivity to the "ostensive" cues that signal the teacher's communicative intention to manifest new and relevant knowledge about a referent object.[65]

This is one of the earliest teaching situations, which, as all mothers intuitively know, requires all kinds of strategies to get the infant to pay attention and to grasp that something is to be learned. Much of this so-called scaffolding, as Vygotsky called it, was studied by Bruner and his associates at Oxford during his brief stay there, and was summed up in the acronym LASS or Language Acquisition Support System, in contrast to Chomsky's LAD or Language Acquisition Device, which according to Bruner is only half the story; but, as we have seen by reference to Donald, it is much less than that. Bruner's work was, of course, based on Vygotsky, but strangely Frith refers to neither of these classic authors on this subject.

Learning by instruction thus requires a teaching situation that calls for a uniquely human interaction between a teacher and a pupil in which each knows how to assume a role in relation to the other. In the first place, as in language learning, these are the quasi-natural roles of mother and infant; but later, as the child matures, strangers will be accepted in the teaching role. As Frith notes, "human infants learn about the world by observing their mothers, rather than strangers. However, at 14 months they will learn from the familiarized stranger and by 24 months strangers are used as a source for learning."[66] As the child develops and learns to speak, so the teaching situation changes and both teacher and pupil assume new roles; from then on instruction can become purely verbal and is no longer always ostensive. Later still, when literacy is acquired, new possibilities for self-instruction are opened up as the pupil learns to learn from books. But simpler forms of self-instruction are already present in learning almost right from the start when the pupil practices alone what has been taught. This inherently human capacity to repeat and correct a performance till some kind of perfection is reached Donald traces to the mimetic stage of human evolution. In the course of learning by teaching it takes ever higher and more sophisticated forms of autodidacticisim.

The human mind forms itself as it learns and is educated, that is, as the individual advances through ever more socially complicated and culturally more advanced teaching situations. It is through this pedagogic process, starting in infancy, that the child becomes conscious in the uniquely human form

65 Ibid, 160.
66 Ibid, 155.

of self-consciousness and acquires the rudiments of the other cardinal qualities of mind, namely, Ego, reason and free will. All these depend on learning through instruction, an inter-subjective social process without which neither human culture nor the specifically human qualities of mind could come to be. Frith points to this general conclusion:

> It is this ability to deliberately share knowledge that makes the human mind unique. The cognitive essence of this ability is to recognize that certain signals are deliberately emitted and intended to instruct. This kind of cognition is sometimes called meta-cognition. It requires that we reflect on our own cognition, in this case the process of expressing and receiving signals. Meta-cognition is intimately associated with self-consciousness.[67]

But there is still a long evolutionary road to be travelled from such rudimentary forms of self-consciousness, as well as Ego, reason and free will, to their achieved realization in mature minds, present in all societies and cultures. However, we can begin to get a glimpse from what primitive sources they derive and how they develop. Donald also makes this point when he states:

> But there is strong reason to adhere to an evolutionary approach to consciousness. Consciousness exists in some form in many animals, but the expansion of the powers of conscious processing is mostly about improving the capacity for learning, especially the capacity to integrate complex new material that must be assimilated by the brain.[68]

The kind of learning this must be and the kind of consciousness it promotes we have just studied by reference to Frith's notion of learning by instruction.

This culture of teaching on the social side and learning on the psychological side are inherently bound up with each other. The two sides are in mutual symbiosis; the one could not exist without the other; and as the one develops so does the other. Teaching and learning are the lynch-pin joining the two sides together. As the one side expands, so does the other reach more advanced human forms. This holds not only for self-consciousness, but also for the other

67 Ibid, 161.

68 Merlin Donald, "Consciousness and the Freedom to Act", in Roy F. Baumeister, Alfred R. Mele and Kathleen D. Vohs, *Free Will and Consciousness: How Might They Work?* (Oxford: Oxford University Press, 2010), 15.

cardinal qualities, Ego, reason and free will. Donald makes this point in terms of his conception of knowledge as providing a "world model":

> The ultimate product of wider conscious integration (this principle applies to many species) is an expanded "world model" by which an organism can guide action including its own future cognitive activity (thought is always potentially manifest in action). By implication freedom of choice is linked to the breadth of an organism's worldview. The reach of consciousness and freedom is graded and flexible.[69]

Like any other human quality of mind, free will is also an emergent outcome of evolution. It is closely linked to consciousness and assumes higher forms as consciousness expands to self-consciousness in humans. It is at its most rudimentary in infants and young children who are only aware of very limited possibilities of choice. But through learning and the opening up of an ever growing world of opportunities, bringing with it the need for ever more sophisticated decisions, so the scope for choosing is vastly enlarged, and the possibility of freedom increases. But as the possibility of freedom increases, so, too, does the possibility of unfreedom. For individuals can always succumb to or even deliberately choose to become thrall to other people or ideas or ideologies or habits or overpowering emotions or addictions or any other compulsions that will place them in bondage. Bondage or a surrender of will itself can also be willed. We shall study this paradox further in connection with the philosophy of Spinoza.

Free will together with reason, self-consciousness, and Ego are traditional philosophical concepts, each with a hoary past. Each has a long history behind it and much disputation as to what it really means and even whether it exists. It helps to understand what all this contention is about if we first outline their etymology and philosophical origins. There are those who prefer that such apparently antiquated terms be ditched in favour of newly minted scientific synonyms, but these neologisms are never completely equivalent. So we deem it better to keep the old terms that carry their history with them, for this guards us against falling into some of their old traps and subterfuges.

Reason is the oldest of the four; it originates in the key term of Greek philosophy, *logos*, first introduced into philosophy by Heraclitus. Later it was translated into Latin as *ratio*, from which our terms rational and reason derive. Logic comes directly from the original Greek. Both logic and rationality gained much of their currency from Aristotle. Aristotle defined Man as the rational

69 Ibid, 15.

animal, distinguished from the other animals in the *genus proximum characteristica specifica* manner of Aristotelian categorization. Reason is the universal attribute of Man, but it is a historical universal in that it has a history of developmental progression which in its latest stages becomes more or less coextensive with the history of science. Man becomes more rational as science progresses, and in that progress numerous stages and types of rationality can be distinguished. The full and comprehensive history of Reason has yet to be written.[70]

Much the same holds for self-consciousness, which is also a historical universal, that is, a universal attribute of Man, which has a history in the numerous stages of human cultural development. Different cultures manifest different forms of self-consciousness and personal awareness. The history of the specifically human emotions constitutes a large part of the history of self-consciousness. The etymological origins of the term as a theoretical concept are somewhat murky and obscure. The words consciousness and conscience are cognate terms deriving from the Latin *conscious* and *conscire*, possibly loan-translations of the Greek *syneidesis*, which assumed a theological meaning in the teaching of late Christianity around the time of St Augustine. The term *conscientia* was a medieval scholastic derivative standing for both conscience and consciousness indifferently. They were only clearly differentiated in modern philosophy from Descartes onwards when both terms assumed their modern meanings.

Ego is, of course, the ordinary Latin word for the first person singular pronoun. It gained its currency as a philosophical concept from Descartes' *Cogito, ergo sum* where the Ego is implied. However, from Descartes onwards it was paired with the term Subject, which is a much older philosophical concept deriving from the Latin *subjacere*, to throw under or underpin; which in turn was a translation of the Greek *hypokeimenon*, a term that featured in Aristotle's metaphysics as *hypokeimenon hyle* or underlying matter. The much later subject-object opposition of modern post-Cartesian philosophy has a long and convoluted background in medieval scholasticism based on Aristotle. To treat the Ego as Subject, that is, as the basic underlying ground of existence, was a fundamental metaphysical innovation that became decisive for modern philosophy where subjectivity became preeminent from Descartes to Husserl, as we shall show in Chapter 6.

Free will is also a concept of late Latin derivation planted in Christian theology and also coming to fruition around the time of St Augustine. It had not played any role in Greek philosophy prior to this point. It became crucial in

70 Harry Redner, *The Ends of Philosophy*, op. cit, Appendix 1.

interpreting the Biblical idea of a choice between good and evil associated with the myth of the Fall. St Augustine was the first major thinker to write extensively on will and to uphold free will despite his belief in predestination. All later discussions of free will start from there.

Obviously, such terms can no longer be invoked in anything like their traditional meanings, for as we have just shown, these are bound up with metaphysical and theological assumptions that are at odds with contemporary science and philosophy. Nevertheless, they must not be completely dismissed or discarded since they point to essential truths about human being. However, the four key terms do need to be revised and given a contemporary interpretation more in keeping with what science and philosophy now teach about human nature and the origins of the human mind.

In the first place, it is of primary importance to realize that the four fundamental concepts, defining attributes of the human mind, necessarily entail each other. This follows from the holistic principle that we previously invoked regarding mind, which is pre-eminently exemplified by the human mind. In its definitive state the human mind must be rational, self-conscious, Ego directed and have free will. And, furthermore, these attributes of the mind necessarily entail each other in that no mind can have any one of these attributes without the others, at least to some commensurate extent. No mind can be rational and not self-conscious; Ego directed and not have free will; or self-conscious without the other cardinal qualities, and so on.

But, of course, in the reality of human psychology with its vast diversity and disparity, this fundamental defining logic of the mind only holds good for normal minds, those that are sane and balanced in the psychological sense. This holds for the great majority of human beings in any given society; for taken overall most people have normal minds, since if they did not then human society could not function. Human evolution has seen to it that a basic normative standard applies to the mind, just as a similar one holds for the body. But naturally, there are exceptional cases that constitute a small minority in every society. The fact that they exist at all is dependent on the normal workings of society and on the normal human beings who constitute it. As we previously expressed it, abnormality is in some sense "parasitic" on normality.

In every society there are individuals who are mentally impaired or sick in all kinds of ways. There are imbeciles and morons who only have minimal rudiments of rationality, self-consciousness, Ego and free will. There are people suffering from the late stages of Alzheimer's degenerative disease who are also in a similar predicament. Yet we consider these to be impaired or abnormal human beings and not as sub-human or treat them as animals; we certainly do not allow them to be exterminated, as the Nazis sought to do with their euthanasia cleansing campaign. We do so on ethical grounds that are well justified. But purely logically considered, such people have diminished attributes of humanity. There are also many others where a lesser incapacity is evinced, as in people with severe mental problems or diseases who still retain some or most of the attributes of humanity but fall short in one or two.

However, as distinctive and uniquely human as the four qualities undoubtedly are, they are, nevertheless, evolutionary derivations of animal capacities. This is most apparent when we compare their human manifestations with their very rudimentary incipient origins in the mental abilities of the higher apes, especially among chimpanzees. Such creatures evince an elementary rationality in their high intelligence as shown by their learning capacities and problem solving abilities. Köhler's experiments on Gestalt insight in apes and Jane Goodall's observations on the social behaviour of chimpanzees in the wild have shown how close to humans they can come in this respect. Nevertheless, a huge difference still separates the two species mainly due to the fact that apes have no capacity to learn by instruction and so no access to language and culture on which proper rationality is based. Whether one considers this to be a difference in degree or in kind depends on one's general outlook on animals, as we saw previously.

Like rationality, self-consciousness is also a fundamental human trait, one that is lacking in animals who are merely conscious. Nevertheless, there is an evolutionary sequence that leads from animal consciousness to human self-consciousness. All animals to whom mind can be ascribed must *ipso facto* be also granted some degree of consciousness. Animals vary in the degree and kind of conscious awareness they evince. On the whole, predators tend to be very conscious, for such a hunter must stalk its prey, selecting and attending to one victim rather than all the others, be alert to its changing behaviour as it seeks to escape and even anticipate its future moves. Animals that hunt in co-operating groups are even more conscious because they must also be aware of the moves of their partners. However, even such animals lack self-consciousness for they are not aware of themselves as distinct individuals. But this is not altogether true of the higher primates who can identify others in their troupe as individuals and recognize themselves in mirror reflections, whereas most animals see such images as if they were other animals. However, even such apes do not have an awareness of their own identity on which self-consciousness depends. Humans do so because they enact symbolic roles in order to fit into a social environment in which their identity matters crucially to their behaviour which is rule governed and appropriate to their place in society.

The reason that they are able to do this is because they have an Ego or a subjective sense of Self. Animals have some sense of themselves as distinct from others in so far as they possess what has somewhat grandiloquently come to be called in psychology circles a "theory of mind". This means that the animal can know something of the minds of others, e.g. what the other sees, intends, threatens etc., and so can in this way react appropriately to others. Among apes this goes much further than in other species; furthermore, chimpanzees brought up in a human environment are even capable of some limited forms of symbolic self-expression. But as such experiments in bringing up chimpanzees have revealed, even the most intelligent among them, such as the famous Kanzi, still fall far short of the sense of self or Ego of small children. But what we see in apes shows us how our sense of ourselves as agents and subjects, which constitutes the Ego, derives from primitive animal origins.

A very similar evolutionary story can be told for free will. There is no doubt that animals have wills, they can strain and exert great effort when required to do so. They can even exercise considerable willed control of their appetites and emotions. Experiments in postponed gratification, especially with apes, show that animals do not always act directly on urges or in response to stimuli, but can deliberately restrain their desires and delay gratification for some extended time. But that is not enough to grant them free will. Only with humans is there a capacity to resist external and internal compulsions, both the external pressures and temptations coming from others as well as one's own internal drives and needs. This is what we now understand by the concept of free will. It is not what the traditional idea specifies, that spelled out in terms of metaphysical philosophy and theology; it is not a causeless cause activated by an immaterial mental substance on material bodies, as held by Descartes and other philosophers.

Having thus briefly sketched out the quartet of cardinal qualities of the human mind, we can next proceed to attempt to provide a scientific explanation for each one of them. They do stand in need of scientific explanation for otherwise they remain purely metaphysical and so have little bearing on the scientific account of mind. Hence, even free will must be explained in terms of the neurological and psychological sciences; which does not mean that it is either negated or abandoned in favour of determinism. Of course, this requires that the idea of free will must be redefined in a way that is consonant with the possibility of being accounted for scientifically, and this means in some senses of the term deterministically. Something similar is true of the other fundamental human attributes as well.

Rationality is dependent on the human capacity to learn through instruction and thereby acquire culturally mediated knowledge and participate in cultural life. This goes beyond the mere ability to speak and to fit in with the norms of one's society. There are some autistic human beings who can do this, yet are scarcely rational. Rather, rationality depends on an ability to engage in discourse and entertain discursive thought. Hamlet refers to this when he states, “He that made us with such large discourse, | Looking before and after, gave us not | That capacity and godlike reason | To fust in us unused!” (Act iv, iv). Reason is related to “large discourse”, which is obviously different from “small discourse” or small talk, a non-rational use of language. Large discourse involves “looking before and after”, that is, weighing up possibilities and consequences and thereby reasoning things out.

When we speak of rationality in respect of the everyday activities of ordinary people we are mostly referring to what is called practical reason, which Aristotle called *praxis* in contrast to *theoria*. It is the ability to set oneself a goal in the light of ones preferences, wishes, hopes, beliefs and values, and then to plan how to achieve that goal by setting out the means necessary to attain it, ordering the intermediate sequence of constituent acts or sub-units necessary to execute the whole action. Such goal-directed planning can involve quite trivial activities, such as how to get from here to there by the roads and means of transport available, or it can require consistent striving over a long time in attaining one’s large goals. In all such practical activities what has to be relied on is “discourse of reason”.

What Hamlet calls “large discourse” has a history; it is the succession of rational modes of explanation which begins as myth and ends as contemporary science. It begins in tribal societies at the stage of development that Donald refers to as oral-mythic culture. Such discourse also involves systems of classification that anthropologists study under the general rubrics of folk physics, folk biology, folk psychology, elementary technology and so on. People can be considered rational in so far as they are capable of making use of the rational standards of their society and apply them to current exigencies and concrete situations.

There are higher and lower forms of reason prevalent in different cultures and societies; and we can judge this on rational grounds. Thus we deem science to be a higher form of reason than myth when it is a matter of explaining the origin and nature of things, though myth might be preferable in other evaluative respects. In science, there is no ultimate, absolute truth, for all theories are hypothetical and provisional. Yet, going by the most developed and advanced forms of science, such as are current in advanced contemporary societies, we can judge the rationality of all previous forms of knowledge, such as the folk wisdom of pre-literate tribal societies, as well as that of the more historical societies of the ancient civilizations.

Higher forms of reason in society enable higher forms of self-consciousness in individuals. It is undeniable that a certain incipient level of self-consciousness can already be found among the higher primates. However, full self-consciousness only occurs among humans in the social context of language and culture. The difference between the two types becomes evident in a comparison of animal and human emotions such as Damasio has undertaken. According to Damasio, animals feel basic emotions, such as fear, anger, disgust, surprise, sadness and happiness, which humans also share. But humans also experience emotions that involve self-consciousness, such as embarrassment, shame, guilt, sympathy, envy, gratitude and admiration. All these emotions entail a sense of self in relation to others that can only be cultivated in a human cultural and social context. They involve the full panoply of human relations, norms, status roles, authority hierarchies and also the mastery of rational discourse. But self-consciousness can even precede discourse, for the child feels emotionally self-conscious even before it can fully engage in reasoning.

However, as Damasio notes, "social emotions are by no means confined to humans."[71] But there is, nevertheless, as huge a disparity between human and animal emotion, as there is between human and animal societies, although the former evolve from the latter, as Darwin was keen to emphasize and as Damasio, too, acknowledges:

> The nested incorporation of components from lower tiers is apparent. Think how the social emotion "contempt" borrows the facial expression of "disgust", a primary emotion that evolved in association with the automatic and beneficial rejection of potentially toxic food.[72]

Emotion is only one mental sphere where human self-consciousness distinguishes itself from animal consciousness. This happens right across the mind in all its faculties and domains. Donald discusses how this takes place in action as humans came to integrate their actions in ways that go beyond what animals are capable of coordinating:

> The key change is the extension of voluntary attention control into the domain of action, especially the control of limbs and the vocal tract. This change might qualify as the central defining characteristic of human conscious capacity. It not only made our perceived world larger, since it placed us squarely in the centre of it, but revolutionized skill. The perceptual ego-centre became an action-based control device exercising considerable discretion over the shape of action. This is a paradigm shift of

71 Antonio Damasio, *Looking for Spinoza*, op. cit, 44.
72 Ibid, 45.

> the first magnitude. In other mammals, with the possible exception of some apes, consciousness seems dominated by the perception of everything but one's own actions.[73]

This paradigm shift from animal to human action is the key to the move from animal brute will to human free will, as we shall presently see. But it is also, as Donald states, "the central defining characteristic of human conscious capacity", namely, self-consciousness. Donald expresses the difference between animal and human consciousness as one between what he calls level-2 and level-3 conscious capacity:

> But level-2 conscious capacity cannot manage much control over action, especially over the long run. It cannot even gesticulate. Level-3 awareness changes all that. Prior to it the self-conscious actor did not exist. Once the physical self entered the conscious world model, a whole new world of possibilities opened up.[74]

Among these possibilities were the self-conscious human emotions which we have just discussed and the forms of human perception which we dealt with in the previous chapter.

Gerald Edelman reaches an analogous conclusion regarding the two forms of consciousness. Animal consciousness, which he calls primary, only allows the animal to live in a "remembered present":

> Primary consciousness is the state of being mentally aware of things in the world, of having mental images in the present. It is possessed not only by humans but also by animals lacking semantic or linguistic capacities whose brain organization is nevertheless similar to ours. Primary consciousness is not accompanied by any sense of a socially defined self with a concept of past or future.[75]

Primary consciousness, shared by humans and animals, lacks self-consciousness that only come with "a socially defined self", which is the prerogative of human beings. It is a higher form of consciousness that "higher primates, to some

73 Merlin Donald, *A Mind So Rare*, op. cit, 196.

74 Ibid, 196.

75 Gerald Edelman, *Wider than the Sky: A Revolutionary View of Consciousness*, (New York: Penguin, 2004), 9.

minimum degree, are assumed to have, and in its most developed form it is distinctly human",[76] as Edelman goes on to state:

> In contrast, higher-order consciousness involves the ability to be conscious of being conscious, and it allows the recognition by a thinking subject of his or her own acts and affections. It is accompanied by the ability in the working state explicitly to recreate past episodes and to form future intentions. At a minimum level, it requires semantic ability, that is, the assignment of meaning to a symbol. In its most developed form, it requires linguistic ability, that is, the mastery of a whole system of symbols and a grammar.[77]

But to have what Edelman calls "a socially defined nameable self, and to have a concept of the past and future" is much more than a matter of language alone, as Donald contends. It is the outcome of the whole human cultural context which Donald calls "mythical culture". It is only in such a context that a socially nameable self can develop and the human individual become a Subject or Ego.

The Ego is perhaps the most crucial factor to human existence as such, that is, to the human mode of being. In a purely formal sense Descartes recognized this when he coined the *Cogito,* "I think, therefore I am". But he completely failed to investigate the precedent conditions that enable such an Ego to arise in the first place. He simply assumed it was already there, fully formed in the mind; he did not inquire as to its origins. He did not think this was necessary since thinking alone guaranteed this as a necessary consequence. It is true that as a matter of pure logic human thought entails an *Ego cogito*, but this only holds for discursive thought, which Descartes held to be the only kind. But since we now know, as Gestalt psychology shows, that there are also non-discursive modes of thinking, as in problem-solving activities some of which even animals can accomplish, it can no longer be assumed that the Ego can be deduced from thought as such. But to deduce it from discursive thought, as in the *Cogito*, is only to postpone the necessity of accounting for its origin and emergence. And that is synonymous with the origin and emergence of the Ego itself.

Where does the Ego come from and what explains its nature and role? We have, in fact, already indicated this in our account of self-consciousness. Self-consciousness in emotion and action has its origins in animal consciousness, in basic emotions and in animal agency. Analogously, the Ego in general comes

76 Ibid, 9.
77 Ibid, 9.

from the integrative functions that are already present in animal minds. As its most basic this is the so-called binding problem investigated on the neurological level by Francis Crick and Christofer Koch. It is the problem of "how does the brain coordinate and integrate the workings of its highly specialized parts to create the apparent unity of perception and thought that constitute the mind."[78] Koch and Crick believed that they have solved this problem at the neurological level by "the synchronicity of neural firing at the 40-Herz oscillations per second of assemblies of neurons."[79] However, David Hubel disagreed that any answer has as yet been found:

> The surprising tendency of attributes such as form, colour and movement to be handled by separate structures in the brain immediately raises the question how all the information is finally assembled, say, for perceiving a bouncing ball. It obviously must be assembled somewhere, if only at the motor nerves that subserve the action of catching. Where it is assembled, and how, we have no idea.[80]

This is the modern form of the same problem first raised by Aristotle of how the sensory objects from the different sense organs come together, which he tried to solve by postulating a common sense. It does not seem as if all that much progress has been made in solving it since Aristotle.

The unity of human self-consciousness established by a Subject or Ego is an issue of still higher degree of difficulty, and almost nothing is known about it in neurological terms. Descartes believed that he could locate the Ego in the pineal gland, but thus far, despite a vast increase in imaging capacity, no location of it in the brain has been found. Perhaps a clue to how the integrating functions of the Ego are carried out in the brain might be found by a study of short-term working memory. Donald points in that direction when he states "that the attribution of conscious capacity to higher mammals hinges on short-term working memory, which is where controlled processing as we know it in humans probably begins."[81] He points out that "theories of consciousness tend to combine research in selective attention and short-term memory."[82] And he goes on to add:

78 Quoted in John Horgan, *The End of Science*, op. cit, 22.
79 Ibid, 22.
80 Ibid, 22.
81 Merlin Donald, *A Mind So Rare*, op. cit, 185.
82 Ibid, 186.

> In this sense, working memory is equivalent to the older concept of a central processor, with most of the basic executive capabilities we identify with advanced mammalian controlled processing. These include evaluation, selection, problem solving and response choice.[83]

However, Donald hastens to qualify this assertion by stating that "the concept of a central processor is a strictly cognitive one, and the evidence for it is cognitive not anatomical… it is at the top of the control hierarchy."[84] In other words, "it does not imply that there is a single neural locus, a 'consciousness module' somewhere in the brain."[85]

The human Ego lies somewhere at the peak of the evolutionary ascent, which begins at the lower levels of the unity of consciousness among the higher mammals, especially the social ones like apes. Among these we find "the ability to cultivate and remember individual relationships within a working social ground" and "a related skill, variously called mindreading, perspective taking or having a theory of mind…"[86] But the human Ego vastly exceeds these capacities for it has to function in a human social context with its myriads of symbolic meanings that have to be retained over long and short periods.

In contrast to its role in long-term memory, in short-term memory the Ego plays a vital role in providing a coordination of all the current on-going activities in mind and body at one and the same time. The Ego provides what Damasio calls "orientation" which makes all these concurrent activities cohere together, as he explains:

> The sense of self introduces, within the mental level of processing, the notion that all the current activities represented in the brain and mind pertain to a single organism whose auto-preservation needs are the basic cause of most events currently represented. The sense of self originates the mental planning toward the satisfaction of those needs.[87]

The role of the Ego is keeping all activities on track and supervising them so that the aim or goal they are designed to fulfil is kept constantly in sight. The Ego leaves some of the component processes on automatic hold and devotes

83 Ibid, 186.
84 Ibid, 117.
85 Ibid, 117.
86 Ibid, 129–30.
87 Antonio Damasio, *Looking for Spinoza*, op. cit, 208.

its attention to others, switching its concentration from one thing to another. Adults can even divide their attention from one task to others in multitasking activities of which children are barely capable and animals not at all. But this requires an extremely developed Ego.

Operating in terms of long-term memory, the Ego ensures the identity of the individual and its persistence over time, most frequently over a long life-time. Short or even long breaks in self-consciousness, such as while asleep or periods of coma and other forms of unconsciousness, generally make no difference to the persistence and perseverance of the Ego over time. On awaking or coming out of a coma one usually comes back to oneself with all ones capacities and memories intact. Only severe brain injury or senility or mental illness can disrupt this sense of self-identity. But for this to be possible, the human Ego has to have formed itself through the mastery of the vast number of accomplishments that go into the socialization of the human child that enables it to acquire an Ego in the first place.

People are not born free, and neither are they born with an Ego. At birth, the neonate is merely a human animal that has only the potential to acquire subjectivity. It does not even possess a human brain, properly speaking, for that only constitutes itself in the process of acquiring an Ego through socialization, language learning and enculturation. How, this happens is the mystery of ontogenesis which is no better understood than that of phylogenesis; and since Ernst Haeckel in the nineteenth century the temptation has always been there to confuse the two and to argue that ontogenesis recapitulates phylogenesis. But this is an egregious error in biology as well as in educational theory.

How ontogenesis takes places is the subject of numerous rival theories of development of the infant and child. At one extreme are those that focus on innate genetic factors and at the other are those that place more weight on socialization; between these are theories of epigenetic development that seek to combine both nativist and empiricist factors working in tandem with each other. Thus at the nativist pole are theories such as Chomsky's account of language learning as based on an innate Language Acquisition Device (LAD) which his followers such as Jerry Fodor and Steven Pinker have developed into a modular account of the mind to which we shall return in Chapter 7. At the empiricist pole lie the theories of child development that emphasize interaction with other human beings in a social context, such as Lev Vygotsky's "zone of proximal development" and Jerome Bruner's Language Acquisition Support System (LASS). At the intermediate position there is Jean Piaget's theory of stages of maturation that are epigenetically determined: sensory-motor, preoperational, concrete operational and formal operational; this was subsequently taken up and modified by his student Annette Karmiloff-Smith.

Every major school of psychology has its theory of maturation and Ego development. This is, of course, pre-eminently true of the Freudian school, for not only Freud himself, but also his daughter Anna and her great rival Melanie Klein were much concerned with infant maturation and personality formation. The role of the Ego or *das Ich* in the dynamics of repression and the unconscious is an extremely complex topic since Freud expounded it at various stages in his career in different ways. A mere summary account in Laplanche and Pontalis covers 13 pages.[88] The celebrated formula "*Wo Es war, soll Ich werden*" (where it (Id) was shall I (Ego) become) has been interpreted in numerous ways at great length. The various stages of libidinal development, such as the anal, oral and genital, and their role in the developing Ego, have also been extensively commented on. Unfortunately, we cannot go any further into the question of how the Freudian Ego relates to that discussed in other developmental theories or philosophies, nor can we enter into their varied accounts of will and freedom.

People are not born free, they attain freedom or, better put, they have freedom thrust upon them by the very process of becoming human. For as they mature and develop an Ego, they also attain free will in their actions and activities. They acquire an ability to resist compulsions; at first simply by postponing gratification; later by other means of self-control. And this capacity to resist compulsions, either external or internal, is the key to human freedom. External compulsions are those which others impose on one, and they come in an almost inexhaustible variety of forms. Internal compulsions are those which derive from one's own bodily or psychological self, be it appetites, habits, dispositions or character traits. The ability to act against both forms of compulsion, to act in the teeth of the pressure emanating from the outside or the inside, is the mark of the free individual.

Free will can only be ascribed to a free agent, that is a person who is a self-conscious Ego or Subject capable of being rational in accordance with the norms of a given culture and society. A free agent is, therefore, also a responsible agent on whom all kinds of social or personal expectations can be placed and of whom corresponding demands can be made. This is the basis for the legal order, for moral life, and for all other social practices in which people are held responsible for their deliberate acts.

How all these factors are defined and judged varies from culture to culture. For as with all social matters, there is also a historical and developmental dimension to the ascriptions of responsibility, and, indeed, to free will. As

88 J. Laplanche and J.B. Pontalis, *The Language of Psycho-Analysis*, trans. Donald Nicholson-Smith (New York: Norton, 1973), 130–43.

societies become more developed and move away from their primitive origins, they also become more sophisticated in their legal and moral conceptions of will and freedom. This means that the exercise of will also assumes a higher form and becomes subject to more refined norms of judgement. What is responsibility, what is deliberate action, what compulsions a person should be able to resist and therefore what can be expected from a free agent, all these matters are redefined as society progresses to higher legal and moral standards. Thereby, free will is also adjusted accordingly, since this, too, is a historical matter. Historically considered, there is also a relation between free will and political freedom in so far as the mandating of free will makes for individualism; and that in turn bears on politics, particularly on the rise of liberalism and democracy.

Free will and freedom in general is not an all-or-none matter, but one of degrees and kinds. Some acts exemplify much greater freedom than others. Trivial acts, such as lifting or not lifting a finger, where nothing hangs on either choice, are only free to a purely nominal degree. But the more an act matters, and the more constraints it encounters, the greater becomes the freedom of the will exercised in carrying it out against such compulsions. Thus, for example, the refusal to reveal important information or betray others when subject to extreme pressure – torture of either a physical or psychological kind and usually in such cases both at once – is under normal circumstances an expression of a high degree of freedom and a paradigmatic instance of the exercise of free will.

To succumb to compulsion means, in effect, to lose a commensurate degree of freedom and no longer be able to exercise free will to that extent. This invariably has consequences for all the other attributes of free agency. It results in not being able to act fully rationally, and in time even to think rationally, as is apparent in the lives of addicts. Inevitably, too, this will have severe psychological consequences for the stability of such a person's Ego, for weakness of will in time leads to loss of control in other respects, for example, no longer being able to exercise restraint on desires and other mental impulses. Thus, the hopeless addict, whose freedom is severely constrained by his drug dependency, ceases gradually to be a coherent Self or Subject in the way necessary for normal human living. Such a person becomes less capable of self-consciousness as he or she loses any sense of shame or even fear in extreme cases. In aggravated stages of dependence, the individual ceases to be a free agent altogether and can no longer be held responsible for his or her acts.

However, one must not take the general requirement of rationality for free agency as requiring the rational motivation of every free act, for that would unduly restrict what is meant by freedom. Not all freely willed acts have to be rational in this narrow normative sense. There are not only reasons of the head, but also, as Pascal put it, reasons of the heart that reason knows not of. One freely chooses to love someone even though one relies on nothing that could be called rational for doing so, since love is a matter of heart not head. It is possible for blind passion to become obsessive and a compulsion that entails the loss of freedom; but normal love is the opposite, it is an expression of emotional freedom in the giving of oneself to another.

Having reasons for doing or not doing something does not guarantee that it is freely done, since the supposed reasons might turn out to be mere rationalization, for as Hamlet puts it, "reason panders will". Hamlet is a case study of someone whose reasons for acting or not acting are continually at odds with what he actually does or fails to do. Thus in one crucial scene he cannot bring himself to kill Claudius even though he has the perfect opportunity for doing so and thereby fulfilling his quest as avenger. Later he berates himself, saying "I do not know | Why yet I live to say, 'This thing's to do' | Sith I have cause and will and strength to act" (Act IV, iv). But Hamlet deceives himself, for as the scene demonstrates, he does not have the will to act. This is the crux of criticism as to why Hamlet delays. Freud and his disciple Ernest Jones had psychoanalytic answers based on the Oedipus complex. But perhaps the answer is conveyed in Hamlet's prior retort to Guildenstern that he cannot "make you a wholesome answer – my wit's diseas'd" (Act III, ii). Diseased wit or will seems to be the problem and the death of his father, whom he excessively idolizes, the cause. But however one diagnoses Hamlet's predicament, it is almost a character study of a person being in bondage to something and he himself does not know what it is, and is thereby unable to act freely. We are now in a position to bring together Hamlet and Spinoza as Damasio does in his book, which we adverted to in the *Prelude.*

The account of free will we have presented is quite in keeping with Spinoza's view of freedom and bondage. But it goes against the metaphysical and theological view of acts of free will as uncaused causes, according to which such acts are not determined in any way and are not to be causally explained; yet they in turn act as causes on the body in enabling limbs to move and acts to be performed. Such a conception of free will is a philosophical velleity deriving from a pre-emergentist view of mind, such as Cartesian Dualism. According to Descartes – who was only expressing a traditional metaphysical-theological doctrine in physiological terms – the will as pure thought acts on the pineal gland in the brain, which in turn causes the bodily limbs to move in such a way as to execute the intended action. Contemporary Dualists have sought for scientific sounding explanations of this uncaused thought acting as a cause on the brain. They have called it downward causation. Some physicists, such as

Roger Penrose, have invoked the laws of Quantum physics to show how it is possible for acts of will to affect the firing of neurons. We shall seek to refute such accounts in the next chapter.

However, in a preliminary way we might indicate that acts of the free will, like all other acts of will or any other mental entities, are emergent properties. They emerge out of deterministic conditions in the physical world, that is, from the brain-body-environment complex. Hence, in that emergent sense they are determined, though they are not identical with the material factors that give rise to them. They can also be explained as effects arising from all kinds of other types of causes, such as psychological, sociological, Freudian, ideological, and so on. Hence, they are not causeless causes, but scientifically explicable emergent phenomena of the human mind. Nevertheless, they can be considered free in the Spinozist sense of freedom.

The metaphysical-theological conception of free will is precisely what Spinoza sets himself against. He encountered it in his time pre-eminently as expounded by Descartes, and it is, therefore, against Descartes' account that he turns. He states quite decisively that "the Body cannot determine the mind to thinking, and the Mind cannot determine the Body to motion, to rest or to anything else (if there is anything else)."[89] Later he puts it even more strongly that "the forces of the Body cannot in any way be determined by those of the Mind."[90] Consequently the metaphysical or theological conception of free will is quite incoherent, if not meaningless:

> So it follows that when men say that this or that action of the Body arises from the Mind, which has dominion over the body, they do not know what they are saying, and they do nothing but confess in fine-sounding words that they are ignorant of the true cause of the action, and that they do not wonder at it.[91]

Spinoza directs himself explicitly against Descartes' Dualist account of the action of the will on the pineal gland in the brain. He argues that there can be no such action, for the will does not act as a cause on the brain any more than the brain can act as a cause on the mind. As we have already argued at length, the relation involved is one of emergence, which is logically much more

89 Baruch Spinoza, *Ethics,* Part 3, proposition 2, *The Collected Works of Spinoza*, vol. I, edited and translated by Edwin Curley (Princeton: Princeton University Press, 1985), 494.

90 Ibid, 597.

91 Ibid, 495.

fundamental than causation. We shall see in the next chapter what is the logic of emergence and how it relates to causation.

Spinoza was, of course, unaware of emergence, nor did he have any knowledge of evolution or any of the other sciences from which the idea arose. Such notions were far beyond the science of his time, when Cartesian mechanistic physics was in the ascendency and biology hardly existed as a science. Nevertheless, his account of the body and its actions is quite compatible with contemporary science and with the kind of emergent interpretation of free will here set out.

Spinoza provides an account of the development and growth of mind and, therefore, of the emergence of will and freedom, in terms of whether a body is capable of few or many things. He states it as a basic proposition that "he who has a Body capable of a great many things has a Mind whose greatest part is eternal."[92] The term "eternal" should not trouble us too much for Spinoza is quite explicit that it has nothing to do with duration, for only what he calls common men "confuse it with duration, and attribute it to the imagination, or memory which they believe remains after death."[93] The mind that arises out of a body "capable of a great many things" is a mind that "has a power of ordering and connecting the affections of [the] Body according to the order of the intellect, and consequently… of bringing it about that all the affections of the Body are related to the idea of God."[94] Provided we are not too troubled by the term "God", which Spinoza considers synonymous with Nature, *Deus Sive Natura*, this more or less matches our view of free will as a developed capacity for acting against compulsions, or the body's affections, as Spinoza would view them. Thus the capacity to exercise free will belongs to a body capable of many things, namely, to a mature individual with intact mental faculties.

By contrast to such a person, the mind of a mentally deficient person or a child and certainly of an animal is not free because the body of such a being or creature is only capable of very few things: "he who, like an infant or child, has a Body capable of very few things, and is very heavily dependent on external causes, has a Mind which considered solely in itself is conscious of almost nothing of itself, or of God, or of things."[95] Namely, such an undeveloped or limited mind is in thrall to compulsion and is not self-conscious, rational nor free. Freedom comes with human maturation, thus "in this life, then, we strive especially that the infant's Body may change (as much as its nature allows and

92 Ibid, 614.
93 Ibid, 611.
94 Ibid, 614.
95 Ibid, 614.

assists) into another, capable of a great many things and related to a Mind very much conscious of itself, of God, and of things."[96]This is also, of course, a body possessed of the freedom of will; which in no way contradicts the fact that this very same body is bound by the laws of nature and, therefore, in that sense also determined. But this very fact is a condition of its freedom, for if it were not a natural body capable, through growth and maturation, of performing a great many things, it could neither acquire a mind nor exercise free will. Freedom must be based on a form of natural determination governed by the development of the body – that is the lesson of Spinoza's teaching.

Spinoza was, of course, severely limited by the science of his time in accounting for how such natural determination takes place. And he was consciously aware of this limitation:

> And, of course, no one has yet determined what the Body can do, i.e. experience has not yet taught anyone what the Body can do from the laws of nature alone, insofar as nature is considered to be corporeal, and what the body can do only if it is determined by the Mind. For no one has yet come to know the structure of the Body so accurately that he could explain all its functions… Again, no one knows how or by what means, the Mind moves the body, nor how many degrees of motion it can give the body, nor with what speed it can move it.[97]

It has taken over three centuries for science even to begin tackling these problems that Spinoza set for it, and though very much more is now known than in his time, as the previous chapters have shown, there is still a long way to go in providing any satisfactory answers.

But at this point we must leave science for philosophy, for the issue now is not to review what we know scientifically but to interpret what it means philosophically. Spinoza's philosophy provides a pointer and guide to such an interpretation. We shall be returning to it again in Chapter 6.

96 Ibid, 614.
97 Ibid, 495.

PART 3

The Philosophy of Mind

CHAPTER 5

On the Logic of Emergence

1 Emergence and Causation

We can now return to Hamlet's assertion that Man is the quintessence of dust armed with somewhat more up-to-date scientific information. Not that it is by any means conclusive knowledge, for, as we have seen, the science of origins, especially that of mind and humanity in general is constantly changing, hence what it means to say that Man is "the paragon of animals" is also changing as we discover new facts about our animal origins. Nevertheless, we are much better informed than was Hamlet, now we can base our view of Man and the human mind on much more reliable knowledge than he had at his disposal. Accordingly, our approach is very different from that dreamt of in Hamlet's philosophy.

Hamlet's philosophy was deeply immersed in the old world view. He implicitly invoked the Golden Chain of Being model which placed Man at the centre of creation and the earth at the centre of the universe. It is this kind of view that is put in the mouth of Ulysses at the opening of *Troilus and Cressida* and that the contemporary magus Robert Fludd would continue to expound for decades later, as in his *Utriusque Cosmi Historia* (1617–19) whose popularity in Paris so upset Mersenne and Descartes. They sought to counter it with their new mechanistic philosophy based on the Copernican model of the world. Hamlet's philosophy was based on the previous medieval picture of a closed universe, derived from Aristotle and Ptolemy, which predates the open universe of Bruno, Kepler, Galileo and Descartes, namely, what we now call modern science. This does not mean that it is mythological or completely false. As we shall see, there are aspects of the Aristotelian philosophy of mind which fit contemporary scientific knowledge much better than Descartes' philosophy of mind or that of his successors, our contemporary Cartesian Materialists and Cognitivists.

We have interpreted Hamlet's view of Man as the quintessence of dust in terms of the five stages of emergence starting from the origin of matter at the start of the universe. These we can take to be the fields of science that cover the hierarchy of matter: the sub-atomic, solid-state, living, sentient animal and human. We might call this sequence a contemporary Great Chain of Being, which we designated as a power law of emergence from E^1 to E^5. The problem on which we were most intent is that of establishing in scientific and philosophical terms "what a piece of work is Man". To do so we relied on the notion of "emergence", especially as this functions within Darwinian evolution. Above all, we studied the emergence of human mind from animal mind; and to understand that we backtracked one stage further and also considered the emergence of animal mind from animal life as such. In our symbolic terms, it is the relation between E^5 and E^4 that concerned us most, but we also had to glance back briefly at the relation between E^4 and E^3.

In the previous chapters we reviewed the sciences of emergence, so now we will proceed to review the philosophies of emergence, starting with their historical development. Emergence has had a punctuated evolutionary history in two major periods, the first from the second half of the nineteenth century till the early 1930s, or roughly speaking, starting with J.S. Mill and finishing with Samuel Alexander and C.D. Broad.[1] Then it fell into abeyance for almost half a century and was only revived once again in the 1970s beginning with the neurologist Roger Sperry and the philosopher Donald Campbell. Since then the philosophies of emergence seem to have gone from strength to strength; though there has also been much opposition, scientists complain that it still smacks of mysticism and philosophers that it contravenes basic deterministic principles of scientific explanation, by which they mean that it runs afoul of reductivism.

However, there is one point to which less attention has been paid by both philosophers and scientists, and that is what might be considered the logic of emergence. This concerns the relation between that which emerges, which can be called the emergent, and that from which it emerges, which can be given the more technical designation of the emerger. The emerger is the base or ground which gives rise to the emergent, whatever that may be. Thus the emerger might be the elementary quantum theoretic basis of solid-state matter; or it might be the brain, body and environment which is the material basis of mind, viewed as an emergent entity, phenomenon or property. In between these two extremes there are all kinds of other emerger-emergent relationships, which, as we have argued previously, are all different and cannot be subsumed under a general theory of emergence applicable to all scientific realms.

Nevertheless, despite the impossibility of a general theory of emergence, there is a general logic to emergence. The logic of emergence is that of an indissoluble non-identity. This means that the emerger and emergent cannot be dissociated; but they cannot be identified either. They are distinct but not separable. If two things can be separated then the logic of emergence would break down and the concept of emergence would not apply. And likewise, if they

1 See Paul Humphreys, *Emergence,* op. cit, Chapter 3.

could be identified then the logic of emergence would also fail and emergence would equally not apply. Hence, the fact that the concept of emergence applies in a given case means that the logic of emergence holds and this entails an indissoluble non-identity between an emerger and emergent.

However, the fact that the concept of emergence obtains in a given case is ultimately, in a very broad sense of the term, an *empirical* matter. It can in principle always be falsified and shown to be otherwise, namely, not a case of emergence at all but something different, as we shall illustrate with many such conceivable counterfactual instances. But once it has been established that emergence does hold, the relation that is thereby posited between the emerger and emergent, the two sides of emergence, is no longer empirical but necessary. It is an intrinsic, not an extrinsic relation, that of an indissoluble non-identity. These features of the logic of emergence will become clearer as we consider specified cases to which it applies.

To illustrate the workings of the logic of emergence, we can show that the inseparability condition of the logic of emergence makes it philosophically impossible to argue for qualia or to represent the problem of consciousness as a "hard problem" in David Chalmer's sense.[2] For if emergence holds in a given case, such as that of the human brain-body-environment complex, then the very same physical conditions must necessarily generate the very same mental phenomena, since the emerger and emergent are inseparable. It is not logically possible for emergence to apply to two identical physical complexes, and yet for one to generate mind or consciousness and the other to fail to do so.

This is only possible on a Dualist conception of the mind-body relation, for according to this metaphysical view the two are separable and logically independent of each other such that the one side need not entail the other. Hence, it is possible for Dualists to assert that the very same physical conditions can in one humanoid animal be accompanied by qualia or "consciousness", conceived of in the Dualist manner, and in the other identical one fail to be accompanied by any such qualitative phenomena. Imaginary creatures of the latter type are colloquially known as "zombies". Hence, such zombies can only exist on the basis of an implicit Dualist metaphysics of the mind-body relation; they are not possible on any emergentist view.

It follows from this that Chalmers's so-called "hard problem" is only a problem on an implicitly, usually hidden, Dualist assumption that goes counter to the inseparability of the mental and the physical. Furthermore, the interpretation of consciousness as qualia makes it epiphenomenal – namely, it is seen

2 David Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, (New York: Oxford University Press, 1996).

as separable from and as irrelevant to the constitution of the brain-body-environment complex; and this also is a logical impossibility on the emergentist approach, since if the emergent were altered then necessarily the emerger must change as well; that is, if consciousness is not there, its physical basis cannot be there either. Both sides of emergence are inextricably tied to each other, the one cannot be epiphenomenal in relation to the other. Hence, the qualia view of consciousness only makes sense on the presupposition of a Dualist metaphysics. If that is dispensed with, then so must the whole idea of qualia and Chalmers' "hard problem" along with it.

But before we can proceed in explicating the logic of emergence any further, there are a number of misconceptions about it that need to be clarified. Emergence is most often confused with causation. Those who invoke emergence take it for granted that the emerger in some sense or other causes the emergent. What invariably comes under dispute is whether the emergent also causes the emerger. The former is known as bottom-up causation, and the latter as top-down causation. Few dispute bottom-up causation; but top-down causation is the source of furious arguments among scientists and philosophers, with some insisting that it goes against the laws of physics and others just as determinedly denying this. There are literally scores of papers and a number of books devoted to this topic. We shall seek to show that this argumentation is misconceived because emergence is fundamentally not a causal matter.

The nexus between the two sides of emergence, the emerger and emergent, cannot be causal because it is logical. The causal relation between any two entities, the cause and the effect, is an empirical matter; but this is not so in the relation between the emerger and emergent. What is a quasi-empirical matter, for it is subject to disproof, is whether the concept of emergence applies to a given system or situation. Once it can be established that it applies and that there is no other way of accounting for or explaining the case, then the relationship it sets up between its two terms is no longer empirical but logical. Hence, it cannot be causal.

Of course, when we speak of the application of emergence to systems or situations as being quasi-empirical and therefore subject to possible falsification, we do not mean this in a narrowly factual sense. One cannot falsify emergence by experiment. However, in a wider sense it is possible to disprove and so falsify the claim that emergence applies in a given case. This can be accomplished if an alternative non-emergent explanation can be given; it might be possible to show that something held to be a case of emergence is really not so at all and to do so on a number of grounds. It may be that science advances to the point where what we now take to be an emergent system or phenomenon comes to be explicable in a purely reductivist manner. Or it may turn out that what we now take as the emergent property based on an emerger turns out to be a separable entity after all, and so capable of existing independently of its emerger basis. In the first of these counter-factual cases it is the non-identity clause of the logic of emergence that is defeated; in the second, it is the inseparability clause that fails.

We cannot now know the future of science, so we cannot tell which phenomena we now take to be emergent will remain so in the future or prove to be capable of reductive elimination. However, no matter how limited and narrow the scope for emergence proves to be in the future, it is inconceivable that it should be completely eliminated; so that some scope for such an account will always remain. It is in principle impossible for all explanations to be reductive and that everything should be deducible from the one kind of fundamental entity, whether this be atoms or quarks or strings. It is impossible for a Theory of Everything to be in fact the explanation of *every thing*, for that would mean that everything could be deduced from the one set of equations. It follows from this that there must always be scope for non-reductivist explanations, which will in fact be emergentist ones. And where an emergentist explanation holds the logic of emergence, that of indissoluble non-identity, also obtains.

It is inconceivable to us now that the logic of emergence should not apply to the mind-body relation as we have previously set it out. By the term "inconceivable" in this context we do not mean unimaginable, for it is all too easy to imagine situations in which emergence would break down, and we will explore one such imaginary scenario presently. Rather, inconceivable means that on scientific grounds it is impossible to assume that such imaginary counterfactual things could happen. For if they did happen, this would make scientific explanation impossible. The very possibility of science guarantees the validity of emergence in most situations to which it now applies, and most surely to the mind-body relation. We shall see in the next section how this line of argument works itself out in relation to the emergence of mind.

However, before we can demonstrate this in relation to the main topic of concern, that of the emergence of mind, the most complex and difficult matter of all, we will first refer to some much simpler and straight-forward cases. Even though there is no general theory that covers all cases of emergence, yet the same logic of emergence applies in very basic systems as in ultra-complex ones. In this respect complexity makes no difference. Indissoluble non-identity holds equally for water as for mind. The former we have already dealt with in Chapter 2 in the discussion of hydrodynamics and statistical mechanics. At the risk of a certain degree of repetition, we can once more turn to water in order to provide a simple example of how the logic of emergence works in practice.

Water is an emergent property of the interaction of vast clouds of molecules made up of hydrogen and oxygen atoms, as inscribed in the chemical formula "water is H_2O". By the logic of emergence, this means that water and the atoms of hydrogen and oxygen are indissolubly linked to each other; it is not possible to have the one without the other, for wherever masses of such molecules arise, so, too, does the substance water; and wherever there is water, there are also these multiples of molecules. In a scientific sense of entailment, the one entails the other. But the two cannot be completely identified for water, as Anderson and many others have argued, has all kinds of physical properties, such as laminar flow, turbulence, density, viscosity and many more that cannot be derived from the known properties of its constituent atoms nor can it be deduced from the equations governing their quantum behaviour. In other words, there is a non-identity between water and its chemical composition; the word "is" in the chemical formula "water is H_2O" must not be read as if it were the "is" of identity or mathematical equivalence, as if it said that water is equivalent to H_2O. What the chemical formula asserts is a much more complex relation that follows the logic of emergence.

The non-identity of water and its constituent molecules is imaginatively illustrated in a *Gedankenexperiment* devised by Steven Weinberg which we have already quoted previously:

> If you knew everything about water molecules and you had a computer good enough to follow how every molecule in a glass of water moved in space, all you would have would be a mountain of computer tape. How in that mountain of computer tape would you recognize the properties that interest you about water, properties like vorticity, turbulence, entropy and temperature? There is in the philosophical literature a term, emergence, that is used to describe how, as one goes to higher and higher levels of organization, new concepts emerge that are needed to understand the behaviour at that level.[3]

On this issue of emergence, Weinberg and Anderson, two former adversaries, are in agreement. Anderson contends that it is not the case that water is nothing but hydrogen and oxygen atoms. It is that, of course, but it is also something more, for this, too, is a case where the whole is more than the sum of its parts.

3 Steven Weinberg, "Newtonianism, Reductionism, and the Art of Congressional Testimony", op. cit, 151.

However, the fact that the logic of emergence holds for water or that, indeed, water is an emergent phenomenon arising out of clouds of H_2O molecules is fundamentally an empirical fact. All kinds of scientifically weird and impossible situations can readily be imagined in a vast variety of *Gedankenexperimente* that would force the conclusion that water is not an emergent phenomenon or that if it is, it does not derive from hydrogen and oxygen atoms but from other ones or not from atoms at all but from some kind of plenum.

It is conceivable that Descartes was right and that water derives from vortices, swirls and eddies in some primordial ether; or perhaps even that Thales was right and that water is that elementary matter itself from which everything else derives. It is also conceivable, as Hilary Putnam imagined, that on some other planet a substance that behaves just like our familiar water was composed at the atomic level in some quite other way. In all such imaginary situations the reduction of water to the chemistry of H_2O would prove false, and so, too, *a fortiori*, would be the truth that water emerges from H_2O. Hence, the concept of emergence as previously set out would no longer apply to water, or whatever else we called this substance, and the logic of emergence would break down.

However, if the logic of emergence does apply to water and it is empirically true that water is H_2O, then no causal relation obtains between the emerger and the emergent, the microscopic molecules and the macroscopic substance. The molecules do not cause water; and water cannot causally affect the molecules or interact with them in any such way. The emerger and emergent inhabit, as it were, two parallel physical domains of science that are synchronically congruent. The molecules exist in the physical realm of atomic matter, while water exists, in the very same space and time, in the physics of solid-state matter. Diachronic emergence does not come into it for this is a very elementary form of emergence.

The issue so fiercely debated in the philosophy of mind whether there is downward causation or whether only upward causation has currency does not make much sense when applied to such a simple case of emergence as water. Does water cause the molecules of hydrogen and oxygen to exist, or to move or to arrange themselves in some way? Clearly none of this is possible or even scientifically conceivable where the emergence of water is in question and where the laws of physics apply. If this were not a case of emergence or if some other imaginary physics was in force then this question could make perfect sense. But water being what it is and the laws of physics being as they are, it makes no sense in science or philosophy to ask such questions. And the same holds for the inverse question whether hydrogen and oxygen molecules cause water. Of course, these molecules explain water, but not all scientific explanation is causal, it can also be emergent. This is the reason that so much of this ongoing argumentation concerning downward causation is misconceived and results from a failure to understand the logic of emergence.

The philosophical position known as "strong emergence" is based on the premise of downward causation. As Paul Humphreys explains it, "an entity, such as a property or event, is strongly emergent only if it has the ability to causally influence entities at a lower level by means of features that belong to the emergent entity but are not found in structure aggregates of lower level entities."[4] This is fundamentally denied by the exponents of "weak emergence" who contend that "as new patterns emerge, the fundamental causal processes remain those of physics."[5] The main issue between them is whether downward causation is possible. Nobody seems to object to so-called upward causation. The question whether a causal relation operates in either direction, down or up, is not raised. But this is the nub of the matter. Is emergence a causal relation at all? We have argued that it is not, but something quite other, having a different logic to that of causation.

This is not to say that the concept of cause and effect does in no way apply to water or even to its constituent molecules. Of course it does and is in constant use both in scientific and common sense discourse. We speak of the causes and effects of waves, flows, turbulence, vortices and all the other phenomena of water. On the somewhat higher scientific level, as in electrolysis, we explain how electrical currents in water break the molecular bonds between the hydrogen and oxygen atoms and cause these to be detached as gases that can be separately collected. With the equipment now available in many laboratories, it is even possible to cause single molecules to be released from tiny drops of water and to be used to bring about all kinds of effects in other respects.

Cause-effect discourse and the explanations it affords is in constant use in relation to water. But it, in turn, is explained by the emergent law that water is H_2O. It might be put that cause-effect explanations of specific events or phenomena in regard to water are construed on the back of the emergent nature of water as a substance. Generally, cause-effect accounts are invoked in practical situations when something has to be achieved or prevented; in relation to water, when it is more a matter of practical hydraulics than theoretical hydrodynamics. In general in science, cause-effect explanations are offered when dealing with processes and procedures, frequently those involving technical instrumentation that affords human intervention possibilities. This, of course,

4 Paul Humphreys, *Emergence*, op. cit, 50.
5 Ibid, 50.

can happen at any level, microscopic or macroscopic, when dealing with substances or single atoms.

The discourse of cause and effect and the discourse of emergence operate alongside each other; both can be uttered in the same breath, as it were, but they must not be confused with each other. The latter concerns how something can be explained in the fundamental terms of science; the former how something works or can be transformed in terms of practical engineering procedures. Thus when we speak of determinism, of something determining something else, we must be very careful to specify whether this is meant in the emergent or causal sense. The logic of the two senses of this word are very different and only confusion and needless argumentation arises from overlooking that difference. Hence, there are numerous papers and books for and against downward causation.

The example of the emergence of water out of its constituent elements is, of course, a very simple case. It does not explain everything about emergence. Focussing on it too intently will make one overlook fundamental distinctions between types of emergence that arise in more complex situations, such as that between synchronic and diachronic emergence which Humphreys points to:

> Diachronic emergence primarily, but not exclusively, involves the emergence of novel phenomena from preceding phenomena as a result of a temporally extended process; synchronic emergence involves the simultaneous coexistence of novel "higher level" objects or properties with objects or properties existing at some "lower level", where the "levels" talk is best interpreted metaphorically. I note that it is possible to have both diachronic emergence and synchronic emergence occurring in a single process.[6]

The final point is particularly relevant to the emergence of mind where there is not only the synchronic emergence of conscious mental acts out of the physical basis of brain, body and environment to be considered, but also how the two forms of diachronic emergence, the ontogenetic and the phylogenetic, explain the emergence of mind in the first place. As we intimated in previous chapters, how these three forms of emergence relate to each other is the crucial issue in the science and philosophy of mind. The key to this issue is to show how the two forms of diachronic emergence, the ontogenetic and

6 Ibid, 43.

phylogenetic, account for the synchronic emergence of mental phenomena from their material basis in the physical world.

Diachronic emergence brings in the temporal aspect which is relevant to all developmental and evolutionary processes. It even applies to such "historical" questions as to how water first arose on earth. It is, of course, of particular relevance to the question of the origination of organisms or how life emerged and evolved, which we considered in Chapter 1. In this case we know much more about the synchronic emergence of life than the diachronic. We know a great deal about how a living cell functions on all levels of organization from the atomic, as in molecular genetics, to the biological, as in the study of forms and formations of whole organisms and how they grow and develop, such as D'Arcy Thompson so beautifully described and illustrated.[7] But we know very little about how life emerged in the first place.

When we speak of life emerging from the interaction and activation of complex chemical molecules we mean this in the synchronic sense of emergence. What it involves is explaining how a living cell or biologically more developed organism maintains itself so as to remain alive by continually processing complex chemicals, such as the DNA and RNA molecules to produce proteins. In other words, the cell is a little chemical plant, in something like the industrial meaning of that word as well as the biological. But the living cell and its chemical plant, and the molecular processes going on within it, are not two separate and distinct things, they are in the emergent sense the very same thing. The living cell and its molecular chemistry are indissolubly linked together. This is so provided the emergent relation holds and the science which postulates and is based on such a relation, namely our genetic and biological science, remains true. But ultimately that, too, is an empirical matter, though one of very great and basic generality, it could conceivably be falsified at least in theory.

If emergence holds, then by the same logic the molecular processes and the organism as a living entity are not identical. Life or the science which studies it, biology, cannot be reduced to chemistry. This is demonstrated by the fact that there is no way of deducing the phenotype from the genotype; knowing all the genes of an organism does not tell us how it will develop for there is much more to that than the genes themselves, there is their expression which is affected by all kinds of epigenetic factors, some of which are environmentally determined and can be pure matters of chance.[8] It is even possible for two genetically identical organisms to develop quite different phenotypes, as Cohen

7 D.W. Thompson, *On Growth and Form* (Cambridge: Cambridge University Press, 1942).
8 See Siddhartha Mukherjee, *The* Gene, op. cit, 107.

and Stewart point out.[9] In this case, it is not only that the whole is more than the sum of its parts in a static sense, but that the whole, the organism, acts continually in producing the parts, its own constituents that it needs to keep itself functioning, that is, to keep itself in being and remain alive, as Spinoza had determined. It is in this sense that biology is teleological, it works with an end in view, as Kant had maintained, and as Maturana and Varela do so now in terms of their notion of autopoiesis. But its atomic complement, chemistry, is not teleological.

It is somewhat analogous with the emergence of mind, except that this is a vastly more complex matter than the emergence of life. As we have previously stated, it is a matter of emergence raised to a still higher power in the power-law account of emergence from E^4 to E^5. What emergence raised to the fifth power, what, following Hamlet, we have called the quintessence of dust, amounts to is our main burden of explanation. But since the logic of emergence applies to it as well as to all the earlier examples discussed, we can draw some analogies from these to illuminate the much more obscure notion of the emergence of mind.

As we saw previously, the three types of emergence are bound up with one another and closely interrelated. Together they constitute the scientific explanation of the mind. At present they are under the aegis of different sciences. Thus synchronic emergence comes under neurology, psychology and, as some maintain, also artificial intelligence. Evolutionary diachronic emergence or phylogenesis belongs to palaeoanthropology, ethology and, as some maintain, also evolutionary psychology. Developmental diachronic emergence or ontogenesis is the province of embryology, early development psychology, linguistics, educational theory and, as some maintain, also modular psychology. Due to disciplinary boundaries and the division of academic and scientific labour, exponents of all these disciplines do not cooperate with each other as much as they need to do. This is also because they do not realize how much they depend on each other. The emergence of mind is the one vast topic with many scientific aspects to it, not a series of separate sciences each pursuing its own problems and paradigms.

The emergence of mind can only be scientifically explained through the integration of the three basic modes of emergence: the synchronic and the two forms of diachronic, phylogenetic and ontogenetic. Each provides vital clues and models for the pursuit of further investigations in the others, as the findings in the study in one kind of emergence clarify the issues in the other kinds. Hence, some of the best research has come from people who are competent in

9 Jack Cohen and Ian Stewart, *The Collapse of Chaos*, op. cit, 299.

or mindful of a number of different sciences of mind. These are the people whose work we discussed intensively in Chapters 3 and 4 on animal and human minds. There are, of course, many others whom we have omitted, but unfortunately this is in the very nature of an investigation where there is so much going on at once, particularly at present when it has become the leading field of science and, as we shall see, also of philosophy.

Whether the problems we now encounter in scientific research into the three modes of emergence might or might not be overcome in future science, we have no way of knowing. One such problem we raised at the very start of this work concerns diachronic emergence in its phylogenetic form: the anthropological origin of mind, that is, the evolutionary problem of emergence. This is unlikely to be overcome in any foreseeable future since the missing links between humans and our nearest animal relatives, the chimpanzees, have been almost completely obliterated. Nevertheless there are continually new discoveries made which slightly close the gap. Such are the most recent findings in the Rift Valley in Kenya which predate the origins of symbolic art by some hundred thousand years to that previously available; and art is crucial to human mind, as we argued earlier.

Most of what we know about the human brain has been learned through a study of animal brains. The detailed functions of neurons are the same in all animals, hence it can be studied even in the lowest of creatures, like slugs and snails. Even the elementary functions like memory and learning can be studied in their most primitive forms in such simple animals, as neurologists like Eric Kandel have proceeded to do with considerable success. Whether their findings can be upscaled as it were to higher animals remains in doubt. But it is not to be disputed that such work on lower animals and on the emergence of the basic faculties of mind in higher ones is of great relevance for humans as well. After all, our brains and minds have emerged from those of animals in the course of evolution. It is, as previously discussed, a hotly debated topic as to what survives in our brains and minds of their animal origins. Are there little crocodiles and other beasties lurking somewhere in the hidden recesses of our brains? And if so, what effect do they have on our minds? Putting it in proper scientific terms, how do such vestiges of one or two dimensional mind/brain systems function in our three dimensional one? These are still largely unanswered questions. But even to be able to pose them correctly is a step forward.

Finally, we come to the sciences of synchronic emergence of mind: how do the myriads of electrical and chemical activities in the brain simultaneously in parallel generate the experiences we have moment by moment? How is consciousness produced? What is the neural machinery behind a stable Ego? Where and how are our memories stored? What enables us to access them?

These and innumerable other such questions all concern synchronic emergence. Few of them have as yet been answered in any satisfactory way. Nevertheless, as research continues, now perhaps more extensively and with better instruments than ever before, new knowledge is being won. What future science will bring is not possible for us to predict.

The neuroscientists are, of course, the leading researchers in respect of synchronic emergence. They have already accomplished a tremendous amount but vastly more still needs to be done. We are only scratching the surface in understanding the workings of the brain. What has so far been accomplished is that much is known about the details, but as yet very little about the whole. A great deal is known about how single neurons function and how they are linked to each other and how impulses travel along those linkages. But very little is known of how larger patterns emerge in the brain that subtend both our permanent faculties and mental functioning from moment to moment. Network neuroscience is still a fledgling discipline.

Each of the major fields of the emergence of mind still has huge shortcomings and lacuna which might or might not be overcome in the future. But regardless of the final outcome, which is purely a speculative matter, since one cannot predict future science, progress is being made in respect of all three modes of emergence and their respective sciences. Such progress will no doubt continue in the future. Does this mean that we are asymptotically converging nearer and nearer to some final limit of truth? It might mean that or it might mean something quite other, for it is possible that the final picture, the "truth" as it were, might never be arrived at, as Crick seems eventually to have pessimistically concluded in an interview with Horgan:

> Just because the mind stems from deterministic processes, Crick continued, does not mean that scientists will be able to predict all its meanderings, it may be chaotic and thus unpredictable. "There may be other limitations in the brain. Who knows? I don't think you can look too far ahead".[10]

It might also turn out that the final "truth" is a receding vision, that the more problems we solve, the more new problems will be opened up in a never concluding process, one that might only come to an end for the pragmatic reason that it is no longer worthwhile to go on for much the same factors appear over and over again, as in a fractal Mandelbrot set, where as our magnification increases ever newer bewildering details show up, but always close to the same

10 John Horgan, *The End of Science*, op. cit, 182.

form. There might be no end to our quest for an explanation of the emergence of mind, which might continue as long as we consider is worthwhile to pursue our researches.

However, regardless of whether there is or is not any final "truth", it is, nevertheless, worthwhile at present to pursue this quest for it still does continually yield new knowledge. Most of this knowledge is practically useful and enables us to deal better with the pathologies of brain and mind and in other psychiatric respects. And as for philosophical speculation, which we must continue to engage in even though it has no immediate practical use, such theorising satisfies our need for understanding and can have profound cultural consequences.

How profound the cultural consequences of a scientific study of the mind can be we can grasp if we step back to the time when the play *Hamlet* was written, just prior to the scientific study of the body and brain inaugurated by William Harvey, Descartes and many others of the early medical physiologists. We need only compare our understanding of mind, brain and body with that current in Hamlet's time to see how far we have travelled in our quest for knowledge and what science has taught us. It would be foolish to deny the term "progress" to this difference. Nevertheless, as we saw, Hamlet was not completely wrong and still has something to teach us.

2 On Ghosts and Minds

At the start of the play *Hamlet* the ghost of Hamlet the king appears. First it shows itself to Barnardo, Marcellus and Francisco, officers of the watch; then to them and Horatio, who at first is highly sceptical, being a Stoic by conviction; and finally to all of them and Hamlet as well, who greets it rapturously, becoming almost hysterical. When Horatio dares to question him, he utters his put-down rebuff: "There are more things in heaven and earth, dear Horatio, than are dreamt of in your philosophy." If Horatio's philosophy is a naturalistic Stoicism, then what Hamlet is saying is that there are things beyond naturalism, namely, supernatural things like ghosts. What if Hamlet was right, as, indeed, most people of his time and most probably even now believe? What if ghosts were real and could not be discounted as mere superstitious beliefs or as mass delusions, as most philosophers, scientists, intellectuals and religious sceptics were wont to do when they ruled out the reality of ghosts, poltergeists, and other such supernatural occurrences.

If ghosts were real, then this would have profound consequences for our philosophy of mind. If ghostly apparitions began to wander the earth and haunt the living, like the ghost of Hamlet the king, then we would have to dispense with any thought of a scientific account of the emergence of mind. Any philosophy of mind based on emergence would have to be abandoned as well and instead a philosophy approximating to Gilbert Ryle's travesty of Descartes' Dualism, that of a ghost in the machine, adopted. This would render mind scientifically inexplicable, and doom key aspects of the whole scientific enterprise regarding human beings. Animals could, perhaps, be considered as mere bodies without any mind since they have no ghosts. Ultimately, it is a mere matter of fact that we inhabit our world and not the world as depicted at the start of the play *Hamlet*. But if that obtained, then it would amount to a kind of falsification of our whole approach to mind, both scientific and philosophical.

We say "a kind of falsification" because philosophical theories cannot be falsified or verified in any straight-forward way, unlike most scientific theories. Nevertheless, there is a sense in which the term "falsification" does apply; for if empirical phenomena were utterly different, if ghosts really walked abroad, then the whole emergentist approach to the mind-body relation would become inapplicable, it would simply cease to be relevant. In the face of the supernatural, so, too, would any rational naturalistic approach to the mind as well. A scientific explanation of mind would become impossible, and the origin and nature of mind would be the mystery that some philosophers, such as Colin McGinn, even now claim it to be. Thus emergentism and science cohere together, if the one falls so, too, does the other. This large claim requires further and more closely argued justification which we must now proceed to provide.

In an era of emergent science we are justified in offering an emergent account of the mind-body relationship as part of a more general theory of the emergence of things. If this holds, then the logic of emergence will apply to this relation and it can be characterized as an indissoluble non-identity. The reason why we are bound to offer such an account is because the failure to do so or the breakdown of emergentism would be both philosophically and scientifically unacceptable. We can, however, hypothetically imagine all kinds of possible situations in which the logic of emergence would not hold and emergence could not be applied to the mind-body relationship.

The apparition of Hamlet's ghost is but one very crude instance of the kind of empirical possibility, which, if true, would completely destroy our naturalistic vision of reality and make any kind of rational science impossible. However, there are all kinds of other imaginable empirical contingencies, which would not have such drastic consequences for science in general, but would, nevertheless, make an emergentist approach to mind untenable, and in that sense falsify it. We can imaginatively explore some of these possibilities.

One kind of imaginary situation would arise if beyond a certain point no further detailed correlations could be discovered between what occurs in the mind and what can be found in the brain and body. Numerous variants of this are possible. There could be very gross inconsistencies such that nothing specific can be detected even with the most refined instruments when thoughts are going through the mind; or, in a more refined way, as Crick suggested, it might be the case that only very chaotic activity occurs with no way of differentiating between one electrical storm and another. In other words, what this would amount to would be a complete dissociation beyond a certain point between thinking and bodily processes, in particular, neural activity.

Such imaginary findings would invalidate the first principle of the logic of emergence, namely, the inseparability of emergent and emerger. Where there is such an indissolubility then this cannot happen, there must always be overall correlations, the degree of specificity only governed by the resolution capacities of our instruments, such that technological progress in instrumentation always ensures that new complexes on ever larger scales will always be discovered. Looking into the brain and body is like looking into the cosmos as a whole, as our telescopes improve so ever larger structures will always be revealed. So, too, as our instrumental means of detection and recording improve so, too, will the extent of our capacity to take in what is going on in the brain on the largest scale. If, *per impossible*, this would cease to happen and the brain became an opaque black box, then emergence would not hold and the whole attempt at an emergentist explanation would break down. The logic of emergence would also no longer apply to the mind-body relation.

But not only would emergence break down, so, too, would science as a whole in this field. It would not be possible to carry out any further meaningful scientific research with any explanatory potential. The neurosciences would be over and so, too, would the various types of psychologies linked to neurological data. It would, for example, become impossible to account for all kinds of pathologies in behaviour, in language production and understanding or perceptual illusions which are neurologically based. In short, it would amount to a catastrophe for science which would fundamentally alter our conception of human nature and reality as a whole.

Scientists and philosophers would no doubt respond in various contrary ways in the face of such a catastrophic eventuality. Some would simply refuse to accept that there was an observational barrier or limit to any further investigation as to what goes on in the brain. They would live in hope that further improvements in instrumentation would at some point reveal further linkages. Others, those who accepted that there is no more to be seen, would resort to versions of Dualism and could argue for a causal relation between isolable and inherently independent substances of matter and mind. They would then be faced with the old Cartesian dilemma that the causal mechanism of this relation would be inexplicable and inherently mysterious. There are, indeed, mysterians at present who argue that there are limits to our capacity to understand the mind-body relation.

What our imaginary examples reveal is that the failure of emergence entails the failure of science. To most of us it is little consolation that this would signify the success of Dualism, though there are some scientists, like the neurologist John Eccles, and some philosophers, such as his co-author Karl Popper, who would no doubt be glad at this outcome, without fully realizing what it would mean for science.[11] If one believes in the continuity of ever further scientific research based on ever newer and better data then one must not fall for the dead-end of Dualism. There is no mystery of mind but a continual unfolding in ever greater scope and specificity of the relation between emerger and emergent in this field. And as long as science lasts, this research will continue with no foreseeable end in sight.

However, it is possible to imagine another outcome where this research would come to an end; another counterfactual hypothesis where emergence would break down. In such a *per impossibile* thesis a completely reductive account would make emergence unnecessary for it would overrule the second premise of the logic of emergence, that of the non-identity of the emerger and emergent. If reductivism were to obtain, then the two sides would be identified and it could be maintained that mind is nothing but brain and body. Materialism would be the great beneficiary of such an unlikely development, though a later somewhat different variant called Functionalism that derives from the computer model might contest that honour. We shall discuss it further in Chapter 7.

The success of reductivism might be imagined in numerous variants. It is conceivable that a simple underlying formula governs all the operations of the brain. This would be like the infinite variety and complexity of Mandelbrot's set generated by the endless recursive iteration of the one simple equation; or like the cellular automaton game of life where a few simple rules produce an unceasing play of patterns; or like deterministic chaos where the one equation recursively repeated produces an ever changing outcome. In fact, the generalists of emergence, such as John Holland, Christopher Langton and their colleagues at the Santa Fe Institute, hope to be able to capture the essence of the mind in some such computer generated technique of juggling with codes in their programs.

11 J.C. Eccles, and K.R. Popper, *The Self and its Brain* (Berlin: Springer, 1977).

Should any such scheme succeed, it would also put paid to emergence and bring the sciences of mind to an end. For there would be no longer any need to carry out any further experimentation or observation *in vivo*, as it were, it could all be done *in vitro* on computers. If one had the right mathematical settings and could devise an effective program, then a computer powerful enough could reproduce any mental phenomenon in computer code, provided it was given long enough time and a large enough program to do so. This would be Artificial Intelligence with a vengeance for it would amount to artificial mind or "virtual" mind.

But, in fact, no such neural code has been found and it is most unlikely that it ever will be found. Even such a hopeful researcher as Christof Koch, Crick's former collaborator, now concedes that "it is most unlikely that the neural code will be anything as simple and as universal as the genetic code."[12] Horgan goes on to quote Koch as warning "that the neural code may never be totally deciphered", and the reason he gives is like that given by Edelman, the enormous variability of different brains and the "mutability (or 'plasticity' to use neuroscientists' preferred term) [which] immensely complicates the search for a unified theory of the brain and the mind."[13] However, it is also likely that the very idea of searching for such a unified theory of brain and mind, that is, a neural code or a program for translating electrical pulses and currents emitted by brain cells into mental entities, is a mistaken endeavour promoted by the computer metaphor.

As a matter of fact, it so happens that neither neurology nor psychology nor any other science of the mind can be conducted in this manner. The reason for this is that no such set of rules of coding will ever be found since the operations of the brain and their emergent correlates in the mind cannot be formalized. Whatever is discovered is always novel, surprising and unpredictable, as is the case in emergence. And such discoveries can only be made by detailed empirical investigations. This is how all we know about the relation between the emerger and the emergent was discovered in the past and this is how so much more will be found in the future. There can be no short-cuts to the scientific work in the laboratory that must be carried out, most of it lengthy and incrementally productive, with the occasional serendipitous finds, such as Rizzolatti's mirror neurons, which came as a complete surprise. There are bound to be more such in the future.

12 Quoted by John Horgan, "The Final Frontier? Are we Reaching the Limits to Science?" op. cit.

13 Ibid.

When the current phase of this scientific project began, largely under the auspices of Crick, it was naively assumed that consciousness was to be associated with some relatively small number of cells in the brain with set characteristics of behaviour, such as neurons firing in synchrony at 40 hertz, or that it was associated with pyramidal cells in the neocortex, or on-off switches in the claustrum. These ideas have been since abandoned and it is now assumed that consciousness relates to large neural networks. How large and where they are located is a matter of intense dispute among neuroscientists.

At present the two dominant contending theories are the Global Workspace Theory advocated by Bernard Baars and Stanislas Dehaene and the Integrated Information Theory advocated by Giulio Tononi and Christof Koch. The former theory locates the correlates of consciousness largely in the cortex at the front of the brain where a neural competition among incoming information takes place with the winner gaining the spotlight of selective attention in the theatre of consciousness. According to the latter theory the correlates of consciousness are located in the cortex at the back of the brain where there is a very large degree of interconnectivity between groups of neurons or modules. It is not possible here to go any further into the differences between these two theories, nor can a layman be presumptuous enough to try to adjudicate between them. There are now numerous books on the subject by the primary exponents of these theories.

It is impossible to predict in advance which present attributions of emergence will hold in the future and which will be falsified, since we cannot, in principle, predict the discoveries of future science. Nor is it possible to divine what new situations of emergence will be discovered in the future. However, it is a safe bet that some cases of emergence are bound to remain for any foreseeable future, and that the mind-brain relation is bound to be one of these since the brain-body-environment interaction is the most complex single entity that we encounter in the universe, short of the universe itself. And that entity is us as conscious beings or minds. And since it is almost inconceivable that we will ever be able to reductively do away with ourselves, it follows that we are bound to see ourselves in emergentist terms.

The scientist who comes closest to affirming the logic of emergence in respect to the mind-body relation is Gerald Edelman, though he is not quite consistent about it, as we shall see. Edelman does not avail himself of the concept of emergence. Nevertheless, he almost captures the logic of indissoluble non-identity that governs emergence. He postulates a non-causal relation between mind and body in that he grants that mind cannot have a causal effect on the brain. He takes the opposite view to that of Roger Sperry who espouses downward causation: "it is noteworthy that, more recently, another gifted scientist, Roger Sperry, has taken the position that consciousness can actually affect a neuronal firing. Obviously, I have taken a contrary position…"[14] Unfortunately, Edelman is not careful enough to rule out the inverse position, that of upward causation, namely, that neural firing has a causal effect upon consciousness. This leaves him open to charges of epiphenomenalism, which he does his best to rebut, but not in a convincing way. We shall explore each of these points in turn.

Edelman asserts the non-identity thesis in a direct and unequivocal manner:

> I have attempted to show that consciousness has arisen, at least in this little speck of the cosmos, at a particular historical time. That it emerges from definite material arrangements in the brain does not mean that it is identical to them, for as we have seen, in its higher order, consciousness depends on relations with the environment and, in its highest order, on symbols and language in society.[15]

This is reaffirmed in later works, as in his joint book with Giulio Tononi:

> …consciousness arises within the material order of certain organisms. However, we emphatically do not identify consciousness in its full range as arising solely in the brain, since we believe that higher brain functions require interactions both with the world and with other persons.[16]

This is restated in his last work:

> All these factors account for the "irreducibility" of consciousness and the subjective state. While some feel it necessary to "reduce" conscious experience by identifying it with neural action, this reduction leads to a category error. The origin of qualia as properties of neural processes having high-order discriminatory power, does not eliminate the subjective experience they represent.[17]

The "some [who] feel it necessary to reduce conscious experience by identifying it with neural action" obviously refers to Edelman's main rivals in

---

14 Gerald M. Edelman, *Wider than the Sky,* op. cit, 84.
15 Gerald Edelman, *Bright Air, Brilliant Fire,* op. cit, 198.
16 Gerald Edelman and Giulio Tononi, *A Universe of Consciousness,* op. cit, Preface.
17 Gerald Edelman, *Wider than the Sky,* op. cit, 124.

neurology, such as Francis Crick and Jean-Pierre Changeux. They are evidently identitarian materialists whereas Edelman refers to his views as "qualified realism, sophisticated materialism, selectionism and Darwinism."[18] By contrast, Changeux, who is also a selectionist as far as brain physiology is concerned, expressed an opposed identitarian materialist point of view on mind:

> The combination of possibilities provided by the number and diversity of connections in the human brain seems quite sufficient to account for human capacities. There is no justification for a split between mental and neuronal activity. What is the point of speaking of "mind" or "spirit" … It seems quite legitimate to consider that all mental states and physiological and physico-chemical states of the brain are identical.[19]

It is, however, ironic that Edelman never mentions Spinoza, whereas Changeux does refer himself to Spinoza, and prefaces the chapter from which this very quotation is drawn with two passages from the *Ethics*. He also puts forward in the very same paragraph as the quotation an apparently Spinozist point of view: "it is only that there are two 'aspects' of a single event, which one may describe in terms taken from the interpretative language of the psychologists or from the language of the neurologist."[20] But this "double aspect" view is quite different from the view expressed in the same place that the "language of the neurologist" referring to neurons is all there is in reality. Clearly, Changeux is somewhat confused about Spinoza and his own position in relation to him.

Edelman is not confused, he sees clearly that mental processes and neural events are not the same, that the former cannot be reduced to the latter; but that at the same time they are not dissoluble, that the former cannot be separated from the latter. In that respect he is neither a Materialist nor a Dualist. Yet he fails to grasp the full force of the logic of emergence that obtains in this connection because he does not see the relation as the two sides of emergence, the emerger and the emergent, even though he often invokes the word in a colloquial way; unfortunately he does not subject his use of the term "emergence" to any theoretical definition.

This lack of any explicit theoretical awareness of what is involved in emergence leads Edelman into making statements that lend themselves to the charge of epiphenomenalism. He espouses a causal relation between the

18 Ibid, 164.

19 Jean-Pierre Changeux, *Neuronal Man: A Biology of Mind*, trans. Lawrence Garey (New York: Pantheon Books, 1985), 275.

20 Ibid, 275.

physical and the mental, but not between the mental and the physical. Such a one-way causal nexus is characteristic of epiphenomenalism, the view that the mind is like the smoke from the locomotive, a product of its combustion that in no way affects its motion. Edelman does his best to fend off such charges but not very successfully since he does affirm such a one-way only causality and signally fails to see that in essence no causal relation is involved either way at all.

Over and over again Edelman insists that the neural events in the brain cause mental processes, but that these, in turn, cannot possibly have any effect on the brain or body or physical world in general:

> How can the firing of neurons, however complex, give rise to feelings, qualities, thoughts and emotions?… The key task of a scientific description of consciousness is to give a *causal account* of the relationship between these domains so that properties of the one domain may be understood in terms of events in the other…[21]

But only the physical domain is causally efficacious, not the mental. As he puts it: "the extended theory claims that underlying any conscious state C is a set of neural processes C^1. Given the causally closed nature of the world, it is C^1 [the neural processes] that is causal, and not C [the conscious state]."[22] As for the "causally closed nature of the world", he affirms it as a fundamental assumption that the "principle of physics must be strictly obeyed and that the world defined by physics is causally closed".[23] Namely, the mental cannot have any effect on it.

However, Edelman eschews the charge of epiphenomenalism:

> I must answer the accusation that I have submitted to the paradox of epiphenomenalism … I believe that the difficulties with this notion have arisen because of the failure to attend to the neural correlates of conscious properties. Inasmuch as the neural process C^1 that entails consciousness C is *causal and reliable*, we do not find ourselves faced with a paradox.[24]

---

21 Edelman, *Wider than the Sky*, op. cit, 11.
22 Ibid, 115.
23 Ibid, 114.
24 Ibid, 145.

The terms Edelman uses to explain the relation between C and C^1 are oddly dissonant and even contradictory: "correlates" is different from "entailment" and both of these are opposed to "causal". A few sentences later Edelman speaks of "entailment" as if he meant this in its proper logical sense as expressive of a necessary relationship: "the relationship of entailment between C^1 and C clarifies the issue and helps define qualia as higher-order discriminations with specific neural bases. A consciousness-free zombie, on these grounds, is logically impossible – if it had C^1 processes they necessarily entail C."[25] This is, indeed, correct but, unfortunately, Edelman does not grasp what entailment in the logical sense means: if "C^1 processes necessarily entail C" then the relation between them cannot be causal, for causality is always a contingent empirical matter not a necessary logical one. If C^1 causes C then it is always logically possible that C^1 occurs without C, the cause without its effect, and this is what it means to be faced with the possibility of zombies, purely physical beings with no consciousness, that Edelman quite correctly denies are logically possible.

If only Edelman had stuck to the notion of entailment in this context and dispensed with that of cause, he would have arrived at something approaching the logic of emergence. And at one point he almost does this, as when he states that "it is essential to recognize that, *in a strict sense*, C^1 does not cause C – there is no temporal lag in the expression of C upon the occurrence of C^1."[26] In other words, if the effect must follow and be consequent upon the cause, then C cannot be caused by C^1 since they are both synchronous, hence *in a strict sense*, which is the logically proper sense, "C^1 does not cause C".

However, we must grant Edelman and all others who speak of a causal relation between mind and brain that there is a loose sense of "cause" in which this is true. But it is not the kind of causation that they assume it to be. Rather, this is "cause" used in a purely pragmatic sense, almost as a kind of *façon de parler* which we find useful to employ in scientific as well as every-day discourse. In this special sense, we can refer to a kind of "downward causation", but it is a conceptual construct based on the logic of emergence. Thus it is both scientifically unavoidable and necessary in neurology to speak of long-lasting mental practices and endeavours, such as solving mathematical problems or learning music or another language, as exercising a causal role on the brain in altering its structure. This, of course, does not imply that the mind is a causal force that can affect the firing of neurons. But from a long-term perspective, there are good grounds for asserting that the mind exercises a causal effect on the brain.

25 Ibid, 145.
26 Ibid, 116.

Here the concept "cause and effect" is invoked in a context of emergence that does not contradict it. It is a special application of the term that means something quite other from what it means when we speak of billiard balls hitting each other or ignition bringing about combustion, or hosts of other such cause-effect relations.

An elucidation of causation in respect of the mind-body relation along the lines we shall develop further is provided by Evan Thompson in his article "Contemplative Neuroscience as an Approach to Volitional Consciousness".[27] He puts it quite explicitly that "the term 'downward causation' is a misnomer. Complex-system causality is not a matter of a higher level acting downward on a lower level."[28] Instead, he adopts a pragmatic or working definition of downward causation based on the idea of access or intervention, which he derives from Michel Bitbol whom he quotes as follows:

> Making sense of upward and downward causation does not require a metaphysical distinction between higher and basic levels of organization. Neither a substantial distinction, as in genuine dualism, nor a distinction between properties or structures, as in the current popular picture. It is enough to assume a duality of modes of access, or modes of intervention. If one intervenes at a higher level of organization, some effects of this action can then be detected by a mode of access specifically aimed at the lower level. This is downward causation. Conversely, if one intervenes at a microscopic level, some effects of this action can then be detected by a mode of access specifically aimed at a higher level of organization. This is upward causation.[29]

Thompson goes on to conclude from this definition of downward causation that "this formulation allows us to say in a perfectly coherent way that contemplative experience acts downwardly on the brain by providing a distinct way of psychologically intervening in neurobiological processes."[30] It thus makes sense to say that intentions and acts of will exercise a downward causation on the brain if one utilizes them to intervene in the brain in order to generate certain specific neurological processes, as one might in an experimental set-up in a psychological laboratory where recordings are being made of what goes on

27 Evan Thompson, "Contemplative Neuroscience as an Approach to Volitional Consciousness", in Nancey Murphy, George F.R. Ellis and Timothy O'Connor, eds. *Downward Causation and the Neurology of Free Will,* (New York: Springer, 2009).

28 Ibid, 193.

29 Quoted in ibid, 195.

30 Ibid, 195.

in the brain while the subject is making up his mind to act, as in the demonstrations carried out by Benjamin Libet. Libet, unfortunately, draws from this the one sided conclusion that it is always the brain that causes acts of will to occur, and he does not allow for a two way causation. We shall discuss this in greater detail in the next chapter.

There is one location in his numerous texts where Edelman does in fact adopt this pragmatic locution of cause in reference to the mind. He calls it a case of "as if causation" in speaking of C (consciousness) as a cause of C^1 (its physical correlates):

> This relationship [between C and C^1] allows us to talk of C *as if it is causal*. For most situations this is not dangerous, given the reliability of the relationship. Only when we are tempted to abrogate physics or give to C mystical power is this procedure hazardous.[31]

But exactly the same can be said of C^1 as well, for it, too, can be spoken of as an *as if cause* of C in scientific contexts. Yet it can also be spoken of as an effect as well, as when we say that constant reading of music can cause the brain to acquire certain neuronal connections, whereas it would not make much sense to speak of these neuronal connections as causing one to read music, only as enabling one to do so. As we shall see in the next chapter, cause and effect relations between mind and brain are sometimes reversible and sometimes not, as in this case. It all depends on the context in which they are invoked, which is a pragmatic matter, not an ontological one.

No such inversion of cause and effect or talk of causation as such is in itself dangerous or confusing provided one realizes that it is built upon a relation of emergence which in a strict sense is not causal. Causality is conceptually constructed on emergentist foundations according to criteria which vary depending on the context in which they are invoked. Invoking causal terms for the mind in an everyday practical context is very different from doing so in a scientific context. Thus, for example, in an everyday sense it is the emotion, e.g. the fear, that causes physiological and behavioural reactions; but in a scientific context, it is vice-versa, the physiological and behavioural responses cause the feeling of fear, as Damasio explains and as we will discuss further in Chapter 6.

It is, of course, to be understood that all the previous criticisms of Edelman refer solely to his philosophy of mind and have little if any relevance to his science of mind, his neurological theory of group selection. His theory of how the brain functions has come under heavy criticism from other neurologists and

31 Ibid, 195.

psychologists such as Crick and H.B. Barlow. Boden quotes Barlow's view that Edelman's work is "a hopeless muddle" and is herself scathing of it.[32] Crick mocks Edelman's term "neural Darwinism", which he maintains should instead be called "neural Edelmanism". Alas poor Gerald! Nevertheless, there is a point to Edelman's invocation of Darwin, since he has developed his theory of the brain on the model of his highly successful theory of immunology which is vaguely reminiscent of Darwinian selection of the fittest. Whether an analogous selectionist approach will also work for the brain is a scientific matter on which we cannot presume to comment.

Edelman has also attracted a host of critics from the Materialist camp, especially from the "mind as machine" exponents, such as Boden. In turn, he has been highly critical of their advocacy of Information Theory, Artificial Intelligence and Cognitivist science, and the philosophy of mind based on such theories, as we shall see in Chapter 7. However, before we can turn to his critiques of current forms of Materialism we must first review historically how they were arrived at. In the next chapter we will try to show that, paradoxical as this sounds, they are derivatives of Cartesianism. Indeed, we will demonstrate on historical grounds that there is such a thing as a Materialist Cartesianism and that this rose to prominence in the period following the Second World War. The main reason for this was the huge technological development that took place as a result of that war and the research carried out in the context of the Cold War. Pre-eminently, this led to the invention of the electronic computer and the utilization of this instrument as the model of an electronic "brain", which in turn had drastic consequences on the science and philosophy of mind.

Edelman is one of the main exponents at the end of the twentieth century of a reaction against this whole centuries long scientific and philosophic trend. This has made him enemies among those who are still ardent adherents and advocates of such views. Even though he is aware that these positions ultimately derive from Descartes, he is, nevertheless, highly appreciative of Descartes' own achievement, he writes of it as follows:

> If Galileo is the founder of modern science, Descartes is the founder of modern philosophy. His thoughts are proof that genius, even genius leading to wrong conclusions, can be of continuing major significance. We still wrestle with the questions that Descartes posed.[33]

32 Margaret Boden, *Mind as Machine,* vol II, op. cit, 1204.
33 Gerald Edelman, *Bright Air,* op. cit, 255.

In the next chapter we shall seek to establish what these "wrong conclusions," to which Cartesianism has led us, are and how and why they have ineluctably arisen in the course of the historical development of philosophy and science. After we have dealt with Descartes and the latter-day Cartesians, we will turn to Spinoza in search for an alternative view on the problems that Descartes has endowed us with.

CHAPTER 6

Historical–Philosophical Interlude

1 From Descartes to the Present

In the history of Western philosophy, three approaches to the mind stand out as preeminent, those of Aristotle, Descartes and Spinoza. They were paradigmatic in that they featured in all kinds of variants over long periods. Aristotle's philosophy of mind was current in the ancient world and was dominant throughout the medieval period well into Modernity. The Cartesian philosophy of mind rules over the whole Modern period and in various forms is still present, as we shall see. Spinoza's philosophy of mind has been the least influential, but it has had its supporters at all times, especially in Germany among its literati, philosophers and scientists, including Lessing, Goethe, Herder, Fechner, Schopenhauer, and Einstein. At present it is coming into its own because it is compatible with an Emergentist approach to mind. As we shall see, it has been explicitly taken up by many contemporary scientists, including the biochemist and medical researcher Henri Atlan and the neurologist Antonio Damasio; both have devoted books to Spinoza.

Spinoza's philosophy of mind is the closest among the Classical philosophers to that which we have outlined in the previous chapter. It, too, advocates an indissoluble non-identity relation between mind and body, and explicitly eschews a Cartesian causal relation between them. Spinoza has been interpreted in numerous ways, but the latest scholarship favours this kind of reading, as we shall see. This underscores the anomalous role that Spinoza has played in the history of philosophy in that he is a Janus figure who looks back to the past as well as forward to the future.

Atlan's work brings out well the role of Spinoza as an intermediary between the most ancient and most modern.[1] He draws support for this view from the extensive recent French literature on Spinoza, with books by Martial Guéroult, Robert Misrahi, Giles Deleuze, Alexandre Matheson, J.F. Moreau and Pierre Macherey. Macherey's statements are particularly telling: Spinoza is a "combination of archaism and the avant garde"; "[he] skipped over or rather under his time"; "[he] has never stopped being, always in the present, an object of

1 Henri Atlan, *The Sparks of Randomness,* vol. I: *Spermatic Knowledge,* trans. Lenn J. Schramm (Stanford: Stanford University Press, 2011).

DOI:10.1163/9789004426863_008

fascination and reflection…".[2] Atlan makes the same point when he states "that Spinoza occupies a special place in the history of philosophy has been sufficiently underlined both by the modern philosophers who reject him and those who, on the contrary, attempt in diverse ways to rework him in modern terms."[3] Among those who rejected Spinoza are such philosophical luminaries as Kant, Hegel and Bergson, whose objections Atlan discusses. Atlan himself, obviously belongs among the latter kind who seek to incorporate Spinoza into their own thinking and try with his help to straddle the great divide separating ancient from contemporary thought. But perhaps he takes this attempt too far as when he states that "in Spinoza, as well as in some kabbalists, Stoics, and Neo-Platonists we find, more than anywhere else, the elements of an epistemology that is associated with an original monistic ontology, neither idealist nor materialist, and that is particularly suited to a natural philosophy that is aware of what contemporary biology, as the physics and chemistry of organized beings, teaches us."[4] Building bridges between diverse modes of thought is a worthy undertaking, but linking physics to kabbala is surely a bridge too far.

It is not in this spirit of all-inclusiveness that we wish to introduce Aristotle into our discussion, and, in any case, he will only play a minor role compared to the modern thinkers Descartes and Spinoza. Nevertheless, there is more than just a passing resemblance between Aristotle's account of the mind and the one we previously outlined. Aristotle's three psyches or souls, the vegetative, sensitive and intellective, echoes in a distant way our previous idea of one, two and three dimensional mind/brain complexes; and in both cases a relation of encapsulation or nesting obtains; for just as in Aristotle the lower psyches act as the matter for the form of the higher psyches, so, too, the three dimensional mind/brain system encapsulates those lower in the hierarchy as evolutionary vestiges.[5] It would be anachronistic to drive these parallels too far, nevertheless, there is some structural analogy in the architectonics of mind between the very ancient and contemporary accounts, and this we might explore further.

According to Aristotle, *nous* or intellect is the third and highest of his souls, that which in later philosophers came to be called mind. *Nous* is translated into Latin as *mens* from which the English term "mind" derives. But *nous* is only one of the constituents of what we now mean by mind, thanks largely to the

2 Ibid, 215.

3 Ibid, 215.

4 Ibid, 15.

5 Aristotle, *De Anima*, Book II, Chap. 2 and Book III, Chap. 2, *Introduction to Aristotle*, ed. Richard McKeon, trans. J.A. Smith (New York: The Modern Library, 1947).

philosophies of Descartes and Locke. Aristotle regarded the three souls as a unity that together defined human being as a whole, both body and mind, as we would now term it. He explained the relation between the souls in terms of his metaphysics of *hyle* (matter) and *morphos* (form). The relation between the souls was hierarchical with each lower soul constituting the matter of which the higher soul was the form. Whether the highest soul, the *nous*, was pure form that could exist without matter, and was, therefore, immortal, was a controversial point among Aristotelians throughout the ages. Aristotle's writings could be interpreted in both ways, both for and against this contention.

The resemblance between Aristotle's three souls and the three types of brain systems can be made even more explicit. The vegetative soul corresponds to the autonomic bodily homeostasis and to purely reflex behaviour which is already to be found in the lower animals. The sensitive or animal soul corresponds to what we have called animal mind, or in human beings to the purely automatic types of perception and action, as well as very basic sensations and emotions. Finally, the intellective soul or *nous* corresponds to the human attributes of mind, such as reason, self-consciousness, Ego and free will. Obviously, these correspondences are only very approximate, but they do constitute a major departure from Cartesian Dualism.

It might be put metaphorically that we have gone back behind Descartes' back to Aristotle, and we have done so via Spinoza. But it is not only Descartes himself who is being by-passed, but also the whole Cartesian era of philosophy which reaches well into our own time. As we shall show in the next chapter, the Cartesian influence is paramount even now if we consider its latest mechanistic variants. But to demonstrate this we first need to outline the history of Cartesianism in all its ramifications from Descartes onwards.

It has long been maintained that Descartes was the founder of modern philosophy, but the truth behind this truism has not always been fully appreciated. This is especially so in the philosophy of mind. Prior to Descartes the philosophy of mind during the Renaissance was extremely heterogeneous as we can readily see by reference to *Hamlet*. A great many variants were simultaneously in circulation, varying from primitive spiritualism to a highly sophisticated medical materialism. There were numerous versions of ancient philosophies at play: variants of Platonism, Aristotelianism and Neo-Platonism; alchemical versions of various kinds, both hermeticist and Paracelsian; varieties of world-souls and astrological and magnetic influences and sympathies; and so on. In this profusion there was much confusion. In reacting against it all, Descartes brought order and clarity, dispelling hosts of cloudy entities to just the rational opposition of thought and extension; and in that simplifying move lies the greatness of his achievement. Hence, before subjecting the

Cartesian mind to any kind of critique, we must first appreciate its enormous historical importance.

Descartes did away with the various forms of Aristotelian souls as well as the spirits and sympathies of many Renaissance philosophies by postulating a distinction between mind and matter as that of thought and extension. But he did not know quite where to locate the difference between them, where to draw the boundary between mind and body. Was it only thought and free will that belonged to mind? Did sensation, perception and emotion really pertain to the body? After all, Descartes did not deny that animals, too, could sense, perceive and emote. Were these merely a matter of "animal spirits" without mind or was mind involved as well in the case of humans? The old medical doctrine of animal spirits kept confusing Descartes, so it is not clear that he ever really settled these problems to his own satisfaction for he kept changing his mind about them.[6] It was not till Locke and later thinkers that anything like a coherent consensus was reached as to what belongs to mind and what to body. But Descartes had made the crucial start to the definition of mind.

Perhaps even more important than his achievement in regard to mind was Descartes' revolutionary contribution to epistemology: specifically, his theory of perception, where he also differed fundamentally from all previous approaches. Unfortunately, commentators have paid far less attention to this aspect of his thought even though its consequences were more momentous than his metaphysics of mind. Cartesian epistemology was of profound historical significance for it reached beyond philosophy into science and eventually into the whole cultural ambience of the modern epoch. We begin by outlining its difference from the previous dominant theory, again that of Aristotle.

For Aristotle perception, or knowledge which came through the senses, was a direct apprehension of sensory particulars, such as colour, sound and smell; for Descartes it was an indirect communication of external objects by means of representations in the mind.[7] As a consequence, the Cartesian epistemology in general is based on the representative theory of perception. And just as with regard to mind, where we are coming closer to Aristotle at the end of the Cartesian era, so, too, in perception some contemporary approaches, such as Gibson's ecological theory, do away with the representative theory of perception and arrive at a view of direct perception that is vaguely reminiscent of Aristotle.

The battle between Cartesianism and Aristotelianism over perception and mind was part of an epochal cultural transition that began at the start of the

6 See Stephen Gaukroger, *Descartes,* op. cit, 389–90, and 366.

7 Aristotle, *De Anima*, Book II, Chap. 7–12, op. cit, 188–206.

seventeenth century. It was a veritable tectonic-plate shift in Western culture that inaugurated Modernity in all cultural spheres and of which Cartesian philosophy and the new science was a crucial part. At the time John Donne registered it when he wrote that "the new philosophy calls all in doubt…" We have dealt with this issue in another work, where we interpreted the transition in terms of a semiotic shift from Mimeticism to Representationalism.[8]

It is clear that the representative theory of perception is a key component of this new culture of Modernity. But as we also showed, analogous developments towards a Representationalist semiotic paradigm also occurred in politics and art as well as science. The political theories of Hobbes and Locke exemplify a new representative idea of political legitimation; and the institutions of representative democracy in America and France late in the eighteenth century show that eventually this had practical consequences. In the arts there was a simultaneous move to a new naturalistic aesthetic away from Renaissance Classicism and the Baroque to what might be called Representationalist Realism. This can be seen in the rise of the new literary form of the novel, first with Cervantes in Spain then with the English novelists a century later. In painting the new style began with Caravaggio and culminated with Velasquez in Spain and Rembrandt, Vermeer and the other Dutch masters.

Descartes' philosophy was thus not alone in breaking with the previous Western semiotic paradigm of Mimeticism, the principle that semiotic relations were based on similitude or some other type of likeness, such as Plato's idea of participation where particulars participate in the Forms. Descartes's denial of any such mimetic relation between the knower and the known was thus decisive in contributing to the cultural revolution that occurred in Western philosophy and science, even though his own specific theories in both these spheres of knowledge did not prove particularly lasting. The representative theory of perception was his key innovation. He did not arrive at it completely unaided, though he pretended that he did, Kepler and Galileo, whom he scarcely acknowledged, were his predecessors and crucial influences. How it all took place needs to be spelled out in greater detail.

Descartes, as is well known, began as a mathematician and natural philosopher before he turned himself into a metaphysician of mind and an epistemologist of science.[9] We have already shown in Chapter 2 how he developed his conception of mind, or thought and consciousness, as the necessary negative complement to his conception of matter as mechanical and inert.

8 Harry Redner, *A New Science of Representation: Towards an Integrated Theory of Representation in Science, Politics and Art* (Boulder, Co.: Westview Press, 1994).

9 Stephen Gaukroger, *Descartes*, op. cit, 320.

As historians of science have shown, his theory of matter developed out of the hydrostatic studies he had undertaken under the guidance of Isaac Beeckman in Holland; and the hydrodynamic and hydraulic model of matter remained with him ever after, for he rejected the corpuscular view that Hobbes, Gassendi and most physicists, later including Newton, adopted. Analogously, in respect of the representative theory of perception, his guiding light was Kepler, literally so for it was Kepler's optics that inspired his first major scientific work, the treatise on light unpublished in his life-time but appearing after his death in 1664 under the title *Le Monde de M. Descartes ou la Traité de la Lumière*. Light, particularly its projection and reception, was its key concern and from that arose the representative theory of perception.

This was a complete break with all previous mimetic theories of perception from Empedocles to Aristotle and the later scholastics, that is, with all premodern epistemologies. Descartes was led to it by the ever more serious problems that were emerging in the Aristotelian theory of perception and the medieval theories of light which subserved it, those of Alhazen, Roger Bacon, John Peacham, Witelo, and all others down to Kepler. Kepler, on whom Descartes relied without fully acknowledging his indebtedness, first propounded a modern scientific account of light and its reception in the eyes, as Stephen Gaukroger explains:

> Now one reason why the Aristotelian account of perception came to fall apart at the beginning of the seventeenth century was that its account of the processes involved in perceptual cognition, particularly in regard to optics, was shown to be flawed, and flawed in a way that could not be really corrected within the confines of a general understanding of vision. Kepler showed decisively in his *Ad Vitellionem* (1604), for example, that the visual image is formed on the retina and not in the chrystalline humour, and that it is an inverted image. Kepler restricts his attention to the formation of the retinal image, telling us that what happens after this is something quite different from what he is dealing with…[10]

Gaukroger's last point is not strictly speaking correct, for Kepler in fact does indicate what happens after the image is recorded on the retina, but he presents it in a largely metaphorical guise as an incipient theory of visual representation, as he writes:

10 Ibid, 220.

> I leave it to natural philosophers [*Physicis*] to discuss the way in which this image or picture [*idolum seu pictura*] is put together by the spiritual principles of vision residing in the retina and nerves, and whether it is made to appear before the soul or tribunal of the faculty of vision by a spirit within the cerebral cavities, or [whether] the faculty of vision, like a magistrate sent by the soul, goes out from the council character of the brain to meet this image in the optic nerve and retina, as it were descending to a lower court.[11]

Descartes took this up from Kepler, but was no longer hesitant as to how the process proceeds, though he still invoked the old notions of animal spirits. In doing so he arrived at the representative theory of perception; but not without the aid of Galileo, as we shall see.

This went totally against the then accepted Aristotelian account of perception. Aristotle, largely in the *De Anima*, puts forward the view that perception occurs in the senses, not in the brain or mind, and that it consists in the reception of the form of an object without the matter:

> By a "sense" is meant what has the power of receiving into itself the sensible form of things without the matter. This must be conceived of as taking place in the way in which a piece of wax takes on the impress of a signet-ring without the iron or gold; we say that what produces the impression is a signet of bronze or gold, but its particular metallic constitution makes no difference: in a similar way the sense is affected by what is coloured or flavoured or sounding, but it is indifferent what in each case the *substance* is; what alone matters is what *quality* it has, i.e. in what *ratio* its constituents are combined.[12]

The crucial point in this account is that the impression or image must resemble and be a formal likeness of the object perceived. Such a *mimetic* similitude is inherent in all theories of perception prior to Descartes: starting from Empedocles, who maintained that "like is perceived by like"; Democritus, Epicurus and Lucretius, who maintained that perception is the reception of *eidola* or *simulacra* which form images or *phantasia* in the soul; Plato and Aristotle, who give a less materialistic and more formal account of what is received, less *eidolon* and more *eidos*, *morphos*, and *schema*. All later commentators and exegetes of the Classical texts followed this mimetic conception without exception till

---

11 Quoted in Machamer, P.K. and Turnbull R.G., eds. *Studies in Perception* (Columbus: Ohio State University, 1978), 188.

12 Aristotle, *De Anima*, Book II, Chap. 12, op. cit, 204–5.

the age of Descartes and modern science, when the quite different epistemological model was introduced that we have called Representationalism.

How did Descartes manage to go against this weight of historical precedent and develop the representative theory of perception? He initiated the revisionist process by following Galileo in rejecting the view that any mimetic relation of similitude obtains between the object perceived and its sensory quality, as Hiram Caton explains:

> Descartes opposes the Aristotelian-scholastic view that sensations are images or copies of the objects which cause them and which the images depict. The fundamental assumption of the copy theory is that since sensations must resemble objects if we are to know them, the resemblance must be of the most perfect kind if we are to know them perfectly; and the most perfect resemblance would be a copy or picture. Descartes objects that proponents of this theory cannot show us how sensations "can be formed by these objects, received by the external sense organs, and transmitted by nerves to the brain."[13]

Thus Descartes insists that there is no possible mimetic resemblance between the perception and the object perceived since the former takes place in the brain and mind and the latter exists in the external material world, and mind cannot resemble matter. Rather, the relation between them is mediated by representations which are more like signs and symbols rather than likenesses and copies. Language is one such mode of representation in which words in no way resemble the objects they signify, and there are many other such symbolic relations that Descartes draws on to make his case.

Having constituted the representative theory of perception and the epistemology of empirical knowledge based on it, Descartes then had to contend with the inevitable problems it created and which continued to plague Classical philosophy from then on. The main difficulty was to establish how we are to ascertain whether our perceptions are veridical, and therefore how empirical science is possible, if representations in no way correspond by similitude to objects. In general, how are we to know that what we take as perceiving an objective world is in fact any such thing, for it might well be a subjective illusion. Descartes devised one set of answers to these problems which his own theory had created and every other philosopher who followed him, including all the great names of Classical philosophy, propounded a different solution to the same Cartesian problem, which was inherent in the theory itself.

13 Hiram Caton, *The Origin of Subjectivity, An Essay on the Discourses*, (Yale University Press, New Haven, 1973), p. 77.

Descartes' own answer was based on a combination of metaphysical-theological and methodological-scientific premises. To the first problem, the issue of how we can be certain that our perceptions are not all delusory, he replied with the former style of premise that we can rely on God, whose existence can be proven, for God is not a deceiver, otherwise he would be a deceiving demon, namely, the Devil and not the God of Christian theology. As to the second problem, the issue of how we can know that our representations are veridical and in some sense correspond with an external reality, he replied with the latter style of premise and his theory of clear and distinct ideas. By clear and distinct ideas he meant those that are intuitively self-evident, like the *Cogito*, or those that are expressible in rational terms, either logical or mathematical, such as the geometric laws of nature that govern extended substance. This was the crucial Cartesian step in the mathematicisation of the world towards whose accomplishment he contributed Cartesian geometry. It is what Philip Davis and Reuben Hersh have termed "Descartes' Dream" and they go on to show the extent to which modern science and civilization have gone towards its fulfilment.[14] According to the Cartesian approach to the world and Reality, only knowledge couched in mathematical form can be considered certain; quantity and not quality is what counts, for only the former is objective, the latter is purely subjective.

In a complementary way to his reconception of objectivity, Descartes also redefined subjectivity or mind. The subject-object opposition took on a new bearing in his philosophy. He defined subjectivity also in terms of mental representations, insisting in the *Principles* "that our mind is of such a nature that motions which are in the body are alone sufficient to cause it to have all sorts of thoughts which do not give us any image of the motions that cause them."[15] This is how Hiram Caton explains "the origin of subjectivity" in his book by that title. The basic idea was not original, as we previously noted Descartes derived it from Galileo.

Galileo had argued that some of the Aristotelian sensory qualities, such as colour, taste and smell, arose purely from the perceptual process itself and had no correlatives in the objective world, whereas others, which were measurable, such as figure, place and motion did correspond to features in the world. This was later referred to by Locke as the distinction between primary and secondary qualities. It formed the basis for the fundamental separation of the objective from the subjective in Classical science, especially in medicine and psychology; and it played an increasingly important role in all other sciences and modes

14 Philip J. Davis and Reuben Hersh, *Descartes' Dream: The World According to Mathematics* (Boston: Houghton Mifflin, 1986).

15 Quoted in Hiram Caton, *The Origin of Subjectivity*, op. cit, 148.

of discourse where that which was considered purely subjective was given a secondary value compared to that which could be established objectively and could be measured.

The impact of Cartesian philosophy on the Western mind in the modern period was deep-seated and profound and so, too, are the pervasive and difficult to eradicate biases it introduced. These can hardly be called "errors" for without them modern science might have been impossible; they were ineluctably inscribed in the development of the modern sciences and other rational cultural expressions of Western civilization. Nevertheless, we now know that they are mistaken or at least historically outworn and stand in the way of our arriving at better conceptions more suited to the current post-reductivist state of the sciences and philosophy where emergence plays the dominant methodological role, as we have argued throughout this work. It is this residual Cartesianism in our thought and culture which prevents us overcoming the rationalistic, objectivist and reductivist assumptions that underlie so many of the implicitly accepted conceptions about science and philosophy. This is not only true of the natural sciences but also holds for all others as well. As Christina Erneling shows, some of the very latest developments in the psychological sciences and their philosophical expressions are still basically bound to Cartesianism through and through:

> Psychology in general and cognitive science in particular fit in with the dominant trends in Western cognitive science, according to which the basic epistemological problem is one that exists for the individual mind in relation to the external world. To be more precise, this tradition supposes that interaction with other people is not a condition for knowledge; and therefore the individual is not someone who fundamentally shares life, knowledge and language with other human beings. In the early modern period Descartes sanctified "I" and "I think" in his search for the ultimate certainty… Hume's associationism and Kant's account of knowledge in terms of synthesizing the unity of individual understanding, although different from each other in many ways, are also similar in this respect. This kind of individualism and cognitivism is prevalent in cognitive science and, in Shotter's phrase, is the centre of its ideology. For example, Chomsky… speaks of I-language – that is, an internal, intentional, and innate language competence – as the only proper object of the scientific study of language.[16]

16 Christina Erneling, "Challenge to Cognitive Science: The Cultural Approach", in D.M. Johnson, and C.E. Erneling, eds. *The Future of the Cognitive Revolution*, op. cit, 275.

Chomsky explicitly declared himself to be a latter-day Cartesian and developed his theories of language and mind accordingly, as we shall see.

However, it is a long and twisted road that leads from Descartes to Chomsky, one whose milestones are the great names of modern Classical philosophy and psychology. People travelled both forwards and backwards along this highway of the mind. Thus, for example, Chomsky came to Descartes by rejecting the associationism of Hume which he first encountered in his own time in the form of chains of conditioned reflexes presented by Skinner as the explanation of language; he showed that Behaviourism in general is utterly incapable of accounting for language. But Behaviourism, too, derived from Descartes, from another side of his multifarious achievements, this time from his physiological studies in which he was also a great pioneer. It is generally acknowledged by historians of medicine that he was the discoverer of the reflex arc, and that this was the progenitor of Pavlovian Classical reflexes as well as Skinnerian conditioned reflexes. This is another example of the many veritable highways in all directions that lead out from Descartes as so many took up different aspects of his teachings in various fields, adapting them to suit their own philosophical and scientific predilections.

This process of differentiation and division already began in his own lifetime among the Dutch Cartesians, much to his chagrin in some cases. His earliest close follower, the medical professor Henricus Regius, took over Descartes' early physics and physiology but set aside his metaphysics, or rather he substituted his own much more materialistic inclined version, much to Descartes' disgust and fury. On the other hand, Johannes de Rey took over Descartes' metaphysics and showed little interest in the more scientific side of his general system. And so it proceeded from that point on as each major thinker and scientist went on to appropriate, but also to alter and correct, what he took over from Descartes. On the whole, his scientific theories and findings were very soon discarded in their specific form but retained as a general approach. Descartes vortices and his hydrodynamic models did not last beyond Newton's critical review in his *Principia Mathematica*, and soon gave way in physics to the corpuscularism that Newton favoured. The pineal gland location of the mind was almost immediately abandoned by the medical physiologists. But Descartes' resort to the old idea of "animal spirits" in order to explain what he held to be the bodily processes of sensation and emotion took much longer to dislodge, perhaps not till Galvani's electrical findings about muscle movement a century and a half later.

The fate of Descartes' philosophy took a different historical turn. Almost all the later schools of philosophy derived from his work, but not without considerable distortion, omission and adaptation. Thus British Empiricism – culminating in Locke, Berkeley and Hume and their successors – also took its point of departure from Descartes; which tells much against the tenor of later British accounts of the history of philosophy where Empiricism is held to be sui-generis and contrasted to Continental Rationalism. The process of transforming Cartesianism into Empiricism started with the French Cartesian Malebranche, as Gaukroger explains in discussing "the conversion of Descartes into a 'philosopher', that is, into an epistemologist, and the founder of a philosophical school":

> The transformation occurred in a number of stages: Malebranche put a more epistemological gloss on Descartes' doctrines turning philosophy in the direction of epistemology in the late seventeenth and eighteenth centuries in particular because British eighteenth century philosophers tended to read Descartes very much through Malebranche.[17]

Locke, Berkeley and Hume all read Descartes through Malebranche and derived key aspects of their Empiricist philosophies either from Descartes' doctrines directly or from Malebranche's adaptations of these. Berkeley, in particular, owes much to Malebranche's original philosophical theory with its theological connotations that ideas come from God.

Perhaps the best account of the derivation of all the later philosophies down to Kant from Descartes is to be found in the now almost forgotten masterpiece by Gerd Buchdahl, *Metaphysics and the Philosophy of Science*.[18] What this work demonstrates above all else is the close connection between philosophy and science in this period. This relation only began to falter after Kant, in a gradual process of dissociation between the two sides of knowledge. In an earlier work we described this gradually deepening process of secession of science away from philosophy which culminated in the positivistic assumption that science is free of metaphysics and independent of philosophy.[19] In reaction, some philosophers in the twentieth century displayed a disinterest in science that amounted to a deep hostility, as in the irrationalist philosophies of Heidegger and late Wittgenstein. We will briefly outline some of the phases in this unfortunate schism.

The process of dissociation of philosophy and science might be seen to begin in an incipient way at the end of the seventeenth century when Locke with

17 Stephen Gaukroger, *Descartes,* op. cit, 5.

18 Gerd Buchdahl, *Metaphysics and the Philosophy of Science, The Classical Origins: Descartes to Kant* (Cambridge, Mass.: The MIT Press, 1969).

19 Harry Redner, *The Ends of Philosophy*, op. cit, Chapter 1.

disarming modesty distinguished his philosophy from the work of the great natural philosophers, or scientists as we would now call them, of his time:

> The Commonwealth of learning is not at this time without master builders, whose mighty designs in advancing the sciences will leave lasting monuments to the admiration of posterity; but everyone must not hope to be a Boyle or Sydenham; in an age that produces such masters as the great Huygens, and the incomparable Newton, it is ambition enough to be employed as an under-labourer in clearing the ground a little, and removing some of the rubbish that lies in the way of knowledge.[20]

Some of that "rubbish" was undoubtedly the metaphysical teachings of Descartes and Malebranche and other such Cartesians. As Buchdahl explains, Locke had come deeply under the influence of the scientists, "emphasizing and epitomizing in his work the leading traits of their scientific approach as they saw it, above all the primary claim of man's sensory *experience* to be the ultimate arbiter of factual truth – experience rather than spiritual authority or some mysterious form of intuition."[21] Locke thus completed the transition of Cartesianism from its metaphysical to its epistemological form, a process that had already started with Regius during Descartes' life-time, to his great annoyance, and which thereby became Empiricism.

This had to be accomplished if philosophy was to be of any use to science, as Locke avows. Descartes' own philosophy was still full of pre-Cartesian vestiges from earlier sources, derived both from Aristotelian metaphysics and Galenic medicine. He was highly ambiguous and inconsistent about the status of sensations, feelings and emotions, leaving them in the limbo of an indefinite region between mind and matter, and relying on the old notion of "animal spirits" in the body to disguise and cover over this equivocation. Descartes, in other words, was not yet the consistent mind-body Dualist that most philosophers now believe him to have been. A proper Dualism of mind and body came much later, perhaps not till Locke.

Among the many internal inconsistencies in Descartes' philosophy is his failure to integrate his epistemology of sensory perception and his metaphysics of rational thought, as Buchdahl explains:

20 John Locke, *Essay Concerning Human Understanding,* "Epistle to the Reader", ed. Raymond Wilburn (London: Dent, 1947), XXIII.

21 Gerd Buchdahl, *The Image of Newton and Locke in The Age of Reason* (London: Sheed and Ward, 1961), 1.

> Descartes fails in the end to make clear the position of the sensory and qualitative realm in his ontology; or, rather, he leaves the matter opaque and dogmatic. God does not deceive me concerning the apparent order of created things: very well! But it does not follow that the realm of sense is ultimately analysable in the way Descartes suggests when he calls it "confused thinking". Here as before in his scientific methodology, with its hard core of irreducible initial conditions and tentative hypotheses, Descartes has not been able to do more than suggest by a chosen language that the detail of "phenomenal" particulars is amenable to the over-all construction of reason. The tension that exists between his rational ideals and the claims of sensory experience is now handed on to his philosophical successors.[22]

Buchdahl's whole book deals with how Descartes' philosophical successors above all Locke, Hume and Kant, stepped into the breach to fill the lacunae and ambivalences in his philosophy and thereby devised their own solutions to the problems he had bequeathed to them, none of them more so than Kant.

It would be futile to even attempt to summarize what Buchdahl has to say about various philosophers in the Cartesian tradition. His accounts are thorough, detailed and exhaustive; they have to be read in full to be followed. Buchdahl leaves us in no doubt of the measure of Descartes' achievement nor of its shortcomings, as is inevitable in such a revolutionary innovator:

> It is a measure of Descartes' pioneering work (not for nothing has he been called "the father of modern philosophy") that in his writings all these elements are still encompassed in their freshness, and with their resulting tensions. Here, these tensions are only half-resolved, for the outlines of the various philosophical concepts are as yet less definite, the resulting syntheses vaguer, but also less rigid than in later philosophers. Moreover, despite appearances to the contrary, *all* the conflicting elements are preserved in Descartes' system.[23]

Unfortunately, Buchdahl's effort to trace the historical course of Cartesianism stops with Kant, he died too soon to be able to take it any further.

Beyond Kant another epoch of philosophy begins to dawn, starting with the German Idealists, who were averse to the natural sciences, epitomised in this period by Newtonian physics, and intent on returning to a more speculative

22 Gerd Buchdahl, *Metaphysics and the Philosophy of Science*, op. cit, 119.
23 Ibid, 81.

state of "natural philosophy". However, Descartes still remained a major presence through his emphasis on subjectivity and the Ego. Kant had already signalled the central role of the Ego by insisting that all mental content is accompanied by an implicit "I think". Fichte, his Idealist successor, expanded the role of the Ego to make it the constructive principle of Reality, much to Goethe's horror for he mocked it in his *Faust*, Part 2. Kant, no doubt, would have been equally appalled. But out of this Fichtean start there emerged the school of German Idealists who dominated philosophy in the nineteenth century, and their various successors continued well into the twentieth. They were generally averse to Cartesianism though they recognized Descartes' role as the initiator of modern philosophy. Hence, we find critiques of Descartes in Hegel, Schopenhauer, Nietzsche and later dismissively in Heidegger. Heidegger derived from Phenomenology which was founded by Husserl. He established it as a synthesis of Cartesian subjectivity, Kantian transcendentalism, Fichtean Idealism, and, most crucially of all, the doctrine of intentionality derived from Brentano. Brentano was harking back to neo-Aristotelian scholastic teaching on the intentional object inherent in mental acts and he generalized this as a definition of the mind. This became the point of departure for Husserl in developing Phenomenology.

Husserl was extremely ambivalent towards Descartes and Cartesian science. Towards the end of his life, in his Introduction to the aptly named *Cartesian Meditations,* Husserl declares himself to be a follower of Descartes, while at the same time rejecting nearly all his specific doctrines:

> René Descartes gave transcendental phenomenology new impulses through his *Meditations*; their study acted quite directly on the transformation of an already developing phenomenology into a new kind of transcendental philosophy. Accordingly one might almost call transcendental phenomenology a neo-Cartesianism, even though it is obliged – and precisely by its radical development of Cartesian motifs – to reject nearly all the well-known doctrinal content of Cartesian philosophy.[24]

This view of himself as a Cartesian *malgré tout* has been both contested and supported by later scholars. Herman Philipse is one scholar who upholds it, as he writes:

24 Edmund Husserl, *Cartesian Meditations: An Introduction to Phenomenology,* trans. Darion Cairns (The Hague: Nijhoff, 1969), 1.

> Husserl endorsed the Cartesian predicament in epistemology already in the *Investigations.* Epistemology should be based on an analysis of the indubitable sphere of my own consciousness. What is more, Husserl's gradual development towards transcendental idealism between 1900 and 1913 may be seen as a progressive radicalization of the Cartesian requirement of finding an indubitable foundation of knowledge by means of doubting. Summarising the argument of his lectures of 1907, Husserl wrote that all he wanted was to "grasp purely and develop consistently what was already implied in this very old intention of Cartesian doubt."[25]

Cartesian doubt is what Husserl transformed into his *epoché*, both deriving from ancient sceptical sources and both serving as the propaedeutic introduction to science. But what Husserl means by science is completely contrary to what Descartes took it to mean, as he explains in his *Cartesian Meditations*:

> Descartes himself presupposed an ideal of science, the ideal approximated by geometry and mathematical natural science. As a fateful prejudice this ideal determines philosophies for centuries and hiddenly determines the *Meditations* themselves.[26]

For Husserl, by contrast, science meant "*strenge Wissenschaft*" achieved by means of eidetic reduction and *Wesensschau* or essential vision which might be taken as a kind of Cartesian intuition.

There has always been a tension between Husserlian science or Phenomenology and Cartesian science or what we now call natural science. This tension is particularly fraught when dealing with the sciences of the mind. According to Phenomenological strictures, the contents of the mind should be studied by an essential vision that brackets away, that is discounts, all naturalistic or scientific knowledge. But according to the scientific view of things, the contents of the mind should be causally explained by reference to neurology, psychology and the Cognitive sciences. Bringing these two clashing perspectives on the mind together seems almost as impossible as squaring the circle.

Nevertheless, a prominent school of largely French Phenomenologists and Cognitive scientists are determined to do so. First at a conference in Paris and then in the printed proceedings they set out their project for naturalizing

25 Herman Philipse, "Transcendental Idealism", in Barry Smith and David Woodruff Smith, eds. *The Cambridge Companion to Husserl* (New York: Cambridge University Press, 1995), 281.

26 Edmund Husserl, *Cartesian Meditations*, op. cit, 7.

Phenomenology, as they put it.[27] Predictably, in their Introduction to the collective volume, the two leading spirits of this enterprise, Jean Petitot and Francisco Varela, begin with the confrontation over science between Descartes and Husserl. They note the similarity between Cartesian doubt and Husserlian *epoché* and arrive at the conclusion that it is only the Cartesian conception of science that stands in the way of a reconciliation of Phenomenology and science. They believe that contemporary developments in the sciences have surpassed the Cartesian view and make Husserl's reluctance to relate Phenomenology to science anachronistic and no longer relevant.

Husserl bases his view of natural science not so much on Descartes himself as on his precursor Galileo, as Petitot and Varela note:

> Similarly, in Husserl's opinion, starting with Galileo, physics attempted for nature what mathematics had already achieved for geometry. It tried to axiomatize the sciences of material reality. But here there is a reverse side to this ideal. Just as with geometry, first the Heraclitean flux of sensible morphologies, and then the sensible qualities across which phenomena are concretely given, had to be excluded. These proto-geometric morphologies and these secondary qualities were considered impossible to mathematize. And Husserl thought this limitation of the Galilean sciences was an insuperable one.[28]

According to Petitot and Varela, "Husserl's position is the result of his having mistaken certain contingent limitations of mathematical and material sciences of his time for absolute ones."[29] But fortunately, according to these authors, the progress of science in our time has overcome such limitations, making it possible "for the physico-mathematical sciences to offer a scientific reconstruction of the phenomenality of the surrounding world."[30] The new sciences they have in mind are the Cognitive sciences and other new developments in the sciences of complexity, such as we previously referred to in Chapter 2. They hint that we now stand on the threshold of a new scientific revolution that is comparable to that which occurred in Descartes' time.

Even though one needs to exercise considerable scepticism about many of their pronouncements, yet there is much to be said for their idea of naturalizing

27 Jean Petitot, Francisco J. Varela, Bernard Pachaud and Jean-Michel Roy, eds. *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science* (Stanford: Stanford University Press, 1999).

28 Ibid, 40.

29 Ibid, 42.

30 Ibid, 40.

Phenomenology. We shall take up this new development in philosophy in the next chapter where we shall attempt to show that it is one of the counters to Cartesian Materialism which has assumed a dominant position over American philosophy and in many other departments. The term Cartesian Materialism might strike one as an oxymoron as it seems to combine two contradictory notions, Dualism and Materialism. However, it is historically justified since over the centuries an amalgam of Cartesianism and Materialism, which began as opposed tendencies, did eventuate. In other words, the philosophies of Gassendi and Hobbes, which were so opposed to that of Descartes' at the start in the seventeenth century, were in time, certainly by the twentieth century, absorbed into it. It happened through gradual stages, as we shall proceed to show next.

Descartes was not only instrumental in shaping the form in which later Empiricism and Idealism were cast, but also, paradoxical though this might sound, Materialism as well. A form of Cartesian Materialism was first enunciated by Julien Offray de la Mettrie; and, as Caton notes, "it is not irrelevant to observe that the author generally credited with having stated the man-machine doctrine, la Mettrie, himself acknowledges having found this doctrine in Descartes' writings."[31] It is precisely the man-machine doctrine that marks this Materialism as Cartesian and, therefore, modern as opposed to all earlier forms of Materialism that derived from ancient atomist sources.

All ancient forms of Materialism are based on the view that mind is a special kind of matter or a different trajectory of matter as compared to all other entities. Thus Aristotle reports of the earliest Materialist, Democritus, "that the soul is a sort of fire or hot substance; his 'forms' or atoms are infinite in number; those that are spherical he calls fire and soul …; the spherical atoms are identified with soul because atoms of that shape are most adapted to permeate everywhere, and to set all the others moving by being themselves in motion."[32] Much later, Lucretius, expounding the Epicurean physics derived from Democritus, also declares that "the mind and soul have a material nature", which is atomic and "surpasses every existing thing in its mobility, in its subtlety, and in the smallness and smoothness of its atoms…"which are also the ones that swerve.[33] Renaissance Materialists, such as Telesio, invoked animal spirits as a superfine matter to account for the soul, which is perhaps what Hamlet also means by "the quintessence of dust".

31 Hiram Caton, *The Origin of Subjectivity*, op. cit, 100.

32 Aristotle, *De Anima*, Book 1, Chap. 2, op. cit, 150.

33 Lucretius, *On the Nature of Things*, trans. Martin Ferguson Smith (London: Sphere Books, 1969), 101 and 103.

The man-machine doctrine, which la Mettrie derived from Descartes, fundamentally broke with all such views that soul is a different kind or motion of matter by affirming the mechanistic hypothesis that all matter is extension and, therefore, the same in whatever form it comes. Descartes view that the body is a machine amounted to a conception of it as an automaton, an arrangement of parts into a causally interconnected self-moving device just like a clock. La Mettrie simply extended this basic Cartesian metaphor of the body as automaton to the mind as well. Man was in principle no different from a marionette, such as Jacques de Vaucanson's mechanical toys, only much more complicated. All this was not very convincing and few accepted it at the time, and would not do so till the design of automata had proceeded much further; this did not happen till the invention of the electronic computer during and after the Second World War, as we shall see in the next chapter.

But for the computer to be accepted as a simulacrum of the human mind and for the view of mind as machine to gain any traction, a large number of developments had to take place, both intellectual and practical, in philosophy, science and technology. As Gaukroger puts it, "materialism did not look remotely plausible before the 1860s, when Helmholtz began his path breaking work on cognition."[34] But even then it was still not very convincing, and it took much further work in physiology and neurology, beginning with Claude Bernard in the 1860s and continuing with Walter Cannon's theory of bodily "homeostasis" in the 1920s, before the body and brain could be seen as self-regulating automata. The continuation of Descartes' discovery of the reflexes, first by Pavlov and then by Watson and Skinner, pointed in the same direction. What finally made Cartesian Materialism acceptable to so many was the development of "logic machines" which seemed to be capable of simulating cognitive functions. Finally, it seemed, machines could think and those who were most active in this endeavour, such as Alan Turing, were convinced of it; Turing even claimed to "prove" it with his famous Turing automaton.

The stage was set in the post-war climate in America and Britain for the thesis of mind as machine to be taken up once again, and armed with the latest logical and technological resources sweep all before it. New mechanistic movements arose, such as first Cybernetics and Information Theory, followed very soon after by Artificial Intelligence and the Cognitivist revolution. The key exponents of the latter were Marvin Minsky and Noam Chomsky at MIT. We shall discuss this whole development in the next chapter.

34 Stephen Gaukroger, *Descartes*, op. cit, 289.

The outstanding Cartesian of our day and age is undoubtedly Chomsky, a self-professed one as the title of one of his books, *Cartesian Linguistics,* proclaims.[35] An exposition and critique of Chomsky's Cartesianism both in his theory of language and his philosophy of mind has been provided by one of his former excommunicated students, George Lackoff, in a book written jointly with Mark Johnson.[36] Obviously there are vast differences between the mid-twentieth century linguist and the mid-seventeenth century philosopher, but in spirit and intent the former follows the latter in essential matters. Thus, as they show, despite the fact that Chomsky does not subscribe to Descartes' two substances doctrine, his account of mind is nevertheless clearly Cartesian in orientation. The computer model of the mind that separates software from hardware made possible a kind of Dualism of mind as Information and language opposed to body as brain mechanisms.

Under a number of headings Lackoff and Johnson present "some of the key components of Descartes' view of mind and reason that were appropriated and adopted by Chomsky."[37] Here we can merely list the main rubrics without the relevant explanations for which the reader must refer to the original text. These headings alone constitute an impressive list of features constituting Chomsky's debt to Descartes: separation of Mind and Body, transcendent autonomous reason, essence, and rationality define human nature; mathematics as ideal reason, reason as formal, thought as language, innate ideas, and the method of introspection. Lackoff and Johnson are highly critical of Chomsky on all these points for they are exponents of an anti-Cartesian bodily philosophy of mind, as they explain:

> Cartesian philosophy is inconsistent with findings about embodiment of mind. The mind is embodied not disembodied… And most of our abstract concepts structured by metaphor – time, causality, even our concept of mind itself – are grounded in bodily experience. Abstract reason is not an autonomous, body-free faculty of mind.[38]

And neither is language itself or its grammar, according to Lackoff and Johnson.

35 Noam Chomsky, *Cartesian Linguistics* (New York: Harper and Row, 1966).

36 George Lackoff and Mark Johnson, *Philosophy in the Flesh, The Embodied Mind and its Challenge to Western Thought* (New York: Basic Books, 1999).

37 Ibid, 470.

38 Ibid, 495.

They go on to argue against Chomsky's Cartesian linguistic theory on this basis:

> The embodiment of concepts also contradicts the formalist philosophy that underlies Chomsky's linguistic theory. Because concepts are embodied and not just symbols, thought is not "linguistic", not just a matter of symbol manipulation. Concepts are not adequately representable by meaningless symbols. The Thought as Language metaphor does not reflect scientific truth. Meaning arises through the body and the brain, not via the disembodied connection of symbols with the world (or a set theoretic model of it).[39]

They present a convincing *exposé* of Chomskian linguistics as an amalgam of Cartesian philosophy and twentieth century formal logic applied to language. On the latter aspect, they contend that "the principal technical idea of Chomsky's theory is that of formal language, which was developed in mathematical logic."[40] They go on to state that "the mathematical theory of formal languages, developed by Emil Post, was the direct inspiration for Chomsky's formal theory of language and was the mathematical setting for that theory."[41]

Cartesianism is very compatible with formalism; hence Chomsky had no difficulty in combining both together. He simply applied ideas from formalization theory to a Cartesian view of language, as Lackoff and Johnson explain:

> Language takes on for Chomsky the role of reason in Descartes' philosophy; that is, language becomes the essence that defines what it is to be human. Language is mathematical in nature, and since mathematics is a matter of pure form, language for Chomsky is purely formal. Language is also universal and innate, an autonomous capacity of mind, independent of any connection to things in the external world. Language must have an essence, something that makes language what it is and inheres in all languages. That essence is called "universal grammar", it is mathematical in character and a matter of pure form.[42]

39 Ibid, 495.
40 Ibid, 472.
41 Ibid, 472.
42 Ibid, 472.

They offer a devastating critique of nearly all of Chomsky's linguistics, showing it to be less of a science of real languages, rather an *à priori* attempt at the formalization of language as such, which is bound to fail, since grammar is inseparable from meaning and semantics and cannot be detached from these, as Chomsky attempts to do. Hence the real grammar of any actual language is always subject to the unpredictable and unregulated vagaries of meaning and language use in concrete situations. "Universal grammar" is a mere fiction that results from Chomsky's philosophy and his formalizing procedures, not something that can be shown to exist by reference to language as a communication medium in society. Though it might be pointed out in criticism that Lackoff and Johnson are themselves not sufficiently alert to the social nature of language and meaning,which cannot be identified with the body, as they tend to do, and as we shall also argue against Thompson in the next chapter.

It is evident that the Cartesian tradition is far from dead and that Descartes' ideas have made a surprising comeback in the guise of logical formalism and the computer model of the mind. How and why such a remarkable transmogrification could take place in the second half of the twentieth century we shall examine in the next chapter. First we shall describe historically when and where it took place, then we shall subject it to a critical review so as to reveal its inadequacies. For only if this moribund Cartesian legacy is dead and buried will it be possible to search for an alternative view of the mind, such as we have sought to provide in the previous chapters.

What is at stake in this endeavour is of some consequence not only academically speaking or in respect of philosophy alone but also for the sake of preventing the worst consequences of the current attempts at a mechanization of mind invading our mentality and culture in general. This is a most daunting undertaking, for just as it is difficult to overestimate the historical consequences of the transition from the Aristotelian to the Cartesian philosophy, so it is conversely easy to underestimate the effort that will be required to free ourselves from the pervasiveness of the debased Cartesian influence. Aristotelianism generated scholasticism and shaped the whole intellectual bearing of the universities from their foundation till well into the nineteenth century; so analogously Cartesianism in its later guises shaped the sciences and took over the universities during the nineteenth and twentieth centuries. It is now inherent in our whole culture and ingrained in our minds. But the mental fight against it must begin even now. As an ally in this battle for the mind we turn to Spinoza, a near contemporary of Descartes, who was strongly influenced by him, but at the same time he initiated the critique of Cartesianism which we might now take up and further develop.

2 From Spinoza to the Present

Spinoza's philosophy of mind is, of course, based on Descartes' from which it originated. But at the same time it goes against Descartes', departing from it and transcending it. It is far more compatible with a contemporary Emergentist approach, and that is how it has been received by a number of contemporary scientists and philosophers, as we have previously noted. But before we can begin to discuss any of these, we must first present a brief outline of Spinoza's thinking on the relation between mind and body.

Fortunately, a new book on Spinoza by Elhanan Yakira enables us to avail ourselves of up-to-date scholarship in this regard. Yakira confirms our previous characterization of Spinoza as a Janus-faced thinker looking both backward into the past and forward into the future:

> A careful reading of the first thirteen propositions of *Ethics*, Part 2 will show, I believe, that what Spinoza says in them shares with the pre-modern thinkers, both Jewish and non-Jewish, a general philosophical concern for these questions, and in this sense his mind-body doctrine belongs to the same fundamental inquiry… However, it does so in a thoroughly unorthodox and original way, both radical and secular (hence modern) and deeply anti-Cartesian. The great iconoclastic significance of Spinoza's doctrine lies in the fact that his *noétique* is launched by immediately turning to the *body*, and by positing it as the foundation of the theory of rational thinking. This gesture determines Spinoza's philosophy as a thoroughly iconoclastic venture vis-à-vis both pre-Cartesian and Cartesian tradition.[43]

As Yakira brings out, there is something of once very ancient and very contemporary in Spinoza's philosophy of the mind-body relation. This has also been stressed by Henri Atlan who writes that "the philosophy of the *Ethics* is an indispensable link between a natural philosophy, re-embedded in the present, and the ancient doctrines that understood ethics on the basis of knowledge of the determining cause of the nondissociated body-mind…"[44] Atlan celebrates what he takes to be the posthumous victory of Spinoza over Cartesian Dualism, as well as the later Idealism of Leibniz and Kant:

43 Elhanan Yakira, *Spinoza and the Case for Philosophy* (Cambridge: Cambridge University Press, 2015), 95.

44 Henri Atlan, *The Sparks of Randomness*, op. cit, 13.

> In the *Ethics* he waxed ironic about this dualism, in which he saw the perfectly incomprehensible image of man in nature as "a kingdom within a kingdom". Categorically denying any explanatory value to the anthropomorphic invocation of final causes in nature, he endeavoured to base the *Ethics* on a physics of the human body. This physics, which he himself described as weak, has given rise to much misunderstanding, because it has been taken for a mechanism or dynamic that should be compared with the physics of Descartes, Leibniz, and Newton. In fact a physical theory of the organism (and thus a first attempt at what we would call today the biophysics of organized systems), it has lost nothing of its evocative significance.[45]

Whether we can take Spinoza's call for a science of the body as a prefiguration of a contemporary biophysics of organized systems is perhaps somewhat anachronistic and questionable from a history of science point of view. After all, Spinoza knew nothing of biology as it later developed or of evolution. But, nevertheless, he might have anticipated a philosophy of mind that only came into force as a result of such scientific discoveries closer to our own time.

The attractions of Spinoza's philosophy to scientists are the opposite of those of Descartes' philosophy. Descartes, with his mechanistic philosophy of matter, appealed strongly to physicists, even though his own physics based on that of motion through direct mechanical contact was soon displaced by the Newtonian idea of action at a distance. However, neither type of physics is well suited to biology, and it is here that Spinoza began to come into his own. His concept of the *conatus,* or the tendency of entities to maintain themselves in being, is what is primarily encountered in biological organisms. His emphasis on the body is also well suited to biology, as well as psychologies that focus on embodiment such as the Merleau-Ponty variants of Phenomenology. Of all the Classical philosophers Spinoza comes closest to our emergentist view of a non-causal relation between mind and body.

Spinoza, of course, had no idea of emergence, which arose much later. However, what he does say about the mind-body relation fits in well with our previous stipulation of the logic of emergence. Spinoza puts it as follows:

> …whatever can be perceived by an infinite intellect as constituting an essence of substance pertains to one substance only, and consequently that the thinking substance and the extended substance are one and the same substance, which is now comprehended under this attribute, now

45 Ibid, 93.

> under that… Therefore, whether we conceive nature under the attribute of Extension, or under the attribute of Thought, or under any other attribute, we shall find one and the same order, *or* one and the same connection of causes, i.e. that the same things follow one another.[46]

This is, of course, a celebrated passage which has been interpreted in various ways by different commentators. Nevertheless, without unduly forcing the meaning we can take it in terms of our preceding discussion of the logic of emergence, allowing, of course, that Spinoza himself could not have taken it in that way.

Expressed in our own terms, Spinoza's statement amounts to the following: Mind and matter – or in Spinoza's Cartesian terminology, thought and extension – are attributes of the one substance. We can now interpret "substance" as emergent system or organism as a whole. The two "attributes" of this system can be taken as the emerger as bodily material base and the emergent as mental outcome. Thus the two sides of the emergent system can be considered as its constituent "aspects" or "parts" making up a whole or substance. They are "parts" of the one system or organism, yet they are not identical, though indissolubly linked. This follows in accordance with the logic of emergence which stipulates that this relation is necessary, not contingently causal.

Spinoza, too, contrary to Descartes, insists that there can be no causal relation between body and mind. There can be no causal nexus since "the Body cannot determine the mind to thinking, and the Mind cannot determine the Body to motion, to rest or to anything else (if there is anything else)."[47] According to Spinoza, each follows a causal sequence of its own; there are chains of material causes and chains of mental causes, each keeping step with the other. But Spinoza does not require them to be strictly isomorphic with each other, as Leibniz later maintained.

It is a feature of Yakira's account that he not only distinguishes Spinoza's mind-body thesis from that of Descartes, but also from that of Leibniz, both of these being dualist positions, which Spinoza's is not. As Yakira puts it:

> Spinoza's great originality consists in the role he assigns to the *body* in his theory of ideas. This is where he parts ways from Descartes, from the occasionalists [Malebranche], and from Leibniz, but also from the empiricists [Locke]. The doctrine of the metaphysical nature of ideas, as it is

46 *The Collected Works of Spinoza*, ed. and trans. Edwin Curley (Princeton, NJ: Princeton University Press, 1985), 451.

47 Spinoza, *Ethics*, Part 3, proposition 2, op. cit, 494.

> expounded in the first part of *Ethics II*, is supposed to be a detailing and concretization of the general metaphysical principle. As such it is a first step toward the elaboration of a theory of man as a knowing being; a sensing, feeling and desiring being capable of wisdom and freedom.[48]

Yakira interprets Spinoza as affirming the thesis that mind or soul is the idea of the body, given his reading of the concept of an "idea". As he puts it:

> At its centre stand two counterintuitive, perhaps absurd claims: (1) the soul is not really a *soul* – that is, a *sui generis* kind of being, a first ontological principle, or a substance – but an idea; (2) this idea's object is no other than *the* "body" – that is, the very same "certain" body that one "senses" affected in many ways, and that one conceives as, or knows to be, "one's own body".[49]

The point of this reading of Spinoza is to affirm his opposition to the Cartesian notion that the mind is a substance, thought, that stands opposed to another substance, matter or extension. It is also, as it were in anticipation, a counter to the later Leibnizian idea of a pre-established harmony; or, as it still later became known through the agency of Gustav Theodor Fechner in the nineteenth century, as a psycho-physical parallelism. For that, too, supposes that there are two separable entities that proceed in parallel with each other.

Yakira, who is also an expert on Leibniz, explores at length the variant positions of Descartes, Spinoza and Leibniz. He is particularly intent on disproving the thesis that Spinoza can be interpreted as advocating a parallelist approach, which is how his famous pronouncement in *Ethics* II, proposition 7 is now often read. He argues decisively that parallelism is a Leibnizian idea which Leibniz in fact developed to counter Spinoza, but which later exponents of this idea, like Fechner, wrongly ascribed to Spinoza himself. Fechner was intent on developing the new discipline of psychophysics, and to do so he borrowed the term "parallelism" from Leibniz to refer to the relation between mental and physical phenomena, which he conceived of in Leibniz's manner as two kinds of distinct phenomena, even though they were "manifestations of one and the same entity – an individual organism."[50] This, Yakira holds, is totally against Spinoza's intent of how the mind-body relation is to be taken. "Parallelism" belongs to Leibniz's philosophical outlook, which is fundamentally dualistic. It

48 Elhanan Yakira, *Spinoza and the Case for Philosophy,* op. cit, 116.

49 Ibid, 116.

50 Ibid, 65.

is based on the formalistic and rationalistic presupposition that unity is a matter of "isomorphism, or identity of structure, existing between two disparate entities or realms of entities, [which] is the basis of Leibniz's doctrine of *expression*, or of the *pre-established Harmony*."[51] Leibniz worked with a "formalistic conception of logic and truth" and sought "to generalize the essential role *structure* and *order* have in the work of reason".

This kind of reasoning has been taken up by many logically and formalistically inclined thinkers and scientists in the twentieth century from Russell to early Wittgenstein and Chomsky, to whom the discovery of isomorphic structural relations such as the mapping or modelling of one phenomenon in terms of another is an adequate scientific explanation. Thus for many contemporary Cognitivists the production of computer models of mental phenomena is all that is involved in accounting for them. The "modelling of mind" enterprises, as in Artificial Intelligence, are thus the distant progeny of Leibniz's formalistic and rationalistic endeavour. We shall return to this in the next chapter and show what is wrong with taking isomorphic modelling as a norm of scientific explanation, for as we shall see two phenomena can share the one formal model yet be utterly different, as for example, in Catastrophe Theory developed by René Thom in which the same set of mathematical equations covers utterly anomalous phenomena which have nothing to do with each other. In such dissimilar cases as dogs attacking and civilizations collapsing both can be represented mathematically in the same way, yet this structural isomorphism explains nothing.

Furthermore, psycho-physical parallelism taken strictly in a literal sense creates unnecessary problems for the mind-body relation which it is possible for Emergentism to avoid. A strict definition provided by Thompson runs as follows:

> Analytical isomorphism is the idea that successful explanation requires there to be isomorphism (one-to-one correspondence) between the phenomenal content of subjective experience and the structure or format of the underlying neural representation.[52]

According to this definition of psycho-physical parallelism, there must be physical events which are isomorphic with every aspect of a mental phenomenon; to be truly parallel what happens in the brain must mirror exactly down to the last detail what happens in the mind. But this might in fact not prove to

51 Ibid, 63.
52 Evan Thompson, *Mind in Life*, op. cit, 272.

be the case, that is, it might not be possible to establish this with any available scientific means. The physical correlates of a mental phenomenon might not match it exactly and precisely due to a number of insurmountable difficulties: what takes place in the brain might be too complex to be specified or followed with total exactitude; it might depend on initial conditions that are irrecoverable; what takes place in the brain is bound up with what takes place in the body as a whole and the environmental context creating a complexity in which no exact and isomorphic parallels to mental phenomena can be discovered. To declare that parallelism must necessarily exist is simply to make an *à priori* assumption of principle that is meaningless in practice.

Emergentism works with a looser fit between mind and brain and does not require an exact mirroring relation of point by point isomorphism. It is enough if close correlates are discovered that permit us to understand how emergence takes place. It is sufficient for the emergence of mind if there is a material basis for it in the brain such that the former varies in accordance with variations in the latter. These far looser forms of correlation are in fact being found, but there is no guarantee that this will continue into the indefinite future. It is conceivable that at some time in the future the discovery of more exact correlations will cease. This would certainly put paid to psycho-physical parallelism, but not necessarily to Emergentism. It would still be possible to contend that mental phenomena emerge from a physical basis without being able to specify an exact isomorphism between the two. But we are still far from such a contingency at present when so little is known about how the brain functions and how it is connected to all that is going on in the body and its environment.

Damasio's account of Spinoza more or less matches our own, as we have already indicated by enclosing a long quotation from his book on Spinoza in the *Prelude*. In that quotation it is highly significant that Damasio insists that "in a strict sense, the mind did not cause the body and the body did not cause the mind."[53] He, too, points to the crucial feature that "thought and extension, while distinguishable, are nonetheless attributes of the same substance…"; and he goes on to comment that "the reference to a single substance serves the purpose of claiming mind as inseparable from body, both created from the same cloth, as it were."[54] And, finally, he also refers to "an entirely sensible "aspect" dualism, [that] rejects substance dualism", that is, Cartesian Dualism.

Damasio backs up his version of Spinoza with a wealth of neuro-psychic research about the mind-body relation. The conclusion he comes to from all these scientific findings is that they are in keeping with the Spinozist thesis

53 Antonio Damasio, *Looking for Spinoza*, op. cit, 209.
54 Ibid, 209.

that "body, brain and mind are manifestations of a single organism."[55] The most striking of these observations is that when patients lose consciousness of the body, they also lose consciousness as such, namely, they fall unconscious. And that "when the perception of *extensive* sections of the body is disturbed, even if temporarily, the cost to the patient is always some measure of mental disarray."[56] This provides indirect empirical evidence against any idea that the mind is localized in or depends solely on the brain or that mental acts can be simply identified with neural processes. There is much more to mind than the brain by itself; what transpires in the body also matters; and even beyond the body, the whole physical and social context counts as well, for the brain is inseparably connected to both bodily and external environments.

As a neuroscientist Damasio is most intent on establishing in neurological terms how the brain keeps track of what is happening in the body, as well as in the external environment. He explains this process in terms of the neural patterns or brain "images" that "are constructed according to the brain's own rules, and are achieved for a brief period of time in the multiple sensory and motor regions of the brain."[57] How such brain "images" become conscious images or feelings in the mind is, as he readily admits, a neurological mystery, for obviously it cannot be explained in such terms. It is, of course, the "mystery" of emergence, a concept that Damasio does not employ and does not seem to understand even while he uses the term in a colloquial sense. He writes that "it is not unreasonable to conceive of the mind emerging from the cooperation of many brain regions"; but immediately goes on to add that "the knowledge gap we now recognize [between brain and mind] may turn out to be little more than a discontinuity in the complexity of the accumulated detail, and in the complexity in the interaction of the brain regions involved in the mapping."[58]

If all that emergence amounts to is merely a matter of such a "knowledge gap", then one could expect this to be filled and, consequently, that eventually mind would be reduced to brain. But if emergence obtains, then this could never happen, for emergence is irreducible; if it were subject to reduction, then it would no longer be emergence. But the emergent relation of mind to body is not one of mere complexity that could ever fall prey to reduction since it is almost inconceivable to us how this might be brought about, apart from some unsupportable and unsustainable eliminationist thesis that mind or our own subjectivity does not exist. The biologist Lynn Rothschild is of the same

55 Ibid, 195.
56 Ibid, 193.
57 Ibid, 199.
58 Ibid, 208.

view when she states: "whether it [the reductionist approach] will ultimately explain everything about consciousness, thereby establishing that consciousness is not an emergent property, is, in my view, unlikely."[59] If emergence obtains, then it is irreducible; which does not mean that it is scientifically inexplicable, for there are other forms of scientific explanation apart from reduction. As we have already argued, the apparent "mystery" of synchronic emergence can be elucidated in terms of the two forms of diachronic emergence, the ontogenetic and the phylogenetic, with which it is bound up. That irreducibility does not mean inexplicability is a point Damasio does not seem to have grasped.

In fact, the main point of Damasio's own book on Spinoza is to explain the emergence of emotions and feeling as, indeed, the subtitle, *Joy, Sorrow and the Feeling Brain*, affirms. In this work Damasio adopts a variant of William James's scientific theory that it is the behavioural and physiological expressions or symptoms which produce the feeling quality of the emotional experience; which runs counter to the common sense view that it is vice-versa, that the feeling gives rise to the expressions and behaviour. As we shall presently show, both are in fact right. Damasio's variant on this Jamesian scientifically accepted theory is that emotions are in effect perceptions of bodily states; and he remarks that objections to this account are "perhaps the last remnant of resistance to William James's conjecture that when we feel emotions we perceive body states."[60] The theory itself is not remarkable, but what is noteworthy from our point of view is that despite adhering to the Spinozist thesis that "in a strict sense, the mind did not cause the body and the body did not cause the mind"[61], Damasio has no hesitation in constituting two way causal relations between body and mind, and construing them as reciprocal causal chains between the mental and the physical. As he puts it:

> The processing of emotions involves this dual track: the flowing of mental contents that bring along the triggers for emotional response, and the executed response themselves, those that constitute emotions, which eventually lead to feelings…Curiously, by the time the process reaches the stage of assembling feelings, we are back in the mental realm – back

59 Lynn J. Rothschild, "The Role of Emergence in Biology", in Richard Clayton and Paul Davies, *The Re-emergence of Emergence*, op. cit, 163.

60 Damasio, *Looking for Spinoza*, 105.

61 Ibid, 209.

> in the flow of thoughts, where, in normal circumstances, the entire emotional detour began.[62]

This is no contradiction. Emergent relations *per se* are not causal; but for practical or scientific purposes causal links can be constituted across the logical gap separating phenomena at different levels; we build causal bridges across emergent divides in order to be able to cross from one side to the other for ordinary pragmatic or scientific convenience. But the two types of contexts must be clearly distinguished for quite different bridging criteria apply. This is particularly so with respect to emotions. In the ordinary pragmatic usage it is the perception of the danger which comes first, the feeling of fear which comes next, and the fight or flight behavioural response follows. Hence, in ordinary speech we say that fear causes one to run away, for it is the over-all situation that matters, not its constituent details. But in the scientific context of the laboratory, where minute bodily reactions are being precisely measured, it is the bodily response that comes immediately after the perception – which might not in the first split second even be conscious – then there comes the conscious feeling of fear, and the behaviour follows. Hence, in a scientific context James and Damasio are right; it is the initial bodily state which causes the feeling of fear.

The reversibility of cause and effect relations between mind and body is particularly prominent in genetic research. Scientists normally assume that it is genes that cause mental phenomena, as when they say, quite correctly, that the mental delusions and hallucinations of schizophrenia are caused by genetic disorders. But sometimes it is vice-versa, as when Horgan reports on the work of Eric Kandel:

> [Kandel] noted in his experiments and others had shown that experience produces physical changes in neurons. More specifically, habituation or sensitization can turn genes on or off and otherwise effect their expression. These findings implied that experience, such as traumatic events in childhood, could cause neurosis through both neurochemical and genetic effects.[63]

In this context it is clear that by "experience" mental events are meant and by neurochemical and genetic effects physical ones; and that it is the former that cause the latter. Such epigenetic findings have been confirmed not only in

62 Ibid, 65.

63 John Horgan, *The Undiscovered Mind,* op. cit, 41.

humans, but also in animals, as Horgan goes on to report on the work of Jerome Kagan:

> According to Kagan, genes only express under specific epigenetic (environmental) conditions. Genes that produce high blood pressure in mice only express if nursed by their biological mothers and not with unrelated females.[64]

What works for mice presumably works for men as well, though the relevant research has not been done. However, the genetic effects of mental trauma have been established, such as the genetic consequences of the hunger year during the Second World War in Holland.

The reversibility of cause and effect relations between phenomena is a particularly prominent feature of biology because, as Lynn Rothschild notes, "the riches of biology contain a multitude of examples and types of emergence, perhaps greater than found elsewhere in nature."[65] She goes on to say that "causality within biological hierarchies may be bi-directional. In other words, there is both upward and downward causation."[66] She provides many examples of this from diverse fields of biology. Most prominent among these in biology are what she calls "feedback loops":

> In fact, one cannot understand biology without understanding feedback loops. They effectively form the constitutive basis for "top-down" and "bottom-up" causality, as well as providing the dynamism that is empirically evident in biological systems.[67]

Of course, when such biological systems become psycho-physical as well then higher order feedback loops obtain, especially when they work through consciousness in animals and humans. The higher the order of emergence that, so to speak, emerges the more reversible causality becomes. It is emergence that takes logical priority and makes the imputation of causality possible. One can base causality on emergence, but not emergence on causality.

Emergence is a logical not a causal relation, but that does not contradict the simultaneous ascription of all kinds of casual connections where it applies. On the contrary, emergence makes it possible to understand what is meant by

64 Ibid, 141.
65 Lynn J. Rothschild, "The Role of Emergence in Biology", op. cit, 152.
66 Ibid, 156.
67 Ibid, 156.

causality in this context, and also how it is to be explained scientifically. A causal relation can be predicated on the basis of an emergent one when various specified conditions governing occurrences at the level of the emerger and emergent are satisfied. Causality must satisfy the logic of an if-then conditional statement where there is an antecedent called the cause and a consequent called the effect. This happens frequently, though not necessarily always, where there is a constant conjunction involving a temporal disparity between two events, such that the earlier event becomes the cause and the later event becomes the effect. But this is not the only way of ascribing a cause-effect relationship; all kinds of other criteria such as circular feedback loops can also be invoked which do not rely on any constant conjunction or temporal disparity or even succession. Where we can identify an initiating agent or Subject or even an enabling condition then that is assigned the role of cause and what is associated with it can be considered the effect. This is most often the case in situations involving mental causation, as in willed action. These different ways of ascribing cause-effect relations can contradict each other, which is what produces philosophical quandaries, especially when dealing with willed action and the will in general.

In respect of willed action, scientific findings seem to totally contradict our practical everyday sense that as agents and Subjects we are the causes of our own freely willed acts across a whole range of activities that constitute our notion of self-determination both in choosing and acting. Such feelings of initiation seem to back up long-established philosophical and theological doctrines about the freedom of the will. However, of late there have been all kinds of scientific findings that seem to go completely counter to any such philosophical doctrines and seem to demonstrate the opposite conclusion that all our actions are completely determined. In this context, the results of the psychological experiments of Benjamin Libet have provoked considerable consternation among philosophers and the informed public as well, in so far as the media have made people aware of the resultant controversies. For all those intent on the metaphysical version of free will, Libet's results are a strange contradiction, one they struggle to resolve, usually without much success.

There has already been extensive discussion of this in the psychological and philosophical literature, so no detailed description need be given of the protocols of Libet's experimental set-up nor of his findings which are neither original nor remarkable.[68] In brief, these experiments show that when a freely willed and spontaneous choice is made or an intention is consciously registered

68 For original sources and the controversies surrounding Libet's experiments, see Margaret Boden, *Mind as Machine*, vol. II, op. cit, 1223.

it is preceded by up to a second by the brain processes from which it emerges in the mind. It is as if the brain has already decided before the mind can make up its mind, so to speak, or become aware of what that decision has been. This strongly prompts the view that the mind's decision is only an after-thought on what has already taken place in the brain. Hence, the feeling of a freely willed conscious choice is only an illusion and there is no such thing as free will for all willing is determined by processes in the brain.

We can certainly grant that such experimental results demonstrate that there is no direct and simple congruence either in structure or timing between the emergent phenomena of mind, such as acts of will, and the emerger from which they arise, the simultaneous processes in the brain. The two are not necessarily strictly isomorphic with each other, such as Leibniz assumed on the à priori grounds of a pre-established harmony like that of two clocks set to the same time, or Fechner postulated as a psycho-physical parallelism; Koehler was also of this view when he set about late in his life in America to find electrical circuits in the brain to explain the Gestalt laws of closure. Libet's results go counter to all such assumptions and show that the relation between mind and brain is much more indirect and complex.

Libet's results must be interpreted in relation to the emergent nature of mental phenomena in general and in particular to the emergence of acts of will. But that need not be taken to mean that they are caused and therefore not free, provided free will is understood in Spinoza's sense, as we explained in Chapter 4 and will now proceed to further elaborate. As we showed previously, free will in this sense is one of the essential cardinal qualities of the human mind without which the others, namely, rationality, self-consciousness and Ego, could not exist. They are all equally real, so that free will can no more be dismissed as illusory than self-consciousness or the Ego, as some philosophers and psychologists have sought to do. Hence, Libet's findings do not contradict this sense of free will, but they do generate complications for metaphysical conceptions of free will and for any theory of synchrony between acts of will and the neural processes from which these emerge, as we will proceed to show after discussing Spinoza first.

Spinoza's conception of free will is not a metaphysical one, like that of Descartes and most other philosophers; it does not require acts of will that are uncaused causes of bodily motions. Such mental acts do not exist, for as Spinoza insists "the forces of the Body cannot in any way be determined by those of the Mind."[69] As for those who believe in free will in the philosophical-theological sense, Spinoza has this to say:

69 Spinoza, *Ethics*, Preface to Part 5, op. cit, 597.

> So it follows that when men say that this or that action of the Body arises from the Mind, which has dominion over the body, they do not know what they are saying, and they do nothing but confess, in fine-sounding words, that they are ignorant of the true cause of the action, and that they do not wonder at it.[70]

Free will in Spinoza's sense means something quite other than uncaused causes of bodily motions, It means act performed by human agents against compulsions, either exterior or interior or both, as set out in Chapter 4.

An act of will that is deemed free is not one that is free of causal preconditions; on the contrary, if it were that, it could hardly be considered an act at all. It must have its causal grounds at least in the character of the person, the agent of the act, and therefore in all the past circumstances that have produced a person that has the strength of will able to resist and act against compulsions. Character formation is the main causal precondition of the capacity for exercising free will. In fact, if there were no such causal preconditions for the freely performed act consistent with the character of the doer, it would not be a freely willed act at all. It might be a kind of *acte gratuite*, such as Gide and other French novelists have portrayed, a purely arbitrary gesture, or an aberration of the will which would have to be explained by reference to unconscious causes. If that attempt failed, then it would have to be considered an inexplicable occurrence that did not flow from the will of a person, and might not belong to the Ego at all. If a human being began to display too many such inexplicable occurrences, this would indicate a break-down or mental disorder; or, even worse, a complete loss of humanity, as with certain severe Alzheimer sufferers in their last stages.

In the light of such a Spinozist account of free will, the findings of Libet do not constitute a challenge to our idea of free will. All his findings show, as Christof Koch puts it, is that "the brain starts to act before the conscious mind decides."[71] But as the brain, or really the neurons in the brain, does not act or start to act, Koch's statement must be rephrased in another way. We might begin by asking, who or what starts to act? Since, *ex hypothesis*, the conscious mind does not register this starting phase of the act, it must be that the unconscious mind is responsible for it. Hence, it is the unconscious mind that starts to act before the conscious mind decides. We must, therefore, refer the starting

70 Ibid, 495.

71 Christof Koch, *Introduction*, in *Downward Causation and the Neurobiology of Free Will*, eds. Nancey Murphy, George F.R. Ellis and Timothy O'Connor (Springer, 2009), 46.

point of the act to a prior unconscious impulse or volition occurring in anticipation of the act itself.

Such an idea is neither odd nor unanticipated in the light of what we have previously revealed in Chapter 4 concerning the interplay between the conscious and unconscious mind. Edelman and Tononi assert that "unconscious aspects of mental activity, such as motor and cognitive routines, and so-called unconscious memories, intentions, and expectations play a fundamental role in shaping and directing our conscious experience."[72] This holds both for action and perception, as Edelman and Tononi indicate by speaking of motor and cognitive routines. But such two-way relations between the conscious and the unconscious take the form of complex interactions, as they go on to indicate:

> However, conscious experience does not just float freely above an ocean of functionally insulated, unconscious processes. Instead, it is constantly influencing and being influenced by many unconscious processes. Indeed, there are thousands of examples in both perception and action, thought and emotion, that demonstrate that conscious and unconscious processes are regularly in touch and that their separation is far from clear-cut.[73]

This is also the case at the very incipient start of such conscious-unconscious routines, especially at the very earliest beginnings in the mind of what eventuate as conscious actions and perceptions. Both have a momentary unconscious start before consciousness kicks in, as it were, and the acts are registered.

In the case of perception, this has been demonstrated by all manner of experimental evidence that what we consciously view has a very brief unconscious preview. It has been amply demonstrated that unconscious perception, which is not consciously registered, occurs all the time. We become aware of this fact from the cocktail party phenomenon. Subliminal perception is also a well-attested phenomenon which involves the split-second projection of an image which is not consciously seen but whose unconscious effects manifest in behaviour. It seems very probable that there is such an unconscious beginning even to conscious perceptions: a virtual glimpse of a whole scene before one's attention lights on the salient aspects that matter, those which are meaningful and noteworthy, which are then registered as the real perception. According to the New Look psychologists Bruner and Postman, this is how our

72 Gerald M. Edelman and Giulio Tononi, *A Universe of Consciousness,* op. cit, 176.
73 Ibid, 177.

senses operate as selecting, filtering and censoring devices. Hence, the initial start of the act of perception is such an unconscious initial apprehension which precedes very momentarily the conscious sequel.

Something like this interaction between the unconscious and conscious mind operates also in freely willed activities. Just as in perception, there is an initial very brief unconscious impulse or volition before the conscious act begins. To rephrase Koch's formulation: the unconscious mind starts to act before the conscious mind decides. The neural correlate of the unconscious impulse prior to volition – discovered before Libet and known as the "expectancy wave" or *Bereitschaftspotential* – "takes the form of a surface-negative wave that peaks just before the action is performed and discharges just as the act is initiated," as Donald explains.[74] In a situation "when the subject freely decides to move, a surface-negative activation wave begins well over a second before muscle contraction is initiated…"; after which it takes "much less than 100 milliseconds to send a motor command from the cortex to the hand muscles."[75] Donald goes on to offer a sustained criticism of the usual interpretation placed on these phenomena and shows that the problem only arises from a Dualist reading of these findings.

Donald's account is fully in keeping with Spinoza's philosophy of free will, for, as Yakira maintains, "instead of trying to circumvent necessitarian determinism, he poses it as the very ground of his answer."[76] Yakira argues that, unlike Leibniz or Kant, Spinoza is not a compatibilist, someone who considers both necessity and free will as real but compatible with each other and not as contradictory. Spinoza does grant that the classical philosophical-theological conception of free will is incompatible with a necessitarian determinism, but goes on to argue that such a notion of free will is an impossible philosophical chimera of thought. As Yakira puts it:

> Reason shows that free choice [in the metaphysical sense] is an illusion. But freedom is not. Freedom is true. Freedom, however, has no sense and is an empty word, *unless* necessitarian determinism is true. There is no need to show how freedom and necessity are compatible – let alone give up on them – because they are one.[77]

74 Merlin Donald, "Consciousness and the Freedom to Act", in Roy F. Baumeister, Alfred R. Mele and Kathleen D. Vohs, *Free Will and Consciousness*, op. cit, 10.

75 Ibid, 11.

76 Elhanan Yakira, *Spinoza and the Case for Philosophy*, op. cit, 241.

77 Ibid, 242.

As Yakira notes and as we have earlier remarked, Spinoza develops his necessitarian argument in opposition to Descartes' traditional philosophical concept of free will as a kind of causeless cause. The act of will arises without causal antecedents in the mind or consciousness and then it acts as a mental cause on the pineal gland which activates a causal chain ending with the motion of a limb. This complex of events constitutes a freely willed act. There are many contemporary renditions of the same general Cartesian picture but with more up-to-data neurological science; somehow or other, and this is never properly explained, the act of will effects the firing of neurons in the brain which then convey electro-chemical impulses to activate the limb. Such is the model entertained in the famous Eccles-Popper theory of the mind.[78] But this Cartesian model must be rejected on both grounds: the will is not causeless, and it does not act as a cause on the brain.

It might seem as if a Cartesian Materialist approach avoids all these problems that Libet's research results create for Cartesian Dualism. Suppose we eliminate any mental acts of will as initiating causes of action; indeed, we might even remove consciousness altogether from consideration. It seems that thereby the Libet experimental problem is solved; but only at the cost of making the initial experimental result almost incomprehensible, for what can it mean on a purely materialist reading to say that the brain begins an act before the conscious subject is aware of it if there is no conscious subject? Suppose there is some reductive way to remove all reference to a conscious subject and replace it by an exclusive reference to what takes place in the brain alone. Hence, it seems that all that Libet-style experimental results show is that some events in the brain, those that Libet recorded, precede others, those supposedly reductively equivalent to the conscious subject. Problem solved? Indeed, the immediate problem of Libet's research is by-passed, but that only opens up all the more difficult to resolve problems of Cartesian Materialism.

78 J.C. Eccles and Karl Popper, *The Self and the Brain*, op. cit.

CHAPTER 7

Cartesian Materialism and Its Critics

1 The MIT-Mind

Cartesian Materialism was a movement which germinated in the seventeenth century in the then opposed philosophies of Descartes and Hobbes. But gradually over the centuries the two approaches grew closer and closer together until in the second half of the twentieth century they coalesced into a combined philosophical, scientific and technological project to mechanize the mind. Since then it has appeared under the general rubric of "mind as machine". Its success marks the furthest possible realization of Descartes' dream to mathematicise the world. This it achieved by extending Descartes conception of matter as extension to mind as well and treating mind, too, in a purely mechanistic manner.

The crucial step in the direction of Cartesian Materialism was carried out in the eighteenth century by la Mettrie, who took Descartes' pronouncement that the body is a machine and extrapolated it to the mind as well. The precedent for that move was already set by Hobbes a century earlier. Hobbes had argued that thinking is mere "reckoning" or computation as we would now say.[1] Furthermore, he contended that this takes place as motions in the body since all mental activities, be it sensing or speaking or thinking, are "but so many motions of the matter".[2] These were all early anticipations which only came to their culminating fruition in the second half of the twentieth century as a result of a complex series of historical developments at once in formal logic, in technology, in science in general and finally in philosophy as well. All these developments were initiated under the shadow of war during this cataclysmic period in world affairs.

The story of the project for the mechanization of mind has already been told by many authors from different points of view, among the earliest were Howard Gardner, Steve Joshua Heims and Jean-Pierre Dupuy, but the most comprehensive account by far is the already frequently referred to work of

1 Thomas Hobbes, *Leviathan; or, the Matter, Forme of a Commonwealth Ecclesiastical and Civil*, ed. A.D. Lindsay (London: Dent, 1957), 18.

2 Ibid, 23.

Margaret Boden.[3] What is clear from all these accounts is that the idea of the mind as machine is not just a thesis dreamt up by philosophers but a historical process that is at once political and technical, scientific and cultural, and that the philosophical ideas it generated ride on the back of powerful forces without which they would have been nothing but dreams. Philosophical ideas do not live in a social and cultural vacuum, they cannot arise or survive if the conditions necessary for them to be incubated and hatched do not exist. This is all the more so the case with the "mind as machine" idea which arose in the context of scientific researches and technological inventions designed for war, the Second World War and the Cold War successively.

The technologies so crucial for the project to mechanize the mind were largely invented in three outstanding war-time sites: Bletchley Park in England, The Radiation Laboratory at the Massachusetts Institute of Technology or MIT, and Los Alamos in New Mexico. Of these, the so-called Rad Lab has been the least noted but perhaps the most important for later philosophies of mind. Bletchley Park is well known for its work on encryption associated with Alan Turing and the Bombe, a precursor of the computer; and Turing's earlier work in formal logic and the Turing Machine and later in formulating the Turing Test was crucial in this intellectual battle for the mind. Los Alamos, though more famous for the Bomb, is also closely associated with computing and John von Neumann, another mathematical logician, who set out the basic principles of the theory and application of the newly invented electronic computer. However, the contribution of Rad Lab, though more difficult to trace and still not fully documented, was perhaps even more decisive for it helped to form so many of the future developments at MIT, and through the person of Norbert Wiener, the originator of Cybernetics, it had an even more powerful influence on the "mind as machine" idea. Turing, von Neumann and Wiener were the key characters in this play of ideas and inventions but there were numerous others involved in the drama that culminated with the new *deus ex machina* or rather the machine that became a god.

Subsequent to the invention of the computer, which took place elsewhere, most of the developments based on it occurred at MIT. This was brought about by the three shapers and shakers of American science policy during and after the war, Vannevar Bush, Jerome Wiesner and Joseph Licklider, all from MIT. Bush, Roosevelt's chief scientific adviser, who oversaw both the Rad Lab and Los Alamos, was himself an engineer, the inventor of an analogue differential

3 Howard Gardner, *The Mind's New Science: A History of the Cognitive Revolution*, op. cit.; Steve Joshua Heims, *The Cybernetic Group*, op. cit.; Jean-Pierre Dupuy, *The Mechanization of the Mind,* op. cit.; Margaret Boden, *Mind as Machine*, op. cit.

analyser; he had a clear insight into future information technologies and conceived the idea of cyberspace. Wiesner, president of MIT from 1973–80, was head of the Research Laboratory of Electronics where the first Artificial Intelligence project was instituted. The RLE also hosted all the seminal disciplines that were to play such a crucial role in the mechanization of the mind, as John McCarthy, one of the early pioneers reports:

> At one stage we had twelve different fields represented in the laboratory. The Linguistic and Psychology departments grew out of groups that were started in the lab. [The RLE's communications engineers were joined by] neurophysiologists and other biologists, linguists, economists, social scientists, and psychologists of various persuasions... They explored each other's fields and slowly began to comprehend each other's lingo... The two decades of RLE were like an explosion of knowledge.[4]

It is ironic that all these disciplines were galvanised into new life by the currents flowing from electrical engineering.

What really made this "explosion of knowledge" possible was the almost unlimited funding at MIT provided by Joseph Licklider, the chief of the crucial patrons. Licklider began as a psychologist at MIT, but under the influence of war-time research he soon branched out into communications engineering; he later sponsored the development that led from the early ARPAnet to the Internet. As the disburser of funds from the Advanced Research Programs Agency (ARPA), an agency of the Department of Defence, of which he was the first director, Licklider made sure that anything of the least military relevance was well supplied and well staffed. He thus furthered the careers of nearly all the prominent names associated with the mechanization of mind.

The Rad Lab, perhaps the biggest of all the war-time projects in terms of scientific man-power, was disbanded at the end of 1945, but it had served to give MIT a head start in the post-war staffing of technical universities with people who took the lead in the mechanization of the mind. Every new endeavour in that project was initiated at MIT, by people who were already there, by students who trained there, or by those who came there from other places. From MIT analogous establishments branched out to other institutions, mainly to technical universities or technical departments, in the first place to Carnegie Mellon University, Stanford Research Institute and Stanford University. Bitter wars were fought in academic politics for a portion of the funding dispersed by ARPA, but the lion's share invariably went to MIT. On victory or

4 Quoted in Margaret Boden, *Mind as Machine*, vol. II, 733.

defeat in these battles for money and the prestige that went with it depended whose projects and ideas would succeed and which would be suppressed. The wrangle between Marvin Minsky at MIT and Frank Rosenblatt at Cornell was one of the most prominent examples of these contests; it was eventually settled by Licklider in favour of Minsky, and thereby Connectionism was left to languish for nearly three decades. Never before in the history of human endeavour have ideas been so driven by money and power, in the electrical, political and academic senses at once.

Both in science and philosophy, the impact of developments at MIT was so predominant that just as we speak of Cartesian mind so we might call the model of the mind that developed in these post-war years the MIT-Mind (this can be pronounced to rhyme with Mighty Mouse, the other great project where the mountain laboured to give birth to a mouse). It was the outcome of the collective efforts of literally hundreds of engineers, scores of scientists and dozens of philosophers not all of them at MIT, but all sharing the basic mechanistic approach developed there. This approach became paramount in many departments, such as computer science, artificial intelligence, robotics, linguistics, psychology and philosophy; from these it spread to many others as well. Eventually it became diffused throughout the American academic system. From America it was transported first to Britain and then to some extent to the rest of the world. It became the dominant intellectual current from then on and it is still nearly all-pervasive.

The history of how all this happened is a long story that only a large work such as Boden's two volume book can convey in detail. But to provide the gist of what took place we might briefly sketch out four major periods of twenty years each: 1936–1955, 1956–1975, 1976–1995, 1996–2015. Obviously, this is a purely arbitrary periodization, but it is heuristically useful in providing a rough temporal framework into which, as in a Procrustean bed, we can truncate ongoing continuous developments.

We choose 1936 as the starting date because that was when Turing published his celebrated paper "On Computable Numbers with an Application to the *Entscheidungsproblem*" in which he described his so-called Turing machine, an abstract automaton which proved that anything computable can be computed by a Turing machine.[5] This has also been frequently and mistakenly taken to "prove" that anything the mind can do can be done by a Turing machine; from which the further conclusion is drawn that the mind is in fact a Turing machine, namely, a computer. It is on the basis of this "proof" that Turing later

5 Alan Turing, "On Computable Numbers with an Application to the *Entscheidungsproblem*", in *Proceedings of the London Mathematical Society*, vol. 2, 42/3 (Nov. 1936).

devised his so-called Turing test which seemed to show that there is no way of differentiating between a mind and a computer. This apparently logical "proof" became the basis and justification for all "mind as machine" theses ever since. But as usual with such "in principle" demonstrations, it means nothing in practice, in the reality of actual minds and machines, as we shall show at the end of this chapter.

However, an even more direct and decisive starting point for the MIT-Mind came in the middle of the war, in 1943, when Warren McCulloch and Walter Pitts published their paper "A logical Calculus of the Ideas Immanent in Nervous Activity".[6] They were then at the University of Chicago; but in 1952 they moved to MIT, to the Electronics Laboratory, as might be expected. According to Boden, McCulloch "was a prime mover in founding both streams of mind-as-machine research: formal computational on the one hand, and cybernetic-probabilistic on the other."[7] At MIT, he joined forces with Wiener who was working on teleological machines. As Dupuy remarks, "the influence of McCulloch, who theorized the embodiment of mind in the machine, was no less than that of Wiener, who conceived of organism and machine as models of each other."[8] The paper also attracted the attention of von Neumann, then the leading mathematical logician, and also a prime exponent of the mechanization of the mind. As Norman McRae von Neumann's biographer notes, "the McCulloch-Pitts model was 'immensely' naïve, but Johnny was 'enormously' impressed with it."[9]

Von Neumann was then engaged on building the first properly functioning electronic computer, EDVAC, and due to his favourable impression of the McCulloch-Pitts model, he described it in language that was directly borrowed from such a mechanical account of mind. He thereby introduced the systematic confusion that has dogged us ever since of speaking of computers as if they were minds and of minds as if they were computers. What brought this about was the idea of the computer as a mechanical "brain" and of the brain as a logic-machine in accordance with the McCulloch-Pitts model. Thus in his March 1945 "First Draft of a Report on the EDVAC" von Neumann writes as follows (using the terms CA for the central arithmetical part of the computer, CC for the central control and M for memory):

6 Warren McCulloch and Walter Pitts, "A logical Calculus of the Ideas Immanent in Nervous Activity", *Bulletin of Mathematic Biophysics 5*, (1943).

7 Margaret Boden, *Mind as Machine,* op. cit, 183.

8 Jean Pierre Dupuy, *The Mechanization of Mind,* op. cit, 79.

9 Norman McCrae, *John von Neumann, the scientific genius who pioneered the modern computer, game theory, nuclear deterrence and much more* (New York: Pantheon Books, 1992), 283.

> The three specific parts, CA, CC (together C), and M correspond to the associative neurons in the human nervous system. It remains to discuss the equivalents of the sensory or afferent and the motor or efferent neurons. These are the input and output organs of the device…[10]

The propensity to this kind of error was, of course, prompted by the fact that "McCulloch and Pitts, as well as von Neumann and Wiener, were deeply committed to Kenneth Craig's mechanistic ideology…", as Heims explains.[11] He goes on at length to show what was behind their thinking that made the equivalence of computers and brains more or less unavoidable, to the subsequent almost inevitable perpetuation of this confusion:

> Warren McCulloch and Walter Pitts were devout adherents to the mechanistic, more specifically electronic, preoccupations in describing organisms. Their most important achievement had been to show that a central nervous system composed of a network of simple neurons, each having the formal property of an electrical switch, could have "memory", could "learn", could "recognize forms", and more. Since neurons are discrete elements with only two possible states, the central nervous system was seen as resembling a digital computer. As Macy participant Heinz von Foerster belatedly point out, the anthropomorphic concepts such as memory and learning reflect semantic confusion when applied to machines. Thus instead of memory, one could more legitimately speak of recording device, and so on, but the anthropomorphic language was not challenged at the conferences.[12]

The Macy conferences Heims is referring to were the series of meetings by invitation only, which took place in the immediate post-war decade and were attended by the foremost exponents of the mechanization of mind movement.

The Macy conferences brought together all the leading participants, drawn from a variety of disciplines, who subscribed to the mind as machine movement, including the key exponents of this mechanistic ideology, Wiener, von Neumann, McCulloch and Pitts. It was Wiener who coined the name of the first instantiation of the movement, Cybernetics. Another crucial participant was Claude Shannon, the principal founder of Information Theory, which he

10 Quoted in ibid, 286.

11 Steve Joshua Heims, *The Cybernetic Group*, op. cit, 204.

12 Ibid, 204.

began as a Master of Science thesis at MIT and later in 1949 published as a book together with Warren Weaver. This, too, was a key contributor to the confusion in likening minds to machines. "Information" in Shannon's technical communication sense has nothing to do with information in any meaningful sense, it is not information *about* anything, namely, it is not an intentional concept but merely a technical measure.

Furthermore, information in the technical sense has nothing to do with meaning; it is, in fact, the converse of meaning, for the more meaningless the message, the more "information" it contains. It is to his credit that Shannon pointed this out repeatedly to the other Macy participants, but they took no notice for the ambiguity of "information" suited their "mind as machine" predilections. The same is true for the other key term of the theory, "message", which has nothing to do with real meaningful communication where the failure to receive a message can be as significant as the reception of one, as the adage no news is good news declares.

Shannon had developed his Information Theory while working at Bell Labs purely to solve a technical problem in electronic transmission. It had solely to do with the conveyance of electronic signals down lines, and as Shannon and Weaver stressed, the "semantic aspects of communication are irrelevant to the information problem."[13] Information in this technical sense is a quantitative measure which could not be converted into anything to do with information as conveyed by language and contained in discourse. But given a telephone exchange model of the brain, such as entertained by McCulloch and many others, it was almost impossible to prevent them using "information" and "message" in both the equivocal senses at once. Hence expressions such as "information overload" and "information processing" soon became common parlance among the Macy conference participants and it was taken up by mechanically minded psychiatrists and psychologists, starting with Gregory Bateson. He also applied to the mind Wiener's Cybernetic concepts of "circular causality" and "feedback" expressed in terms of "information" and "messages".[14]

Wiener developed both the theory and concepts of Cybernetics in the course of his war-time work at the Rad Lab at MIT on automatic anti-aircraft aiming guns, which he conceived of as a problem in mechanistic teleology. He then went on after the war to combine this idea with Shannon's Information Theory of "message" and "code", both of which he interpreted in terms of

13 Claude Shannon and Warren Weaver, *The Mathematical Theory of Communication*, (Urbana: University of Illinois Press, 1949), 3.

14 See Gregory Bateson, *Steps to an Ecology of Mind, A revolutionary Approach to Man's Understanding of Himself*, (New York: Ballantine Books, 1972).

thermodynamic physics to come up with the grand theory of negative entropy as a general measure of order in all the sciences, especially in biology. As he puts this in his own words:

> In fact, it is not surprising that entropy and information are negatives of one another. Information measures order and entropy measures disorder. It is indeed possible to conceive of all order in terms of message.[15]

Taken in this way, "information" identified with negative entropy seemed to provide a way of accounting for order and form in all natural systems, particularly in biology. This idea became the source for fanciful speculations in nearly all disciplines as "communication", "information" and the exchange of "messages" proliferated in the vocabularies not only of the natural scientists but of the social scientists and humanities exponents as well. Those who opposed this scientistic trend accused Wiener of having sold his human birthright for a pot of message. But Wiener was undeterred.

Wiener himself continued this process by referring his idea of "information as a measure of order" to a general study of all organized systems, whether organic or cerebral or social. It was all meat for his omnivorous scientific appetite, for he devoured it all indifferently. Thus, as he stated:

> The most important characteristic of a living organism is its awareness of the outer world. This means that it is furnished with organs of couplings to pick up messages from the outer world that condition its future conduct. To consider this in the light of thermodynamics and statistical mechanics is instructive.[16]

It is also highly confusing. For, taken in this way, living beings, even those with minds, are identified with mechanical automata in a move, which, though it is scientifically much more sophisticated, is in direct continuity with la Mettrie and Cartesian Materialism in general.

In a formal sense the Cybernetic phase of the movement lasted for as long as the Macy conferences continued till the mid-1950s. The transcriptions of the debates at these meetings have since been published as an extensive series and numerous books have appeared about the movement as a whole. When the official Macy conferences were over, the movement went, so to speak, underground; later it was partially revived as a so-called Second Cybernetic by one of

15 Quoted in Jean-Pierre Dupuy, *The Mechanization of Mind,* op. cit, 114.
16 Ibid, 114.

its youngest and latest participants, Heinz von Foerster. What brought it to a close was the popularity among the younger cohort of the succeeding generation for a new departure called Cognitivism and Artificial Intelligence. Even though the proponents of that movement disclaimed and disdained their paternity, they could not have come into being without the Cyberneticist fathers and their leading ideas which they embraced implicitly without openly admitting it, as Dupuy's book reveals.

This new phase began with a bang in 1956 at the start of our second period, another fortuitous date which came to be called the *annus mirabilis* of Cognitive science. It was during this year that two crucial conferences occurred that launched this next movement. In the summer of that year, at the Dartmouth Summer School Research Project, largely organized by John McCarthy together with Minsky and Shannon, all at the time associated with MIT, Artificial Intelligence was founded. Later that year an even more momentous meeting took place at MIT itself which was decisive for linguistics, psychology and many others of the so-called Cognitive sciences. It was when Chomsky first made his mark and established his working relationship with George Miller, then at Harvard; but he had spent his war-years at MIT, and at that time was far more attuned to developments at MIT rather than Harvard, as his colleague Bruner attests.

It began with Chomsky's onslaught on B.F. Skinner at Harvard and his futile attempt to provide a Behaviouristic explanation of language. But very soon this was generalized into an attack on Behaviourism in psychology and elsewhere where it had been applied to human behaviour. But the difference between Behaviourism and the Cognitivism that displaced it is not as large as the exponents of the latter make out. Both rely on mechanistic models, which relate stimulus to response, the Behaviourists by means of hidden variables, and the Cognitivists by means of rules or representations that govern the process whereby input is transformed into output. Edward Reed sees "cognitivism as a set of epicycles on S-R theory", and he goes on to make the critical point that key terms like "information" and "representation" remain undefined in Cognitivist psychology:

> Modern writers confess that they have no account of what information is (e.g. Palm and Kimchi, 1985) or insist that, whatever it is, it cannot be information in the cognitive sense of referring to things in the world (Jackendoff, 1992)… It also might be argued that the concept of representation serves to move cognitivism beyond the framework of S-R psychology. But this concept has remained essentially undefined for so long that the only useful response to such a claim is to ask for a detailed account

> of "representation" that both fits what cognitive psychologists do and offers more than a gloss on the phrase "representations are the contents of cognitive states".[17]

This is, of course, criticism that dates long after the so-called Cognitivist Revolution was inaugurated in the heady year of 1956 when no such problems were foreseen in what appeared to be a new multidisciplinary paradigm in the offing.

Chomsky was the driving force and later became the *eminence gris* behind the Cognitivist Revolution in linguistics and psychology and later still his influence was prominent in many other sciences, as well as philosophy. He attained his sway in large part because it was believed that he had slain the dragon of Behaviourism that had been terrorizing American psychology for a generation or more. In linguistics itself, his own native discipline, he did not attack the reigning Bloomfieldian authorities head on, but, like MacArthur in the Pacific war, he adopted the indirect outflanking approach of simply ignoring them and leaving them to wither on the vine. In any case he was not professionally equipped to mount a full frontal assault on the established linguistics paradigm and critically engage with its concepts and methods. He was, in fact, poorly equipped for any comparative study of languages, the forte of the orthodox approach, since he commanded very few. As he himself latter admitted, he "could not possibly have obtained a position in linguistics anywhere – I really was not professionally qualified by the standards of the field."[18] But at MIT, where he was employed in the Research Laboratory of Electronics (RLE) on the project of machine translation, his incompetence did not matter because there was no linguistics department and no tradition of language teaching. Machine translation was a forlorn endeavour, despite the extensive funding it received, as it soon became evident. Its use is even now extremely limited since it is so unreliable. At the time, it arose largely out of Cold War exigencies for the translation of Russian scientific papers, especially in the post-Sputnik era, and the shortage of competent Russian translators. It reveals clearly the *fons et origo* of Chomsky's linguistics, from where comes its emphasis on formalized grammar or syntax at the expense of meaning or semantics. The fact that it soon proved a failure, did not prevent Chomsky developing his formalistic linguistics at MIT

17 Edward Reed, "The Cognitive Revolution from an Ecological Point of View", in D.M. Johnson, and C.E. Erneling, eds. *The Future of the Cognitive Revolution*, op. cit, 266.

18 Noam Chomsky, *Language and Responsibility*, based on conversations with Mitsou Ronat, trans. John Viertel (New York: Pantheon Books, 1979), 133.

free from censure or outside challenges from established linguists, as he himself attests:

> ...we were able to develop our program at MIT because, in a sense, MIT was outside the American university system. There were no large departments of humanities or the related social sciences at MIT. Consequently, we could build up a linguistics department without coming up against problems of rivalry and academic bureaucracy. This permitted us to develop a program very different from any other and quite independent.[19]

However, from that position of splendid isolation, Chomskians went out to conquer the academic world, first linguistics and psychology departments and then many others as well. Whether this has been to the ultimate benefit of these disciplines or the humanities and social sciences in general is to be doubted.

It is in line with this kind of formalistic and mechanistic bias that Chomsky approaches the study of language, a social and humanistic matter if ever there was one. However, for Chomsky that is not what counts, for his version of linguistics is also highly mentalistic and Cartesian, as Christina Erneling brings out:

> He insists that we need to understand language and other mental functions from an individual and internal point of view. By contrast, details of actual public use of language and its socio-communicative aspects fall outside scientific investigation. This individualistic and internal (one could also add intentional and innate) language Chomsky calls I-language to separate it from E-language – that is, the external and social use of language... Instead of discussing semantics and questions having to do with communication and normative aspects of language, he thinks that he should focus on the language system in the individual mind/brain. Chomsky thus clearly proposes to limit the study of language to a narrow range – that is, giving an account of the individual grammatical competence of an ideal speaker or listener.[20]

This Cognitivist internal and mentalistic aspect of language materialized in the hands of Chomsky as a gadget in the brain, which he called a Language Acquisition Device (LAD). Namely, he postulated that every human child had

19 Ibid, 134.

20 Christina Erneling, "Cognitive Science and the Study of Languages" in D.M Johnson and C.E. Erneling, *The Future of the Cognitive Revolution*, op. cit, 115.

an inborn and genetically inherited grammatical competence built into it as a module in the brain. This was the material manifestation of an innate competence in exercising what he called Universal Grammar, an abstract grammar that is shared by all the grammars of all existing and all possible languages. Its mechanistic and formalistic derivation is explained by Howard Gardner as follows:

> His view of grammatical generation was based on an automaton – a machine in an abstract sense which simply generates linguistic strings on the basis of rules that have been built (programmed) into it. The resulting grammar is neutral – equally valid as a description of linguistic production or linguistic comprehension. Clearly, Chomsky was a *child* of the new era of Wiener, von Neumann, Turing and Shannon, though – just as clearly – some of his specific ideas about how language works ran directly counter to information theory notions.[21]

Chomsky never really acknowledged his debt to his Cybernetics precursors, but his work and ideas would have been impossible without them, as Dupuy's book brings out.

Chomsky more than any other is responsible for constituting the MIT-Mind for it was he who devised the idea that the mind can be at least partly broken up into self-contained and hermetically sealed-off modules, of which his LAD was the initial example. This was an idea eagerly taken up by his students and acolytes at MIT, who ran with it as far as it could be taken. Modularity of mind, that is, modules of perception, of social processes, of innate knowledges, and of much else besides, was preached by Jerry Fodor, Jerrold Katz, Zenon Pylyshyn and Steven Pinker. Once again we will rely on Gardner to explain why this was such a crucial move for generating the MIT-Mind:

> Chomsky gradually challenged the widespread belief in extremely general and broad powers of the mind – powers like learning, stimulus generalization, and the like… He came to think of the mind as a series of relatively independent mental modules, each with its own rules of operation and development, along with prescribed means of interacting with other "organs". Here was a powerful statement on behalf of the autonomy both of language as an organ – and, not coincidentally, of linguistics as a discipline.[22]

21 Howard Gardner, *The Mind's New Science*, op. cit, 188.
22 Ibid, 188.

Of course, the linguistics that is so independent is Chomskian linguistics which detaches itself by this means from every other approach to language and from every other discipline that has any bearing on language. Hence, over his own newly defined linguistics Chomsky could reign supreme and unchallenged. Under Chomsky's inspiration and instigation many others – who also founded purportedly "new" disciplines that came to make up the MIT-Mind – adopted a similar strategy, each claiming to be solely governed by its own module in the brain that was innate. Thus modules of all conceivable kinds proliferated.

Minsky's contribution to the MIT-Mind was not as crucial as Chomsky's but it was also important. For most of his career Minsky specialized in Artificial Intelligence in its so-called GOFAI (Good Old Fashioned AI) form, that is, he abjured all forms of the alternate computational approach called Connectionism, though later he did come around to a version of it, called Parallel Distributed Networks. During the early part of the second phase of the Cognitivist Revolution 1956–1975, Connectionism was advocated by many outside MIT, such as Frank Rosenblatt at Cornell, who became Minsky's main rival and bête noir. Minsky, together with his associate Seymour Papert, launched a no-holds-barred frontal assault on Rosenblatt and the Connectionist coterie and all their robotic works. It was a battle that Minsky was bound to win, for apart from any scientific or technical advantages that GOFAI might have had over PDP or any other form of Connectionism, Minsky had the ear of Licklider, the main disburser of funds from ARPA, and, incidentally, also an original participant in the Macy conferences. Thus, by such means was Connectionism starved into submission for most of this period; though it came back with a vengeance during the third period toward the end of the 1980s and became ever more prominent from then on, especially when the US Navy found it was useful in submarine detection devices. On such factors do developments in the science and philosophy of mind depend in our time.

Perhaps not altogether coincidentally, it was also during this period, specifically in 1962, that Thomas Kuhn published his renowned book *The Structure of Scientific Revolutions*.[23] This had a decisive influence in shaping the self-confidence, manner and bearing of his scientific confreres at MIT. They saw themselves as propounding a new paradigm in the Kuhnian sense, just as Newton and Einstein had done in physics and Lavoisier in chemistry and so on throughout the sciences. As originators of a new paradigm, whether of AI or Cognitivism or both combined, they assumed that they could afford to take no notice of any objections to their endeavours, for these came from the exponents

23 Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago, Press, 1962).

of the previous paradigms who would not understand them and only offer carping criticism. Since alternate paradigms were incommensurate and incomprehensible with each other, there was really no need to debate or argue with exponents of earlier paradigms who would inevitably be superseded. There was no need or possibility of rationally justifying oneself since those who held alternative paradigms could never see eye to eye or even see the very same things, for in science these are determined by theories. Hence, rival theorists could not understand each other much less convince each other. It was this Kuhnian idea of science progressing by such incommensurable paradigm revolutions that brought about a particularly antagonistic intolerance that is so evident in the dogmatic way that Chomsky and Minsky and most of their followers at MIT and elsewhere conducted themselves in relation to their scientific or philosophical rivals, or even their own disobedient students, who were frequently treated as enemies. It also produced a characteristic dogmatism in the way they framed their views, which was particularly damaging as the fields to which they sought to make a contribution are little given to scientistic self-assurance and certitude, but must be explored from many different points of view and debated with due regard to differences.

All this became more evident in our third period starting in 1976, the ascendency phase of the Cognitivist Revolution. Unfortunately for our attempt at exact dating, nothing noteworthy happened in 1976 itself, except that a year earlier Chomsky published his *The Logical Structure of Linguistic Theory*, a milestone in formalist linguistics.[24] It must have seemed to him and all the others that Cognitivism was unchallenged and unchallengeable. As Fodor put it in his colloquial epithet "there is no other game in town". This seemed to be confirmed when ten years later in 1985 Gardner's book *The Mind's New Science* appeared at the very apogee of this victorious ascent. His graphic representation of the memorable hexagon of sciences and philosophy and their numerous interconnections made it seem as if an overarching synthesis of the natural and social sciences concerned with matters of mind had been accomplished. It featured neurology, psychology, anthropology, linguistics, artificial intelligence and philosophy. If only diagrams could substitute for real relations, then an amazing composite field had finally been opened up and merely awaited further detailed exploration, or so it seemed at the time.

All this represented a kind of colonialist expansion of the MIT-Mind which now sought to invade all territories dealing with the human mind and subjecting them to its rule. Modules of all kinds multiplied incessantly. Research on newly born babies and infants by Elizabeth Spelke and others at MIT seemed

24 Noam Chomsky, *The Logical Structure of Linguistic Theory* (New York: Plenum Press, 1975).

to reveal a host of modules with which neonates come equipped into the world. Research on autistic children by Simon Baron-Cohen, who worked at MIT, seemed to show what can go wrong when any such modules are damaged or missing; autism rapidly became a world-wide epidemic. The climax of this search for modularity came in 1982 when Jerome Barkow, Leda Cosmides and John Tooby published their edited volume *The Adapted Mind*, and brought prehistoric mankind into the module scheme of things under the heading "massive modularity".[25] Their metaphor for the MIT-Mind was the Swiss all-purpose army knife with separate gadgets for different needs. We shall critically consider what is wrong with this way of dividing up the mind later in the next section.

Artificial Intelligence also seemed to be celebrating one success after another. At MIT in the early 1980's there was established a transdisciplinary Media Lab at the instigation of Jerome Wiesner, a friend of McCulloch, Pitts and Wiener and also a Macy participant. At its head was Minsky with Seymour Papert in support. Funding for this venture came from commercial sources, such as IBM, Apple and Nippon Telephone and Telegraph, as well as the ever-generous ARPA; both the capitalists and the military saw huge prospects for AI in their different lines of business. They were not to be disappointed. But what this did for the science and philosophy of mind is another matter altogether. At first it seemed as if all human activities could be mechanized; hence the drive to develop programs for so-called "expert systems" which seemed to be able to do the work not only of receptionists, but of doctors and lawyers as well. But success was scanty as basic problems soon revealed themselves. Despite the fact that computers can play chess as well as, if not better, than chess masters, they prove remarkably "stupid" or rigidly automaton- like where ordinary common sense matters are at issue. And it soon became apparent that such problems are not mere technical hitches in programming, but insurmountable difficulties that show-up the inherent limitations of computers when attempting to simulate human activities.

Towards the end of this third period in the late 1980s and early 1990s not only technical shortcomings began to emerge with GOFAI, but theoretical ones as well as Connectionism liberated itself once more at the hands of John Haugeland, D.E. Rumelhart and J.L. McClelland. Rosenblatt's pioneering work and that of connectionist stalwarts such as the neurologist Steve Grossberg was vindicated. Most of this work took place at the opposite side of America to MIT

25 Jerome Barkow, Leda Cosmides and John Tooby, eds. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture,* (New York: Oxford University Press, 1982).

on the West coast, at San Diego, UCLA and Cal Tech where so-called Hopfield nets were developed, but who did so first is a hotly disputed issue of priorities.

Connectionist models of the brain certainly make a difference to how the underlying mechanism of the MIT-Mind is modelled and conceived. But, nevertheless, this is still a mechanistic conception, though it is one that is closer to how brains realistically operate, having brain-like features such as the strengthening of connections and the ability to function despite damage. However, a connectionist network is only as Horgan remarks "a new-fangled embodiment of old-fashioned statistical methods, such as curve-interpolation for coping with incomplete or ambiguous data."[26] Hence, a connectionist machine is, nevertheless, still a machine, and more like a toy compared to the brains of even mindless animals. It does not even begin to approximate human brains. And even the most modest connectionist models can only be made to work as simulations on computers, so they cannot really exceed the capacities of real computers, as opposed to imaginary Turing automatons for which anything is possible.

It was also towards the end of the third period, at about the time when the Soviet Union collapsed and the Cold War was over, that the first signs of cracks in the façade of the MIT-Mind began to appear. Even at the very moment of its triumphalist affirmation, some of its former exponents were expressing regrets at what had become of the movement that they had initiated. Miller wrote of it as entering an end-phase of "analytic pathology"; and his close colleague Bruner – they had together founded the highly influential Centre of Cognitive Studies at Harvard – was also beginning to back away from mechanical reason:

> I think I am suspicious of "formal" models of human behaviour – theories couched exclusively in mathematical terms or in abstract "flow diagrams". I have always been sympathetic to the metaphors of computation and information processing, but resistant to getting trapped in their necessary measurement constraints... I did not, in consequence, get deeply into the Harvard-MIT "cognitive sciences" network...[27]

He refers to computation and information processes as *metaphors*, which shows how far he has departed from the original project. Later he was to

26 John Horgan, *The Undiscovered Mind*, op. cit, 208.

27 J.S. Bruner, *In Search of Mind: Essays in Autobiography* (New York: Harper and Row, 1988), 99.

withdraw even further away from his initial involvement with the "cognitive sciences network":

> It would make an absorbing essay in the intellectual history of the last quarter-century to trace what happened to the original impulse of the cognitive revolution, how it became fractionated and technicized… Very early on, for example, emphasis shifted from "meaning" to "information", from the *construction* of meaning to the *processing* of information… The key factor in the shift was the introduction of computation as the ruling metaphor and of computability as a necessary criterion of a good theoretical model.[28]

So much then for his original "sympathy to the metaphors of computation and information processing"! But this was written after his sad experience at Oxford where he had to endure, in his own words, weekly "Bruner bashing" seminars from his Cognitivist colleagues until he could bear it no longer and departed back to America.

At that point it became apparent that as the MIT-Mind went from strength to strength and appropriated one science after another, so, too, in reaction, opposition to it also began to mount. At the very moment of its triumph, those who had been participants in its birth were having second thoughts and even regretting their part in creating this Frankenstein Monster, which threatened to multiply out of control and produce innumerable clones of itself. Thus already in 1983 Bruner saw that what he and his confreres had initiated had escaped their supervision and was proliferating:

> My generation created and nurtured the Cognitive Revolution – a revolution whose limits we still cannot fathom, for it is at the centre of a post-industrial society that is still in roaring growth… And it is probably not an accident that at the very time the Cognitive Revolution was occurring, America's Gross National Product was for the very first time coming to be made up of the earnings of the industries involved in communications and in the high technology of control, computation and management.[29]

In other words, the MIT-Mind was in keeping with the spirit of America in the post-war years of industrial and social expansion which reached its apogee in the mid-1980s when Bruner was writing his autobiography.

---

28 Jerome Bruner, *Acts of Meaning*, (Cambridge, Mass.: Harvard University Press, 1990), 4.

29 Jerome Buner, *In Search of Mind*, op. cit, 274 and 63.

Unfortunately, despite the utility of the MIT-Mind approach for machine technology and industry, its effects on science were far from salutary. Perhaps no science was more disastrously impacted than psychology, where the Cognitive Revolution was initiated: "what happened, of course, was that psychology's world was turned topsy-turvy not by psychology itself but by modern theories of computation, of linguistics, of anthropology, and, indeed, even of biology."[30] George Miller was also beginning to be sceptical about what the revolution had accomplished, he stated that "in my opinion, however, the use of these mentalistic terms [cognition, attention, memory, intuition, etc.] is still constrained by a positivistic philosophy of science, so that now we have in effect an oxymoron non-mentalistic cognitive psychology."[31] Thus criticism was mounting from the inside as well as the outside.

So with this rising tide of opposition, we enter the fourth and final period from 1996 onwards, though that date is again not significant; but it was from then on that attacks on the MIT-Mind and all its computational and information processing accessories came thick and fast. The collected volume *The Future of the Cognitive Revolution* published one year later in 1997, though based on a conference in Toronto held earlier, is a portend of the time ahead. The general verdict among the participants is that the Cognitive Revolution had run its course; though stalwarts of Cognitivism, such as Boden, still hung on to the prospects of its continuing relevance. That was to be the ground-tone of her near exhaustive and exhausting to read two volume study of the whole of the movement from Turing onwards; we have relied on it to produce the more critical account we have just sketched out.

Boden is well aware of most of the countervailing movements; though not of all, or even from our point of view the most salient ones, for she makes no mention of either Donald or Deacon, which indicates that she does not know of them or does not take their views seriously. Hence, she is unlikely to be impressed or disturbed by any expressed in this book either, and that is more or less what is to be expected from long-time adherents of the MIT-Mind. Nevertheless, Boden does take note of opposed views, she refers to at least three such alternative "closely related recent movements: situated robotics, cognitive dynamics and neo-Heideggerian epistemology. Each of these stresses the situated, embodied nature of thought and action, and rejects commitments to representation and computation of GOFAI and connectionism alike."[32]

30 Ibid, 60.

31 Quoted in ibid, 120.

32 Margaret Boden, "Promise and Achievement in Cognitive Science", in D.M. Johnson and C.E. Erneling, *The Future of the Cognitive Revolution*, op. cit, 57.

Situated robotics has more to do with the technology of automata and so is of no concern to us here except as it reveals the shortcomings of GOFAI even in the specialized domain of building better machines. It derives from work undertaken at MIT in the Media Lab led by Rodney Brooks and his co-workers seeking to overcome the purely technological impasse that GOFAI in its standard form has come to. Relying on top-down programming – in which everything that could conceivably arise in a situation to be simulated has to be explicitly allowed for in the program – leads to unwieldly and ultimately impossible to realize procedures since exceptions are literally limitless. Hence, Brooks decided to abandon this type of GOFAI approach and instead to adopt what he called situated robotics, that is, to design robots that function on just a few operating procedures, like the reflexes of insects, and rely on environmental cues to direct the further behaviour of the machine. This is a bottom-up approach that seems to work better with low-level tasks, especially such as can be carried out with teams of robots working together.

As for Boden's other two alternative movements, cognitive dynamics and neo-Heideggerian epistemology, we shall encounter those at the end of this chapter in Section IV. But before we can come to that, we must first present a more thorough philosophical critique of mechanical reason, that is, Boden's own preferred approaches: GOFAI, Cognitivism, Cartesian Materialism and everything else we have bracketed together under the rubric of the MIT-Mind. We begin with the views of one of its prime exponents, Minsky.

2 A Critique of Mechanical Reason

In 1996, exactly at the start of our notional fourth period, Minsky, by then the grand old master of AI, made a grandiloquent oration in Nara, Japan, on what his discipline had accomplished for mankind. What he said shows how in a mere four centuries Man has come from Hamlet's "piece of work" or masterpiece and gone to pieces, having been ground to bits, as Minsky avows:

> The most important thing about each person is the data, and the programs in the data are in the brain. And some day you will be able to take all that data, and put it in a little disc, and store it for thousands of years and then turn it on again and you will be alive in the fourth millennium or the fifth millennium.[33]

33 Quoted in Katherine Hayles, *How We Became Post-Human* (Chicago: Chicago University Press, 1999), 244.

At last, via data storage, near immortality is promised human beings. Minsky is not alone in extolling informational immortality, many scientists and philosophers have joined him, among them his close associate, Dennett, who expressed himself accordingly:

> If you think of yourself as a centre of narrative gravity [Dennett's theory of an illusory subjective identity]… your existence depends on the persistence of that narrative… which could theoretically survive indefinite switches of medium, be teleported as readily (in principle) as the evening news, and stored indefinitely as sheer information. If what you are is that organization of information that has structured your body's control system (or, to put it in its more usual provocative form, if what you are is the program that runs on your brain's computer), then you could in principle survive the death of your body as intact as a program can survive the destruction of the computer on which it was created and first run.[34]

In such thinking science and science fiction have finally become fused and it becomes difficult to distinguish between them. Mechanical reason finally revealed its irrational propensities in aspirations for immortality.

The distinction between science and science fiction has been elided in another way by those, such as Hans Moravec, who argued that there is really no need to concern ourselves with our own informational immortality because by the time this can be accomplished machines of a much higher order of intelligence will have arrived to displace us. Man is an obsolescent carbon based biological being who will soon be replaced by much better engineered non-biological devices, no longer merely "meat machines", as Minsky called us. Our only role is to outdo and undo ourselves by inventing the machines that surpass us. Beyond that there is no further need for humanity; mankind has reached the stage of the post-human.

But by this late date at the turning of the millennium, there were many who were no longer convinced by this kind of talk of informational immortality or higher machine intelligence. It seemed apparent to them that it can only be entertained by those who espouse an informational dualism whose lineal descent from Cartesian Dualism is unmistakeable. According to Evan Thompson "this notion of information… is ultimately regressive, for it entails a form of thinking that is structurally isomorphic to structuralism and mind-body dualism."[35] Thompson directs his critique specifically against Richard Dawkins and Dennett, the former for holding that information is the soul of life and the

34 Daniel Dennett, *Consciousness Explained,* (Boston: Little, Brown, 1991), 430.

35 Evan Thompson, *Mind in Life,* op. cit, 186.

latter for maintaining that it is the life of the soul. The living cell or whole organism for Dawkins is only the vessel for temporarily containing genetic information, which passes unchanged, save for mutations, from generation to generation. For Dennett the body serves a similar role, it is the storing device of the information that constitutes the mind, which could just as well be recorded on computer tape or disc. Hence, the possibility of keeping it for perpetuity and downloading it at any future date by whoever or whatever might be around then.

The philosophers who think along these lines have fundamentally retained a Cartesian Dualism but subjected it to two paradoxical and seemingly antithetical logical processes: first, a materialization of mind to brain; and second, a dematerialization of this mind-in-brain to information processing. As Minsky puts it in his speech, "the most important thing about each person is the data, and the programs in the data are in the brain." It is clear that by "person" Minsky means something approximating to a Cartesian mind, which is at once materialized by being located in the brain and dematerialized by being identified with programs stored in the brain.

Thus before Minsky, Dennett and others can arrive at informational immortality there are three crucial philosophical positions to which they must subscribe. The first is Cartesianism, the view that everything about a thinking being is contained in a self-enclosed mind. The second is the materialization of mind by identifying it with brain in accordance with Cartesian Materialism. The third is the converse dematerialization of mind by identifying it with information processing in line with Cognitivist science and the MIT-Mind in general. We shall address and try to refute each of these steps in turn.

The Cartesian provenance of this approach is evident, for what takes place when the Cartesian mind is materialized and identified with the brain is that all the peculiar metaphysical features of the former are translated and transferred onto the latter. If the Cartesian mind is hermetically self-enclosed, an autonomous thinking substance existing independently of the world outside it, then the brain, which supposedly contains this kind of a mind, is also an autonomous object that is self-contained and independent of any outside entities. On this view, the brain, like the Cartesian mind, can function quite well on its own apart from the body and its environment; hence a brain in a vat can retain all its faculties and continue to exist seemingly unimpaired. The brain has no need of anything outside itself, it is not dependent of on other people or society, language and culture. Such a mind-in-brain is Descartes' thinking substance materialized.

Such a brain is best seen as a machine, for machines work regardless of what is outside them that is not a necessary component of their functioning mechanism. Hence, the prevalence of the view of the mind as machine, for only a Cartesian mind in which everything that transpires is inherent in itself can be viewed in this way as a machine. What helped decisively to foster this view was, of course, the historical contingency of the invention of precisely such an apparently "thinking" machine, the electronic computer. A computer seems to be able to carry out intellectual functions of mechanical reasoning that mimic thinking, if this is understood in a purely technical sense of specific problem solving activities, which the computer can accomplish since it has been programmed to do so. It is true that it can perform some such tasks much better than humans and can frequently beat humans in formalizable activities, such as playing chess. This, of course, tells one nothing about mind or about reasoning or thinking as performed by humans, but it does reveal what kinds of activities can be formalized and carried out mechanically. Chess is apparently one of these as are other such strictly rule-governed games for which mechanical reason is adequate. In time more such possibilities of mechanizing certain kinds of activities will no doubt reveal themselves; though at the moment there are diminishing returns to such efforts as most of the hopes once held out for Artificial Intelligence have failed to be realized to anything more than a marginal extent. Mechanical reason cannot supplant human reason in most things.

However, at the early stages of computing it was easy enough to entertain high hopes for AI, and to jump from a few brilliantly engineered mechanical reasoning feats to the mistaken conclusion that machines can think. And if machines can think, then why not speak, perceive and act, or anything else that humans can do? In this way, the fundamental differences between minds and machines were obfuscated and erased. A kind of logical sleight of hand deception was practiced right from the very start of computing, as we previously showed, through the linguistic trick of metaphorically transferring mental concepts to computers and mechanical concepts to minds, thereby anthropomorphizing machines and mechanizing humans.

This is one of the most corrupting aspects of contemporary language habits; for people begin by speaking of themselves as machines, go on to thinking of themselves as machines, and eventually end up by behaving like machines. Humans are so highly adaptable that they can even imitate non-human ways of proceeding, such as mindless mechanical ones. But in doing so they display mimetic capacities, which are, indeed, what makes them human and so shows them to be utterly different from machines. Unfortunately, in doing so they distort and debase themselves as human beings, and this is where the main harm of the encroaching mechanization of the mind lies. It has become a social process which threatens our culture and civilization with technological barbarism. It starts with the immersion of children in computer games and goes on from there throughout all levels of education which habituate the young to rely on computers for all their intellectual and most of their social needs.

Everything we have thus far presented in this book concerning the origin and nature of the mind argues against the Cartesian view. It also tells against all current attempts of Cartesian Materialism to identify the mind with the brain and then to dematerialize the mind-in-brain by identifying it with information. The mind thus emerges as a brain that carries out information processing, namely a machine just like a computer. But the differences between brains and computers are so great and fundamental that it makes it very dubious whether the brain can be considered a machine in any clear and well-defined sense of that term. It is a physical system of a very different kind; and, obviously, not all physical systems are machines. We shall try to specify criteria for distinguishing between the two types of physical systems in the next section.

The brain is not anything like a computer because it is a constantly fluctuating dynamic system of such incredible complexity that its mode of functioning is still largely unknown and might perhaps never be fully describable. It is in a continual state of dynamic activity which is both internally generated and externally stimulated, for without the interaction of these two processes no orderly structure could arise at a higher level and no mental phenomena could synchronically emerge. How such interactions between the internal and external dynamics take place is now known only in the vaguest generalities. For example, we know much about how the sensory organs function but very little about how perception takes place, for between the stimulation of the organ and the emergence of perception in consciousness all the rest is still largely unexplored *terra incognita*. There are no known codes or "laws" for explaining such synchronous emergence, though much more is known about diachronic emergence in both its phylogenetic and ontogenetic forms.

We know, for example, that the human brain could not have arisen without a long evolutionary process extending over very many species, and without the correlative evolution of language, culture and human society in general. Analogously, no individual human brain can develop without an exposure to language and culture through interactions with and ministrations of others, usually mothers in the first place. Genetic information coded in DNA cannot of itself grow brains, for brains cannot develop without these epigenetic outside factors being present. And, unlike computers, brains cannot be simply manufactured; nor can they self-organize or self-assemble independently of their nurturing environment. Brains are certainly physical systems subject to all the basic laws of physics and chemistry, but they are of such extraordinary complexity that they cannot be understood as we understand machines.

These facts tell against all attempts to treat brains as machines that are analogous to computers. They also tell against speaking of brains as processing information, perhaps the most egregious error that has imposed itself of late on our conception of what it is to be human or what is Man, as Hamlet has termed it. But the brain cannot be spoken of as processing information, since neither the term "processing" nor the term "information", can be applied to it. In regard to the functioning of the brain, we cannot distinguish between what is "message" and what is "noise", the crucial distinction in any information transmission. We cannot even specify what a "bit" is in neurological terms. And it makes little sense to speak of the brain as "computing" or "processing" anything like symbols or numbers since "the neurochemical and neuro-electric actions of the brain are much more like digestion than number-crunching", as Donald declares.[36] Whether network neuroscience or some other such mathematical-computational technique will resolve some of that complexity still remains to be seen.

These considerations bear critically on the dematerialization thesis of identifying mind with information processing that the Cognitivist Revolution promoted. Information processing is a purely logical procedure of symbol manipulation which can in theory be carried out by any computer composed of any kind of material and even by imaginary computers such as Turing machines. It is conceivable that even ghostly ectoplasm could be a carrier of information. This seems to provide a double advantage, at once of freeing the mind from matter and enabling it to be stored in a multiple number of ways in any substance whatever, as well as being transferrable from one medium to another; hence Minsky's "little discs" that could give us informational immortality.

Minsky was evoking a view of mind developed by Chomsky and his students and acolytes that came to be known as the Functionalist theory of mind. Jerry Fodor and the young Hilary Putnam were most active in this endeavour; though in more mature years, when no longer at MIT, but at Harvard, Putnam turned abruptly against his early indiscretions. The Functionalists took the computer metaphor as a literal truth, as Sherry Turkle reports:

> It has often been suggested that the relationship of bits to program is analogous to that of brain to mind. In the "brain" there are physical things – neurons, synapses, electrical impulses. In the "mind" there are images, concepts, ideas, language and thought. Similarly, within the computer

36 Merlin Donald, *A Mind So Rare*, op. cit, 102.

> there are physical things (the bits), and, then, at the level of "mind" there are programs.[37]

Thus the mind is to the brain as the program on which a computer functions is to the mechanism that enables these programs to be run. The dematerialization of the mind was thereby accomplished without apparent reference to Cartesian Dualism in a way vaguely reminiscent of Aristotle's report about the mystical Pythagorean pronouncement that soul is number.[38] If one takes number to stand for quantity of information, then the Pythagoreans have indeed been vindicated two and a half millennia later.

Once having accepted the Functionalist model, Chomsky and his followers embarked on the road to modularity, eventually to such a degree of proliferation that even he was loath to follow. But the opening move derived directly from Chomsky's linguistic theory. The initial idea was that the human mind/brain contains a Language Acquisition Device (LAD) as an inherent genetic birthright, and that this encodes the basic Universal Grammar of all possible languages. The argument for it depends on the specious premise that the human infant suffers from a poverty of stimulus, namely, the amount of language data to which it is exposed is too sparse for it to be able to learn to speak, on Behaviourist principles of learning. This might well be true, but Chomsky jumps from this to the unwarranted conclusion that most of language must be innate, thereby not allowing for other modes of specifically human learning, such as we studied by reference to Frith, Vygotsky and Bruner in Chapter 4. According to Chomsky, language learning is a mere mechanical process of switching on the genetically endowed Language Acquisition Device (LAD) to the parameters of the specific language into which the infant happens to be born, namely, the actual grammar in use. Thus learning to speak is no different to learning to walk, requiring only a little occasional support to activate the "language instinct" as Steven Pinker dubbed it.[39]

We have previously already criticized this view by reference to Jerome Bruner's Vygotskyan approach to language learning as requiring much more than a LAD, if any such actually exists, but also a LASS, a Language Acquisition Support System provided in the first place by the mother or carer. According to Bruner, "the genetic 'program' for language is only half the story. The support

37 Sherry Turkle, *The Second Self: Computers and the Human Spirit* (New York: Simon and Schuster,1984), 277.

38 See Aristotle, *Metaphysics,* Book 1, Chapter 4, op. cit, 856.

39 Steven Pinker, *The Language Instinct: The New Science of Language and Mind* (London: Penguin, 1994).

system is the other half."[40] By support system he means "formatting, fine-tuned responsiveness, modes of embedding language in action and interaction – all these, I finally concluded, comprise the Language Acquisition Support System, the LASS that makes possible the operation of a Chomskian like LAD."[41] When he wrote this Bruner was not quite ready to go all the way and question the very existence of a Chomskian LAD, but it is obvious from his other comments that this is the way he is heading:

> How puzzling that there should be so much emphasis in the eighth decade of the twentieth century on the underlying *genetic* program that makes language acquisition possible and so *little* on the ways in which culture, the parents and more expert speakers (including other, older children) help the genetic program to find expression in *actual language use*… It is not just the *grammar of sentences* that is at issue, but discourse, dialogue, the capacity to understand spoken and written language.[42]

Language, as we have said earlier, is not just parsing, it is moreover communicating and poesis; and grammar is only ancillary to these more essential functions, as, indeed, Lackoff argues. Donald, as we have seen, believes that it evolved purely culturally and socially, not genetically.

However, Chomsky and Fodor and the others saw it quite differently and went even further in the direction of pre-established genetic programming. Fodor grasped hold of the idea of a LAD, overlooking anything to do with a LASS, and generalized this to various other mental faculties, at least initially with Chomsky's blessing. He called such mechanistic devices "modules" and thereby initiated the great hunt for modularity that has been going on ever since, as newer and more recherché modules were being uncovered in the most unlikely places of the mind, until ultimately even Fodor himself was dismayed. According to Fodor, a module is a kind of mechanistic windowless monad of the mind, a mini-computer carrying out its own dedicated computations in isolation from all other monads with whom it cannot directly communicate or exchange information. Each operates automatically and cannot be directed or influenced by higher thought processes, though it sends its messages to them for more elaborate computational processing. Fodor used the terms "encapsulated" and "domain-specific" to convey these ideas about modules.

40 Jerome Bruner, *In Search of Mind,* op. cit, 173.
41 Ibid, 173.
42 Ibid, 173.

Basing himself on this idea of modules, Fodor elaborated a simple model of the mind.[43] There are two levels to the architecture of the mind, a ground floor where modules reside and an upper storey where the more general cognitive processes take place. Information from the bottom flows directly to the top, there to be integrated so that the more domain-general thoughts and ideas can emerge. This is how common-sense and intellect arise. Unfortunately, according to Fodor, what goes on in the upper levels is inaccessible to science, no psychology can explain it, as Boden recounts:

> For Fodor, then, individual thoughts are inexplicable. It follows (or so he claims) that computational psychology – "the only scientific psychology we've got" – must be restricted to "modular" phenomena, such as syntax and low-level vision. He regarded these as scientifically tractable, being innate competence unaffected by conceptual thought.[44]

Fodor's modular ideas about syntax came from Chomsky, of course, and his ideas about low-level vision derived from the work of David Marr on computer simulation of vision also carried out at MIT.

The identification of science of mind with computing has led to this kind of a dead end where nothing can be scientifically known about the mind, according to Fodor and the others, except that which is computable, that is, for which computer models can be built. On this view, it means that just about the whole history of psychology is a dead letter and that this science could hardly proceed much further if reliable computing procedures were to prove no longer viable. And this is precisely what has happened since the collapse of Artificial Intelligences as the preferred approach in modelling the mind.

However, those who were intent on following Fodor into modularity were not to be restricted by the limits he tried to draw on what could and could not be modulated. Syntax and low-level vision were but the start as modules began to proliferate in all other domains of mind. This turned into a veritable avalanche once the idea was applied in the fields of neonatal psychology and evolutionary theory. Elizabeth Spelke and others at MIT found modules for an intuitive, and presumably genetically programmed, naïve physics, biology, psychology and sociology in babies. Apparently, they already had considerable innate knowledge of the behaviour of objects, of species, of other minds and of

---

43 J.A. Fodor, *The Modularity of Mind: An Essay in Faculty Psychology* (Cambridge, Mass.: MIT Press, 1983).

44 Margaret Boden, *Mind as Machine*¸ vol. 1, op.cit, 424.

human relationships. And where there is innate knowledge, there must be an evolutionary process to explain the emergence of such traits.

Hence, it was not long after Fodor's work that a school of combined anthropologists, psychologists, linguists and many others joined forces to establish an evolutionary psychology movement. It was initiated by a collected volume edited by Jerome Barkow, Leda Cosmides and John Tooby, *The Adapted Mind.*[45] They were soon joined by Pinker a linguist at MIT, a pupil of Chomsky, and Dan Sperber an anthropologist from Paris temporarily resident at MIT. Under their combined labours modules began multiplying at a great rate as a veritable cornucopia of plenty. It would take pages to list all the mooted modules. A brief summary would include modules for face recognition, rigid objects, theory-of-mind, tool use, child care, a body-ratio assessor, a symmetry spotter, a habitat recognizer, a cheating detector, emotion perception, kin recognition, and so on and on. The whole mind was seen as a vast panoply of such gadgets, little wonder that the Swiss army knife, with its numerous special-purpose attachments, became the favourite metaphor. By this stage Chomsky and Fodor were appalled by what they had started, but, like the sorcerer's apprentice, they were unable to halt the rapid proliferation of modules once it got out of their control. What the modularists, such as Charles Gallistel and Sperber, hailed as "massive modularity", Fodor denounced as "modular madness".

Reactions were not slow in coming, and these varied from revisions and adjustments to wholesale rejections. The former is best exemplified by Annette Karmiloff-Smith, a former pupil of Piaget, in her book *Beyond Modularity*; the latter by both Deacon and Donald in works we have already referred to previously.[46] Karmiloff-Smith retained modules, but reversed the relation between modules and development: instead of modules being the innate start of development, it was development that produced modules. Furthermore, the development of modules is to some extent culturally variable, which is, of course, unacceptable to strict modularists and evolutionary psychologists. Nevertheless, Karmiloff-Smith is still broadly within the modularist hypothesis as this has been applied in the work on innate knowledge of neonates carried out by her colleagues Alan Leslie and Scott Atran, this time in London not at MIT. There, too, allied work on autism by Uta Frith and Simon Baron-Cohen also took off from the modular hypothesis. But this has been rejected by those who hold that a modularist interpretation does not lend itself to the great

45 J.H. Barkow, Leda Cosmides and J. Tooby, eds. *The Adapted Mind,* op.cit.

46 Annette Karmiloff-Smith, *Beyond Modularity: A Developmental Perspective on Cognitive Science* (Cambridge, Mass.: MIT Press, 1992).

complexity of childhood development where so many interlinked factors are involved, hence that autism cannot be explained in modular terms.

Opposition to modularism has been launched at a more basic theoretical level by Deacon. He goes directly to the source and attacks the founder of the whole modularist approach, Chomsky himself. He severely criticises the "language in the brain" hypothesis from which LAD derives, the primary module and progenitor of all the others. Deacon castigates this thesis as what he calls the "hopeful monster" story, that is, the idea that "a freak mutation just happens to provide a radically different and serendipitous better-equipped organism", namely, Chomsky's "language organ".[47] The problem is that such "an accidental language organ requires no *adaptive* explanation for the structure of language."[48] In other words, it does not call for, and cannot be given, any evolutionary account of how it came to be; and, in fact, Chomsky shies away from any evolutionary theorising and looks askance at Pinker for flirting with the evolutionary psychology exponents. According to Deacon, Chomsky's LAD short-circuits any sequential evolutionary explanation:

> Another appeal of the hopeful monster story is that it promises a definite and dramatic transition from one stage to another in the evolutionary sequence. It offers a single-stop evolutionary account that is much easier to comprehend and organize in one's thinking than continuous change involving multiple factors interacting and overlapping in time in complex ways. In one step, some ancestor crossed the threshold to humanity.[49]

Such criticisms are in accord with Donald's even more severe strictures against Chomsky and the whole of the Cognitivist and computational movement. The attempt to find specific mechanisms in the brain for all kinds of mental activities Donald calls the "Great Module Hunt", especially that for a language module present in the human brain and absent in the brains of ape species: "these purported new modules simply do not exist."[50] And he offers this warning to all module hunters: "the Big Fact is one that should be inscribed on every cognitive theorist's door: No New modules."[51] In other words, there is nothing of this type that separates human from ape brains; the differences that make for mind must be looked for elsewhere. What does distinguish the human brain is simply

47 Terrence W. Deacon, *The Symbolic Species,* op. cit, 35.
48 Ibid, 37.
49 Ibid, 36.
50 Merlin Donald, *A Mind So Rare*, op. cit, 112.
51 Ibid, 112.

its size and the fact that "the regions of the brain that expand most dramatically in humans are those associated with executive functions… for supervising, governing and metacognitive functions."[52] This is the "executive brain" that subtends the "self-system" or "egocenter".[53] Can there be any place for anything approximating to modularity in such a brain?

To modulate or not to modulate, that is the question for the composers of the mind. Is it nobler in the mind to suffer the pullulating proliferation of modules, or take arms against modularity altogether and be accused of harbouring a homunculus in the brain? The choice appears to be between the two extremes of modularity or homuncularity. It is a new more up-to-date version of the old problem of specialized functions in specific locations in the brain as against that of distributed activity throughout the brain carrying out centralizing, integrating and governing functions.

To some extent Donald tends to go in both directions; he does not abandon specialization and a degree of modularity altogether, but at the same time he plumps for homuncularity as well. Thus he maintains that "the actual sound system of language may be modular, but not the higher intellectual drivers that give language its great representative power."[54] But he also asks rhetorically "if there is no homunculus, how can our experience of conscious unity be explained? Where is our experience of agency rooted?"[55] Many of the Cognitivists scientists and philosophers hold that these are bootless questions, that there is no homunculus and that the mind is a warring system of many agents without any central consciousness or governing Self. And, hence, it follows that our sense that there is a coherent Ego is an illusion that the brain devises to deceive itself; presumably for some evolutionary reason that favour such self-deception. Such are Minsky's "society of mind" theory and Daniel Dennett's "pandemonium of the brain" theory, "replacing the traditional homunculus of Descartes with a committee of equally inscrutable quasi-homunculi called 'agents'."[56]

In opposition to Minsky, Dennett and the other Cognitivists, Donald reinstates something of the "idea of a homunculus" or what he calls a "perceptual egocentre, which has been a staple of the concept of selfhood since the work of William James and Gordon Allport."[57] But this he distinguishes from the Cartesian homunculus of an autonomous and hermetically sealed-off Ego by

52 Ibid, 112.
53 Ibid, 193.
54 Ibid, 204.
55 Ibid, 29.
56 Ibid, 29.
57 Ibid, 134.

treating it as a "body-based homunculus", one that "feeds on the many body maps registered in our brain, [an] ego-centre that ties all those maps together."[58] In the human mind, "all conscious experience is referred to that ego-centre"[59] In humans it is what enables self-representation and thereby affords self-consciousness; and "ultimately it is the foundation of what we call meaning in the subjective sense."[60] It is the basis of self-identity, and, surprisingly, also for the language capacity of humans: "language could be the greatest beneficiary, rather than the cause, of the extended human capacity for self-consciousness."[61] Donald thereby inverts the usual causal accounts to be found in many philosophers and psychologists who regard language as the basis of self-consciousness. As we noted previously, this is where Donald and Deacon differ once again, but we will not pursue this difference any further here.

An even stronger onslaught on modularity and the Cognitivist Revolution than either Donald or Deacon muster came from Edelman, who already by the early 1990s was warning of dangers inherent in the proliferation of the MIT-Mind:

> The blend of psychology, computer sciences, linguistics and philosophy known as cognitive science has grown enormously. As with all vigorous efforts, ill-founded or not, much has emerged that is of great interest to scientists and non-scientists alike. Not the least of the positive results has been the routing of simpleminded behaviourism. But at the same time an extraordinary misconception of the nature of thought, reasoning, meaning and of their relationship to perception has developed that threatens to undermine the whole enterprise.[62]

In that same work of 1992 in the concluding Appendix Edelman provides an epitome of an already on-going counter-revolution to the Cognitivist Revolution in which he quotes almost a dozen authors from varied disciplines.

In great detail and analytical thoroughness, Edelman sets out the insurmountable differences separating brains and computers or ideal Turing machines in an even earlier work from 1989:

58 Ibid, 134.
59 Ibid, 135.
60 Ibid, 136.
61 Ibid, 137.
62 Gerald H. Edelman, *Bright Air, Brilliant Fire*, op. cit, 228.

> I have elaborated on the themes of density dependence, variance, and causality in selective systems because I wish to draw a conclusion regarding functionalism, information processing and computation as a source of behaviour and thinking… More specifically, according to the TNGS [theory of neural group selection], the brain and its processes of perceptual categorization do not represent a Turing machine – to account for higher brain functions, there is no need to posit a "tape" describing a sufficient number of effective procedures as existing either in the observed world or in the genetic structures of the organism, and there is no evidence that such a tape exists. Genetic and epigenetic constraints, the laws of physics, and natural and somatic selection are sufficient. The emergence of consciousness as an evolutionary event and the subsequent emergence of language have led to the *idea* of information. But we should not reify it as a priori or immanent property of the world.[63]

He spells out the critique of any likening of brains to Turing machines still further:

> An analysis of the evolution, development, and structure of brains makes it highly unlikely that they are Turing machines. The tape read by a Turing machine is marked unambiguously with symbols chosen from a finite set; in contrast, the sensory signals available to the nervous systems are truly analogue in nature and therefore neither unambiguous nor finite in number. Turing machines have by definition a finite number of internal states, while there are no apparent limits to the number of states the human nervous system can assume (for example, by analogue modulation of large numbers of synaptic strengths in neuronal connections). The transitions of Turing machines between states are entirely deterministic, while humans give ample appearance of indeterminacy.[64]

In short, brains are not machines but biological organisms of an astounding intricacy that is for all practical purposes inexhaustible. Already in an earlier publication of 1987 Edelman had tackled head on the Information Theory approach to the brain that had been the mainstay of the Cognitivist Revolution:

63 Gerald H. Edelman, *The Remembered Present,* op. cit, 259–60.
64 Gerald H. Edelman, *Bright Air, Brilliant Fire,* op. cit, 223 and 225.

> It is surprising to observe that neurobiologists who disbelieve any resort to interpretative homunculi can nonetheless believe that precise algorithms are implemented and that computations and calculations of invariances are taking place inside neural structures. These beliefs persist despite the presence of enormous structural and functional variances that exist in neural tissue – variances that would doom any equivalent parallel computer to producing meaningless output within short order even with the best error correcting codes.[65]

Backing up his charge that only a homunculus in the brain could implement the Cognitivist information processing programs, he asks rhetorically: "Who or what decides what is information? How and where are 'programs' constructed in situations never before encountered."[66] He points out that according to Shannon's Theory of Information the measure of information must be specified à priori based upon a code being established and the probability of receiving any signal under that code specified in advance. But none of these conditions that apply to technological communication systems hold for biological organisms, where what is noise and what is information can only be arbitrarily determined.

Edelman directs his arguments explicitly against the two doyens of information processing in the brain, Marr and Chomsky, both located at MIT. He charges that their views place utterly unrealistic requirements on the brain:

> This view culminates in discussions of algorithms and computations, on the assumption that the brain computes in an algorithmic manner (David Marr, *Vision: A Computational Investigation into human representation and processing of visual information*, 1982). Categories and natural objects in the physical word are implicitly assumed to fall into defined classes or typologies accessible to a program. Pushing the notion even further, proponents of certain versions of this model are disposed to consider that rules and representations (Noam Chomsky, *Rules and Representations*, 1980) that appear to emerge in the realization of syntactic structures and higher semantic functions of language arise from corresponding structures at the neural level. If statistical variation enters at all into such a view of the brain, it is considered in terms of noise in a signal, which in

65 Ibid, 42.
66 Ibid, 43.

> information processing models is taken to be the main manifestation of variation.[67]

Having to rely on noise for novelty relates to the school of cyberneticists associated with Heinz von Foerster who seek to derive order from noise in general. But this is a far too specialized and technical application of Information Theory to be dealt with here, and it is doubtful whether it has any relevance to the brain. The problems of applying any such theory to the brain is set out by Edelman under seven major headings which are "designed both to sharpen the major issues connecting psychology to neural structure and to expose the inadequacies of various forms of information processing".[68]

Edelman's critique of information processing in the brain, and thereby of a whole Cognitivist project, seems irrefutable, no matter what one makes of his own counter proposal of a "neural Darwinism" based on the general idea of the pruning of neural connections, which we cannot discuss here. That it is not easily countered is implicitly demonstrated by the way that Boden, a stalwart life-long upholder of GOFAI (Good Old Fashioned Artificial Intelligence) deals with Edelman's critique. She treats him to GOFBS (Good Old Fashioned British Style) debunking, castigating him as "obscure, incomplete, one-sided and mistaken on crucial facts… riddled with unexamined assumptions, and crude misjudgements of the views he was ridiculing."[69] Methinks the lady doth protest too much. She provides not a single point of rebuttal to substantiate these *ad hominem* charges, and simply refuses to engage with his arguments.

3 On Minds and Machines

Boden's work presented under the title of *Mind as Machine* raises the old almost perennial question of whether the mind is a machine. But this way of putting it contains an implicit dual assumption: namely, that the mind is the brain and that the brain is a machine. We have already dealt with the identity thesis at length but we have not as yet expressly discussed the mechanistic thesis that the brain is a machine. This is an old assumption going back to Descartes and given a materialistic formulation by la Mettrie. It gained a new lease of life in the post-Second World War period with the invention of the computer

67 Ibid, 38.
68 Ibid, 38.
69 Margaret Boden, *Mind as Machine,* vol II, op. cit, 1203–4.

and the development of the MIT-Mind. Hence, before we proceed with the main topic of the brain as machine we will briefly review where we now stand in regard to the MIT-Mind.

Our own approach to the philosophy of mind differs from the MIT-Mind in at least three fundamental respects. Firstly, the mind is not located in the brain and cannot be identified with the brain or even with the brain plus the body. Secondly, the mind is an evolutionary product of emergence, hence any study of it must be approached in primarily biological and anthropological terms and not in those of formal logic, mathematics, engineering or computer science. Thirdly, the brain is not a machine like a computer; in fact, as we intend to argue next, the brain is not a machine at all on any possible standard definition of "machine", though, of course, it is a physical system. But not all physical systems are machines, indeed, very few are, despite Descartes' mechanistic philosophy which the Materialist Cartesians have continued. To conclude this work we will briefly comment on each of these three points in the light of what we have presented in the preceding chapters of this work.

On the first point, we have strenuously argued against any identity thesis and especially that which identifies the mind and brain. If the concept of emergence applies to this relation, then the two sides cannot be identified with one another, since the logic of emergence mandates a non-identitarian indissolubility. We have argued that this concept of emergence applies throughout the sciences. If it did not apply, then science as we know it would not be possible. But since, as a matter of fact, science is possible, hence emergence does apply. This is the quasi-Kantian argument from the existence of something to the grounds of its possibility we invoked previously. Thus we proceeded to show how emergence applies throughout all the realms of the Great Chain of Being from what we designated as E^1, the emergence of matter, right through to E^5, the emergence of the human mind. Metaphorically this justifies Hamlet's assertion that Man is the quintessence of dust.

On the second point, we have shown that any science of mind must follow the evolutionary route from mindless animals to those with minds, in the first place, and then from animal mind to human mind. In following this route, it is best to rely on the work of biologists, anthropologists, neurologists, psychologists, linguists, and cultural and social scientists in general, rather than the people of the "hard" sciences artificial intelligentsia variety responsible for promulgating the MIT-Mind. At present work in all the above sciences is proceeding at great intensity and much is being revealed about biological life in general; about the specifics of the evolution of the hominin species and humans; about the structure of the human mind; and about the formation of language, culture and group life or "society" that enabled that kind of mind to develop. It is in these areas that the vital questions about the origin and nature of the human mind are being addressed, not in computer simulations.

This is not to say that the present answers are the final ones, for we know very well on historical grounds that they are bound to be revised or totally repealed at some point in the future. How, when and why they will be overthrown, namely, what weaknesses and faults will be shown up, we are, in principle, precluded from knowing. As to whether there will ever be final answers to our questions that, too, is unknowable. It is bound up with the issue of whether science will ever come to an end. We can, of course, speculate about such matters, but it is best to adopt the attitude of the ancient sceptic philosophers and practice *epochḗ* in its original sense of a noncommittal suspension of judgement about such future matters.

On the third and final point much more will need to be said than we can present in this book. Is the brain a machine? This issue, which concerns us most in this section, is a matter for the very "hard" sciences which we previously ruled out as telling us much about the mind. But with the brain it is different; there we must consult the logicians, technicians and engineers who, after all, are the experts on machines. Since we cannot even begin to outline the extensive study necessary to debate this issue, we must perforce conclude with merely a few notes on this final topic.

The issue turns largely on the definitions of what is a machine; though this does not make it a trivial verbal matter since definitions of concepts in science sum up theories, and theories are ultimately what count. But before we turn to strict scientific definitions, we might note that the term "machine" etymologically goes back a long way to the original Latin, as in the expression *deus ex machina*, and that by the time it reached Hamlet it was already being used loosely to refer to the body. Descartes used it in this sense. But by modern times the term "machine" had given way to the more grandiose term "technology", which derived from two Greek words *techne* and *logos* that had originally stood in opposition to each other, but have since combined on the basis of advanced science.

When high technology came of age at the end of the Second World War, the perfect machine for combining technical instrumentation and logic was invented, the computer or logic machine. With this machine it was presumed that all logical operations could be performed, such that anything that could be reduced to computation in its logical aspects could be solved. Hence, anything that could be stated clearly and unambiguously in this sense could be carried out by a computer. Thus the computer became the key to all logical problems and that, it was believed, was synonymous with all problems of thought. For those who thought in this way, as Edgar Morin expressed it, "Computo, ego sum" became the new neo-Cartesian dictum. Computing was taken as the key to all mental activities and the brain was seen as the machine that carried out these mechanical operations.

To those who were opposed to this wholesale technological tendency to reduce the mind to computation, it seemed that the only counter was to find things that the mind could do which a computer could not equal. This was the trap that Hubert Dreyfus fell into in his early days. Among other things, he predicted that computers would not be able to play championship chess. In due course, he was soon proved to be wrong when a computer beat the world chess champion. He made the mistake of assuming that there were some things that computers could not do outright, or if they could, then in a less optimal way, and this was the reason they were inferior to humans. And this is where he went wrong, for there is no single clearly specifiable thing that in principle computers are precluded from performing. The reason for this is that it is, by definition, always possible to write a program for any well-defined activity, especially if it is rule-bound like chess.

However, from this it does not follow that the computer, or even a collection of computers, can do everything that a human mind can do, for what a human mind can do is open ended and not specifiable in well-defined terms. There are no limits to what a human mind can do given the right conditions and learning opportunities – the mind is in this way unlimited and unbounded. By contrast, any program is, by definition, limited for it can only do that which it was designed to do, even when this leaves scope for adaptation and self-correction, which will always remain bounded. Hence, it is, in practice, always possible to distinguish a mind from a computer, for the mind, being a holistic entity, can range freely across all activities, whereas a computer will be restricted to a small select sample.

We must avoid arguments about in principle possibilities that are meaningless as far as human beings are concerned. Hence, to argue that a problem is solvable, but that it will take something approximating the age of the universe to do so, makes it in principle possible but in practice meaningless. The time it takes to solve problems can never be discounted if meaningful results are sought, but this is precisely what the original design of the Turing machine assumes for it makes no allowance for time or any other practicalities. Once this is allowed for, the argument about what computers can do becomes very different from what Turing or other logicians assume it to be by taking no account of time or any other realities.

In actual practice, which is always far behind "in principle" possibilities, it has been found that what can be done by computers is extremely limited. Computers are good at logical and rule-bound activities; but it is not easy and it takes a great deal of *human* ingenuity to reduce even some simple ordinary activities to that state, namely, to write programs for them. Computers are, consequently, very bad at any task that cannot be translated into that form, or can only be done so very schematically. Thus, despite, the huge labours that have been expended at mechanizing perception, such as by David Marr and many others at MIT, it is still not possible to visually track the motion of an object by such means, something that a baby of a few months can perform. Though there has been some success in the simple identification of objects and even faces. Thus the huge promise held out for GOFAI in its early days has been largely unfulfilled, which does not mean that it has proved useless. Indeed, Artificial Intelligence has found numerous applications in both military and industrial technologies in a huge range of automated tasks; and it is now also invading ordinary life activities, which spells danger for human living.

There are two main stumbling blocks that have been encountered in practice that severely limit the problem-solving abilities of AI and hence restrict the range of its applicability: the combinatorial explosion and the "frame problem". The first concerns the impossibility of going through every possibility in the search for a solution because the possibilities can be almost limitless, sometimes exceeding the number of atoms in the universe, which not even the fastest conceivable computers can handle in the time that would make any solution relevant. Even rule-bound and strictly definable activities, such as chess, are prey to the combinatorial explosion, and it takes great human ingenuity to find short-cut algorithms to overcome this hurdle. This is not always possible. Of course, given enough time and computing power every problem that can be programmed is in principle solvable, but is of little relevance in practice where time is of the essence. Furthermore, not every problem is programmable due to the second weakness of AI, the "frame problem".

The "frame problem" is the impossibility of specifying every conceivable exception to the application of a program, when the well-known Murphy's Law – that anything that can go wrong will go wrong – has to be always taken into account. Boden sees this difficulty as constituting two problems:

a) Knowing just which aspect of a situation would be changed by a particular action, and which not. For instance, whether it is true that if a person has a telephone he still has it after looking up a number in a telephone book.
b) The second was the problem of reasoning with incomplete knowledge – where one doesn't know, for instance, whether the page has been torn out of the phone book, or whether the phone system is out of order.[70]

70 Margaret Boden, *Mind as Machine*, vol II, op. cit, 774.

Researchers in AI have spent decades working on these problems, and the general consensus is that the frame problem is unsolvable, as Boden readily admits. She quotes one such researcher, Patrick Hayes, who together with John McCarthy, one of the founders of the discipline, "spent many years trying to defuse the problem, working respectively on non-monotonic reasoning and naïve physics. (Today, Hayes feels that they were over-sanguine in the 1960s and 1970s, and that the frame problem is probably unsolvable…)".[71] Indeed, the initial hopes for the whole discipline were vastly over-sanguine, even allowing for the naiveté of its founders and their faith in formal logic and mathematics, and the excess expectations that are inevitable at the start of a new discipline.

This certainly was true of the great mathematical logician, Turing, who in a paper published in 1950 sought to demonstrate that anything a human can do a computer can do just as well and most often better. For this purpose he devised the celebrated Turing test to determine whether it is always possible to distinguish between a human being and a computer or robot going solely by verbal performance. What he describes is something akin to a quiz where the two contestants have to be identified purely by the questions they answer.[72] Hence, this is an artificially limited situation for it involves no bodily or sensory activity, namely, both embodiment and embedment in a specific context or environment, the two primary characteristics of the mind, are by definition ruled out. Since computers have no bodies, they cannot engage in acting and perceiving in interaction with others in an unspecified and ill-defined environment, consequently they are notoriously unable to perform the kinds of thing for which minds are so eminently suited. However, it is possible that computer performance will improve on present standards in the future provided that much greater computer power will be available. It will also depend on the greater ingenuity of programmers, namely, on human developments that are at present unpredictable. Ultimately, what computers can do is limited by the human ingenuity required to direct them. Like all machines, computers depend on the humans who create and service them.

Even in the restricted and artificially constrained context of most versions of the Turing test where only verbal performance is in play, computers still prove notoriously limited. They lack the vast storehouse of common sense knowledge that humans gain through experience over a life-time of acting and learning in dealing with all kinds of objects, animals and, above all, people in

71 Ibid, 774.

72 Ibid, 1351–6, where Boden reviews the vast literature on this subject outlining the numerous variants of this test.

an unspecifiable range of contexts and situations. By utilizing such a vast accumulation of knowledge and experience, most of it tacit and unconscious, humans can envisage how things and people behave and how situations unfold and develop which gives them vast inferential capacities. Imagination is the means of thinking and knowing that gives them their unequalled advantage. In using their imagination and entertaining countless imaginary what-if hypothetical scenarios they can generally arrive at the right answer to any unforeseen problem, whereas computers can only address problems in terms of the explicit knowledge that is built into their programs. But it is possible that in respect of sheer book learning computers can exceed human capacities, being, in effect, automated encyclopaedic search-engines. Hence, if the Turing test were limited to this kind of knowledge, it is conceivable that computers would do as well as any humans, even the most learned quiz contenders. But unless such artificial restrictions are built in, computers are bound to fail at any other version of the Turing test.

It is ironic that even the most diehard opponents of Artificial Intelligence, and of the arguments derived from it about the indistinguishability of humans from machines, tend implicitly to accept that computers can do just as well as humans on the Turing test without gross restrictions. Thus John Searle bases his Chinese Room argument on the implicit assumption that an automatic process of translation, performed either by a human being ignorant of Chinese or by a computer – for the purposes of his argument it comes to the same – could do just as well as a skilled translator fluent in Chinese.[73] In other words, Searle accepts the initial premise that perfect automatic translation is possible; which, not only grossly underestimates what is involved in translation, but also grants his opponents everything they would ask for and gives the game away, since his opponents can then argue quite cogently that the system as a whole understands Chinese. However, perfect translation by computer, namely, automatic translation, is not possible except to a very limited extent, since even adequate translation of such things as Chinese poetry requires a human being who understands Chinese language and culture and knows very much not only about China, but also about human life in general. Hence Searle's argument cannot even get started, for his initial premise contradicts his conclusion. For if a computer – or human being acting purely automatically without understanding or regard to time or any other practicality – could produce perfect translations then, indeed, it must be able to understand language and life and, so ipso facto, to think. But this is, *ex hypothesis*, a contradiction. One has

73 Ibid, 1384–5.

only to examine the output of machine translations to see how ludicrous the idea is.

This brings us to the crux of the whole argument about whether the brain is a machine. If the brain can do what no computer can do in anything like an acceptable time – of course, a computer can, in principle, do anything whatever if a program can be devised for it and if it is allowed to run for an unlimited time – then it follows that the brain is not a system that is anything like a computer. Is it then some other kind of machine, perhaps not a von Neumann configured computer but a Turing "Oracle" machine, or a Connectionist net, or anything else we might call a machine? It all depends on what we can call a machine, that is, it ultimately comes down to a matter of definition. Definitions in this context are not arbitrary verbal stipulations but conceptual specifications derived from a scientific theory. Hence, everything depends on how we theoretically understand various kinds of systems and how we then proceed to differentiate between these, distinguishing between those that are mechanical and those that are not – for not all deterministic physical systems are mechanical, as Descartes supposed. Hence it is possible to argue that though the brain is a deterministic physical system, it is not a machine under any acceptable theoretical definition of that term. What kind of system it is, in fact, is a matter of on-going research and debate to which we shall presently return. But first we must establish how the term "machine" might be defined.

Perhaps the simplest and most obvious definition of machine is that it is any mechanical device that human beings, namely, mechanical engineers, create. On this definition, Giambattista Vico's principle of *verum et factum convertuntur* applies, that is, that anything we make we can *ipso facto* understand. Thus we can understand how the machines we make function since we have plans and blueprints of how they work. This is true even of computers whose specific output we cannot predict or know in advance, since we still know how this can be arrived at by following a program; even when that program is so vast and cobbled together and patched up at various times by different people that we can no longer tell how it was put together or how it functions in any detail. It is in this sense that computers will always count as machines even if they were nothing like our present machines and operated on different principles, such as quantum computers on which much work is going on at present, or even Turing's fabled "Oracle" machines able to compute numbers that are not Turing computable.[74]

However, it is conceivable that at some time in the future we will be able to create devices that are not computers but can carry out extremely complex

---

74 Ibid, 179.

functions, perhaps approaching closer to human minds. It is thus possible that we might be able one day to make devices whose operations we can no longer survey or represent in the form of a mathematical equation or a plan; such a thing might be imagined perhaps as extremely intricate connectionist networks. If we had such devices, then we would no longer understand how they worked and this would contradict Vico's verum-factum assumption that anything we make we can understand. Would such things be machines or something else for which we have no word at present?

This was a possibility that right at the very start of the computing age von Neumann considered at the Hixon Symposium in 1948, as Dupuy reports:

> …finding himself taxed by the neuro-physiologists (including McCulloch) for not stressing enough the differences between natural and artificial automata, he replied that this distinction would grow weaker over time. Soon, he prophesied, the builders of automata would find themselves as helpless before their creation as we ourselves feel in the presence of complex natural phenomena.[75]

The most complex of all natural phenomena, the human brain, is what von Neumann most probably had in mind as the reference to neurophysiology implies. The issue that concerns him, as Dupuy explains, is what it means to model the brain. With any object that approaches that level of complexity, the model begins to be as complicated as the object it represents:

> The model, which before had been hierarchically subordinated to a reality that it managed only to mimic, stood now as the equal of its referent. Consider the problem of trying to model a natural object whose complexity exceeds the critical threshold posited by von Neumann. The model, if it is to be faithful to what it represents, must exceed this threshold. But then it will be not only a model of its object, but also a model of itself, or rather its own behaviour.[76]

This is the kind of dilemma that modelling any complex object that even begins to approximate the complexity of the brain represents. The fact that we can understand the model as little as we understand the object itself means that the principle of verum-factum ceases to operate and that if we can perhaps

75 Jean-Pierre Dupuy, *The Mechanization of the Mind*, op. cit, 142.
76 Ibid, 142.

one day in the future create devices which we can no longer model or, therefore, understand, these could no longer be defined as machines.

If this were ever to happen, then the very distinction between machine and natural system would, as von Neumann contends, grow weaker and weaker. We would no longer be able to distinguish between those systems which are clearly machines and those that are clearly not so. Hence, the more something even begins to approach the complexity of the brain, the less it can be considered or spoken of as a machine. This substantiates our view that the brain cannot be defined as a machine since we cannot understand it in the way we can understand machines. It is far too complex for that kind of an understanding, as von Neumann intimated. Hence, it follows that "brains" and "machines" are contradictory concepts, and that the expression "brain as machine" is an oxymoron. And, so, too, is "mind as machine", the title of Boden's book. But that does not preclude an imaginary continuum between one extreme and the other and an intermediate stretch where neither concept, brain or machine, any longer applies.

Von Neumann was not the only one to raise such difficulties, many others expressed themselves in similar terms. As early as 1948, Ross Ashby, the designer of the aptly named Homeostat robot and the author of the oxymoronically entitle book, *Design for a Brain*, expressed similar worries when he said that such devices could turn out "too complex and subtle for the designer's understanding."[77] With the much later return of Connectionism in the late 1980s, these fears were to some extent realized. Thus Boden quotes from a workshop report from the winter of 1986–7 by the British computer scientist D.A. Partridge:

> Connectionism and its second coming is the next unavoidable issue [for AI]… [If] subsymbolic networks are necessary to support intelligent systems, then the conceptual transparency of the resulting AI systems is likely to be *on par with that of the brain – i.e. somewhere close to zero.* Homogeneous networks or subsymbolic elements will severely test our abilities to understand our models. So if subsymbolic network architectures are in some sense necessary then that might be bad news for AI.[78]

What is bad news for AI is good news for those who insist on the difference between brains and machines.

77 Margaret Boden, *Mind as Machine,* vol. 1.
78 Ibid, 1167.

Brains are, of course, natural systems completely subject to the known laws of physics and chemistry, but a brain is not a machine because it is of such complexity that we can neither completely model it nor understand how it functions in anything even approximating to the way we do so with machines. Some biologists such as Paul Weiss, had almost intuitively grasped these differences between natural systems and machines since the very start of the Cybernetic movement that was founded on the opposite idea, namely, that there are no such fundamental differences in kind and that the engineering principles, such as feedback and information processing, apply to both kinds of system indifferently. As Dupuy reports, "unlike the cyberneticians, Weiss made a very clear distinction between machines and systems. The latter are natural rather than nominal totalities..."[79] But it was not till neurologists, such as Edelman, addressed this issue that we could begin to appreciate what kind of complex system the brain really is, and how far removed it is from machines:

> Indeed, the circuits of the brain look like no others we have seen before. The neurons have treelike arbours that overlap and ramify in myriad ways. Their signalling is not like that in a computer or a telephone exchange; it is more like a vast aggregate of interactive events in a jungle. And despite this, brains give rise to maps and circuits that automatically adapt their boundaries to changing signals. Brains contain multiple maps interacting without any supervisors [homunculi], yet bring unity and cohesiveness to perceptual scenes.[80]

And it is not only human brains that can do this; animal brains do so as well, hence they, too, have minds. And even mindless animals, invertebrates and lower vertebrates, have brains that are capable of extraordinary behavioural feats that at present no machine can duplicate.

It can now be in little doubt that the whole attempt to view the mind as a machine has turned out to be a disaster in science and philosophy. This does not gainsay the fact that it is in many ways a technological triumph; it has resulted in innumerable automatic devices and working systems, which have been of great economic utility and many of which have the military application for which they were funded in the first place. Whether this is ultimately good or bad for humanity depends on how one views our present technological age. But whatever benefits have undoubtedly accrued, they have at least in

79 Jean-Pierre Dupuy, *Mechanization of the Mind*, op.cit, 132.
80 Gerald Edelman, *Bright Air, Brilliant Fire*, op. cit, 69.

part been bought at the cost of scientific and philosophical confusion with regard to mind. This is also Edelman's conclusion:

> If you consider these extraordinary brain properties in conjunction with the dilemmas created by the machine or computer view of the mind, it is fair to say that we have a scientific crisis. The question then arises as to how to resolve it. For a possible way out, let us look to biology itself rather than to physics, mathematics or computer sciences.[81]

The crisis in science is also a crisis in our whole academic system of disciplines, for the dominance of the "hard" sciences has had deleterious consequences on the "soft" sciences, on the social sciences and humanities, and, above all, on philosophy. How to address this situation is something we shall go on to consider in the *Postlude*.

4 The Phenomenological Opposition

Despite all that we have so far said about the hegemony of the "hard" sciences and, as an outcome of that, of the dominance of the MIT-Mind in many areas of the "soft" sciences and of Cartesian Materialism in philosophy, nevertheless, it did not go completely unchallenged. Opposition to it came from various quarters in the sciences as well as philosophy. We have referred to and quoted from many of these critics in order to marshal our argument against this domination of academia in America, Britain and other university systems. There were even voices of protest at MIT, such as the computer scientist Joseph Weizenbaum and the philosopher Hubert Dreyfus; though Dreyfus did not survive long at MIT; he moved to Berkeley where he joined Searle and pursued his Phenomenological predilections to continue his campaign against AI from there. Opposition grew as the twentieth century drew to its close from other quarters as well, gathering into its ranks one time stalwarts of the MIT-Mind, such as the philosopher Putnam, the psychologists Ulrich Neisser, Bruner and Miller, and the computer scientist Terry Winograd, once a darling of the artificial intelligentsia who was branded a turncoat when he turned against them.

There had been, of course, constant opposition from the so-called Continental philosophers, but in America and Britain these did not count for very much since they were looked upon as minority sectarians or really nuisances. In Europe, however, such schools predominated, above all the Phenomenologists,

81 Ibid, 69.

and they could have no truck with any mechanistic view of mind. As we previously noted, Boden refers to this kind of opposition when she speaks of "neo-Heideggerian epistemology", not realizing that it was really neo-Husserlian and that a wide divergence separates these two Continental schools, as we shall presently show.

Towards the end of the century the neo-Husserlian opposition had organized itself into the Naturalizing Phenomenology movement that we discussed in the previous chapter. It is the attempt to combine Husserl and the natural sciences, especially the Cognitive sciences. This *demarche* has assumed an international dimension with contributors even from the Anglo sphere, America, Canada, Britain and from as far afield as Australia. However, as might be expected, the movement is mainly centred in Paris, at the *Centre de Recherche Épistemologie Appliqué* (CRÉA) headed by Jean Petitot, Francisco Varela and Jean-Pierre Dupuy. They organized a conference whose proceedings were published in 1999 under the title *Naturalizing Phenomenology.*[82] Evan Thompson and numerous others, not necessarily all inclined to Phenomenology, have made contributions to this collective endeavour to which we shall refer.

It is far beyond our scope here to discuss individual papers, or even the general project of reconciling Phenomenology and natural science, or specific issues, such as whether "neurophenomenology" as a science makes sense. We cannot even take issue with the large claims that Petitot and Varela make in their introduction that such new departures in science as systems, complexity, chaos, catastrophe theory, etc. represent a new paradigm or even a new "scientific revolution" comparable to what Descartes and the natural philosophers of the seventeenth century accomplished; though obviously given our account of "chaoplexity", as Horgan refer to it, in Chapter 2 this would be highly implausible. All we aim to show is that this movement goes counter to the equally inflated earlier claims for the so-called Cognitivist Revolution as a new paradigm in science.

In direct opposition to the information processing computer model of the brain, many of the contributors to the volume espouse a dynamic model of the workings of the brain. From the point of the critics of Cognitivism, it has the obvious advantage of obviating any need for internal representations and for computation in general in explaining how the brain functions. It is expounded by one of its leading practitioners, Tim van Gelder:

> In cognitive science, most (though by no means all) dynamical models are neural networks, in which the set of interdependently evolving

82 Jean Petitot et al, *Naturalizing Phenomenology,* op. cit.

> quantities are the activity levels of the neural units. According to the dynamical approach, cognition is not the manipulation of symbols, but rather state-space evolution within a dynamical system, and the emergence of order and structure within such evolution.[83]

In place of the computer, the model of the brain that van Gelder offers is that of Watt's steam-engine governor, which from our point of view is not much of an improvement since it is an even simpler cybernetic mechanical device. Nevertheless, van Gelder believes otherwise, for he insists that "standing behind the dynamical approach is an alternative philosophical tradition…[with] figures such as Dreyfus and Varela, which takes the essence of mind to be an ongoing active engagement with the world."[84]

It is to this alternative philosophical tradition that Boden refers in her *magnum opus* on the history of Cognitive science when she speaks of "neo-Heideggerian epistemology" as another counter to Good Old Fashioned Artificial Intelligence (GOFAI) to which she remains an adherent, despite her willingness to admit criticisms and unfulfilled hopes. She has primarily in mind such critics as Dreyfus and Varela whom she lumps together as Continental philosophers, as is customary in America and England, without realizing that the former is a Heideggerian and the latter a Husserlian, a crucial difference, as is brought out by Thompson. Thompson, who is a Husserlian, objects to Dreyfus's Heideggerian "interpretation of Husserl as a representationalist and protocognitivist philosopher, as well as his Heideggerian critique of Husserl thus interpreted."[85] In what follows we shall seek to bring out this difference by reference to Thompson.

Thompson himself presents an "enactive view that mind is embodied in the active organism and embedded in the world."[86] Elsewhere he states that "according to an enactive approach there is a deep continuity of life and mind, including conscious mentality, and the philosophy of mind needs to be rooted in a phenomenological philosophy of the living body."[87] The phenomenological philosophy of the living body is, of course, Husserl's philosophy. Husserl distinguishes the living body, that is, the subjective sense of one's embodied being, from the body as mere object like any other in the physical world. This fundamental separation of the two senses of "body" is etymologically facilitated

83 Tim van Gelder, "Wooden Iron? Husserlian Phenomenology Meets Cognitive Science", in Jean Petitot et. al. *Naturalizing Phenomenology* op. cit, 252.
84 Ibid, 252.
85 Evan Thompson, *Mind in Life,* op. cit, 414.
86 Ibid, 187.
87 Ibid, 222.

in German by the two words *Leib* and *Körper*. Much of the philosophy of Husserl is concerned with specifying the distinction and relation between the two "bodies", which is expressed by Thompson as the body-body problem, a quite different way of putting it from the usual mind-body problem. In Husserl this extends to the still wider problem of the relation between his so-called life-world and the physical world, which concerns the role of science in relation to Phenomenology. As is well known, through his insistence on *epoché*, Husserl sought to keep the two apart, whereas the Naturalizing Phenomenology movement is designed to bring them together again.

Despite his attempt to keep science out of Phenomenology, Husserl is not averse to science as such, which is what distinguishes him from Heidegger. But this is not the only point of difference. They are fundamentally at variance not only about science but also on issues related to it, such as the body and the life-world. For Heidegger human being is what he calls *Dasein,* a notion that is quasi-theological in origin, as is most of Heidegger's philosophy. *Dasein* makes no reference to any bodily presence or enactment and is completely removed from biology and the natural world of science. According to Heidegger, *Dasein's* being is a matter of mood, above all of *Angst* or fear of death; and it is this mood proper to *Dasein* which situates it in the world not its body, which Thompson criticises accordingly:

> From a neurophenomenological point of view, Heidegger's account is unsatisfactory. For one thing it is strangely disembodied, for the body plays no role in his account of mood and atunement, despite the attention he gives to fear and the "fundamental mood" of anxiety.[88]

It follows from this that in his view of *Dasein* or human being Heidegger is totally averse to biology (except perhaps to racial biology on account of his anti-Semitism) and to any kind of science, be it psychology, sociology or any other scientific explanation of human being as such. Hence, his notion of being-in-the-world is also quasi-theological and is not like Husserl's more naturalistic life-world. Thus Thompson is quite right in rejecting Dreyfus' Heideggerian criticism of Husserl.

However, Thompson himself interprets Husserl in the light of Merleau-Ponty and that is where potential problems arise. Thompson quotes from Merleau-Ponty's masterwork *Phenomenology of Perception*: "But I am not in front of my body, I am in my body, I am my body", and comments on this that "one's self is not merely *embodied* but *bodily*", which he goes on to explain as follows:

88 Ibid, 379.

> Merleau-Ponty also refuses to understand the proposition "I am my body" in a materialist way, as meaning that I am (or my self is) nothing more than a complex physical *object.* Instead, he maintains the original position that I am a bodily subject.[89]

Thus, according to Thompson, Merleau-Ponty identifies the "I" or the bodily subject not with *Körper,* but with *Leib* in Husserl's terminology. But this, too, is an identity thesis. It seems that Thompson, having rejected the Materialist identity thesis, like Francis Crick's "I am my neurons", comes close to following Merleau-Ponty into a phenomenological identity thesis, "I am my body".

But no such identity thesis can work, for the "I" or Subject is always more than anything it can be identified with and goes beyond any such limiting specifications. In brief, it cannot be identified with *Leib* anymore than with *Körper.* Though, it is also true that it cannot be dissociated from both, from body in the two sense of that word. As we have argued, it is an emergent entity which mandates an inseparable non-identity with its body. Hence, to any assertion that "I am my body" it is necessary to add "I am also more than my body", for subjective being or the Ego extends beyond the body. It reaches out to what is beyond the body, to relations with others, language and culture, and in a sense incorporates and encompasses these within itself. Its nature and constitution is determined by this extension beyond the body. "I am you" is as much a truth as "I am my body".

However, despite his uncritical espousal of Merleau-Ponty, in the last resort it does not seem as if Thompson does fall in with a Phenomenological identity thesis for he also recognizes what we have stated above about the constitution of subjectivity. As he puts it in the very last sentences of his book:

> The individual human subject is the encultured bodily subject. In this way, the knowing and feeling subject is not the brain in the head, or even the brain plus the body, but the socially and culturally situated person, the encultured human being.[90]

On this view, the "culturally situated person" is more than just the body in the sense of both *Körper* and *Leib*, which does not chime in with his earlier Merleau-Ponty inspired assertion "I am my body". There is a noticeable disharmony between much of the earlier biologically and phenomenologically oriented treatment of the body and his idea of a body-body identity thesis and what is

89 Ibid, 245.
90 Ibid, 411.

presented concerning subjectivity in the last chapter entitled "Empathy and Enculturation". So much so, that the last chapter can be read as a correction, even if not a contradiction, to the earlier too narrow and exclusive affirmation of bodily being and lack of any specification of all that constitutes the mind that lies outside the body.

Another problem that arises out of Thompson's too narrowly biological and phenomenological orientation on the body is that it leaves no way of distinguishing between animal and human minds. How are we to differentiate between those biological and/or phenomenological descriptions that apply only to humans from those that apply to animals or to humans and animals at once? Thompson does not raise this issue. In fact, he provides no explicit account of animal minds and hence the fundamental disparity between animal and human minds is not discussed. It is only incidentally referred to in passing in the last chapter on empathy in apes and humans.[91] What Thompson does say prior to his last chapter obfuscates such crucial differences for phenomenological descriptions do not allow of any such separations.

In fact, Thompson's discussion of Merleau-Ponty's phenomenological account of being-in-the-world is a clear instance of such an obfuscation by not distinguishing between human and animal action. Thompson presents Merleau-Ponty's view in the following terms:

> To belong to the world in this way means that our primary way of relating to things is neither purely sensory and reflexive, nor cognitive and intellectual, but rather bodily and skilful. Merleau-Ponty calls this kind of bodily intentionality "motor intentionality". His example is grasping or intentionally taking hold of an object… In picking up a teacup to drink from, I identify it not by its location in space but by its egocentric relation to my hands, and I grasp it in the light of the goal of sipping from it.[92]

Most of this account of grasping a cup applies just as well to an ape, particularly a chimpanzee, as to a human being. Indeed, it is difficult not to credit such a creature with most of what emerges from a phenomenological account of perception and action in general. Where the human differentiates itself from the ape is not in "motor intentionality", or Merleau-Ponty's "bodily and skilful" attributes, but precisely in manners, which is a cultural and social and not a phenomenological matter. What accounts for the difference does not lie in what Thompson calls Gibsonian "affordances", which, as we showed in

91 Ibid, 396.
92 Ibid, 247.

Chapter 3, mostly apply indifferently to animals as well as humans, but rather in the specifically human capacities; as in this case, not just to grasp the cup physically and drink from it, but to grasp mentally and understand what is the polite or culturally appropriate way of doing so. This is where the human handling of a teacup distinguishes itself from the grasping of an ape. Managing it in the human way goes way beyond "motor intentionality" and cannot be conveyed in phenomenological terms but only in those of cultural sociology. This is where the practice of *epoché* is such a debilitating handicap in phenomenological descriptions, for by excluding all knowledge of cultural context and situation and concentrating on the experience alone, or what is "given", it ignores the whole human dimension of the act; that which is studied in those human sciences that go beyond Phenomenology to the understanding of social and cultural meanings that are historically conditioned and determined. Thompson is not alert to these inherent shortcomings of Phenomenology which is, according to Husserl, a *strenge Wissenschaft*, or pure science of *Wesensschau,* "essential vision", and that practices *epoché* in the Husserlian sense of excluding all else not immediately given.

Nevertheless, despite the obvious differences between Thompson's largely biological-phenomenological treatment of body and the more generally scientific account of mind presented in this book, there is a large commonality of interest in the two approaches. In Chapter 5 we have already indicated how close Thompson comes to our view of emergence as non-causal, but despite this allows for downward and upward causation as a matter of practical necessity. And even beyond that convergence on emergence, Thompson's account of mind matches closely our tripartite architectonic of mind/brain systems; and incidentally, also Aristotle's three psyches, to which he explicitly refers.[93] Thompson presents his three-fold division as follows:

> One of the guiding ideas of this book is that the human mind is embodied in our entire organism and in the world. Our mental lives involve three permanent and intertwined modes of bodily activity – self-regulation, sensorimotor coupling, and intersubjective interaction. Self-regulation is essential to being alive and sentient. It is evident in emotion and feeling, and in conditions such as being awake or asleep, alert or fatigued, hungry or satiated. Sensorimotor coupling with the world is expressed in perception, emotion and action. Intersubjective interaction is the cognition and affectively charged experience of self and other. The human brain is

93 Ibid, 226.

> crucial for the three modes of activity, but it is also reciprocally shaped and structured by them at multiple levels throughout the lifespan.[94]

Thompson's three types of bodily activity more or less correspond to the three forms of mind/brain systems that are hierarchically nested in the human mind and which we called the horizontal axis of diachronic emergence (see Chapter 4 Section II). But, unfortunately, there is nothing in Thompson even to approximate to what we have called the vertical axis in the structure of mind, the three levels of the conscious, unconscious and nonconscious, which involves synchronic emergence. In fact, there is no reference to the unconscious whatever, for this whole Freudian dimension is missing. But it is hardly possible for Phenomenologists to admit a Freudian side to consciousness as that escapes their *Wesensschau* or essential vision on which their approach is based. In this respect they remain Cartesians.

However, as a critique of Cartesian Materialism and the MIT-Mind, Thompson's work is as good as anything in the philosophical literature. He puts paid to the whole idea of materialization and dematerialization as propounded by Minsky, Chomsky and their acolytes which culminates in the delusion of informational immortality. There is also much to be praised in the work of Varela, with whom Thompson closely collaborated, as well as the other exponents of the Naturalizing Phenomenology movement despite its shortcomings. What they bring out is that the mind is embodied in a body, a biological organism, and this body is embedded in an environment; but what they do not sufficiently emphasize is that the mind of human beings transcends the mere physical world of objects and extends into the human world of society, language and culture.

Embodiment and embedment are the two crucial notions that separate our account of mind from the MIT-Mind. The mind is embodied in the body and embedded in the extra-bodily environment both physical and cultural. In the first place, it is embodied in the brain, the essential organ without which it could not exist for it contains the neural processes to which the mind is indissolubly linked but with which it cannot be identified, as Materialists are wont to do. And these processes cannot be described in information processing terms on analogy with computers, as the Cognitivists have postulated. How the brain functions is gradually being disclosed by neurologists such as Edelman, Damasio, Crick, Tononi and Koch. But what they have discovered so far goes far short of explaining how mind emerges from such neural activities, which remains still largely unknown. But it is not an inscrutable mystery beyond human

94 Ibid, 243.

comprehension since more and more is known about it as research continues. Whether that research will continue to generate ever new knowledge for all time or whether at some point in the future it will cease to do so and come to an end is unpredictable. Whether everything worth knowing about the mind will ever be known is itself unknown and unknowable. If there is anything that might be considered mysterious, this is the only mystery that remains.

Postlude

When Heidegger declared that Cybernetics is the metaphysics of the twentieth century, it was one of the few occasions when he was proved right. For soon thereafter what started off as mere Cybernetics turned into the full panoply of what we have dubbed the MIT-Mind. It was hailed as a Cognitivist Revolution and it came to dominate the academic establishments of America; then it spread by stages to the rest of the world, even to its ideological enemies in the Soviet Union. This mechanistic mentality began in a handful of "hard" formalistic sciences but then gradually fanned out to engulf most of the rest of the academic environment. Information passed into common currency, the coin of the realm in which everyone dealt, its concepts pervading just about all disciplines and even invading ordinary speech: communication, message, noise, redundancy, feedback, bits, coding, data, memory, program and many more became watchwords to be invoked at every opportunity. This is the language of the new metaphysics.

This turning that mankind took since the end of the Second World War was soon being hailed as a new world of Information technology and the onset of an unprecedented technological civilization. It manifested itself intellectually in the MIT-Mind, and Cartesian Materialism was its reigning philosophy. This whole development more or less coincided with the Cold War and the two played into each other, as we have already seen in the disbursement of funding for military purposes to the sciences of Information. The end of the Cold War did somewhat lessen the flow of funds, but this did not bring about any serious disruption to the dominance of the MIT-Mind, for by then it was already well entrenched both in academia and the media. Nevertheless, this eventuality saw the arrival of the first challenges to it; it had begun to falter though not yet to fall.

The challenge the MIT-Mind presents now is still overwhelming since it represents the dominant forces determining our culture, those of a global technological civilization, so called. Hence, it will take so much more than philosophical critique to counter it, and we cannot even envisage when it might be overcome. Nevertheless, critique is a start in mounting its overthrow by exposing its evident intellectual weaknesses and making its certitudes questionable. The struggle against it will doubtless continue into the indefinite future for every new development in Information technology lends it strength and confidence, such as at present the internet, or in the near future the likelihood of quantum computers, the prospect that so delights the artificial intelligentsia.

Yet even while the MIT-Mind and Cartesian Materialism still prevail, it is essential to offer a counter to it, not merely in the form of criticism but also by presenting an alternative conception of mind more in keeping with our current developments in science where emergence is now a leading preoccupation. This is the reason that we began our studies of mind with an investigation of the problem of origin in matter, life and mind; beginning with animal mind which is the evolutionary origin of the human mind. Questions of origin are, of course, issues of emergence, and emergence has been our preeminent topic throughout the work.

Emergence is generally of fundamental concern throughout the sciences, for even as the sciences of reduction are reaching their close, those of emergence are taking off. The emergence of mind from matter or the quintessence of dust has been our dominant theme. We have sought to show that just as in Descartes' time, when a new conception of matter as inert and mechanical led to a new conception of mind as immaterial and subjective, so, too, in our time a reconception of matter as active and self-organizing might be leading to a view of mind as emergent and self-activating. In this quest towards a new concept of mind science and philosophy work hand in hand together. Whether the specific way we have theoretically conveyed this relation in our account of mind is accepted or not is of lesser importance than the overall approach we have initiated in relating science and philosophy.

In briefest possible outline, our conclusion has been that mind is emergence raised to the fifth power or, metaphorically expressed, the quintessence of dust. We went through the various powers of emergence which we can symbolically summarize as follows: E^1 stands for the emergence of matter as such, E^2 for solid-state matter, E^3 for life, E^4 for animal mind and E^5 for human mind. All this was presented in the first chapter. In the second chapter we went on to argue that there can be no general theory of emergence covering all cases at once, for each of these higher powers of emergence is *sui generis*; each is overseen by autonomous sciences which vary in method, procedure, instrumentation and types of experiments. These sciences are conventionally known as elementary particle physics, solid-state physics, chemistry, biology, psychology, sociology and so on.

In all seriousness, we have even entertained Donald's premonition that humanity is now entering a still higher level of emergence, that of a post-human mind that would constitute E^6. On this assumption, the individual human mind would be absorbed into an all-encompassing global technological system in which super-computers or other highly intelligent systems of some kind would be linked up to minds in a new kind of compound entity, a no-longer-human Man-machine symbiosis.[1] If this is where evolution is heading, then it would obviously explain and make sense of the many anti-humanistic developments of our time, including the rise of the MIT-Mind as the ideological herald of this brave new world. However, it is a prospect we must resist if we wish to survive and remain human.

It is still possible to hope that this idea of a sextessence of dust is only a misanthropic assumption, not a scientific prediction, for there is no science of futurological emergence. However, even though there is no unified science of emergence, there is a logic of emergence, that of indissoluble non-identity. This logic applies to all cases of emergence from the emergence of water out of atoms of hydrogen and oxygen to the emergence of life out of a much more complex chemistry, and so on to the eventual emergence of mind through the process of evolution. In the case of mind, the non-identity condition stipulates that no Monist identitarian thesis, such as postulated by Materialists in one way or by Idealists or Phenomenologists in another way, is possible. The principle of indissolubility means that no Dualist account, in any version that can allow for the separation of mind from its physical basis in the brain, body and environment, is possible. On an Emergentist thesis, this relation between mind and matter can be studied according to the three temporal forms that emergence can take: synchronic and diachronic, where the latter divides into phylogenetic and ontogenetic. We have in fact proceeded to do so by examining the evolutionary sequence from animals without minds to animal minds, finally culminating in the human mind.

The philosophy of mind studied in terms of the sciences in this Emergentist way exemplifies a new relation between science and philosophy. It presupposes the abrogation of the old Positivist relation based on reduction, especially that prevalent since the Second World War when the new mechanistic sciences starting with Cybernetics dictated what should be the philosophy of mind. According to that philosophy, mind is based on what Lackoff and Johnson call "first generation cognitive science", which they sum up as a "blend of generative linguistics, information processing psychology, analytical philosophy of language and artificial intelligence".[2] This was in fact the main source of Cartesian Materialism and the MIT-Mind. It dominated American and British philosophy in the second half of the twentieth century. Only since then have

1 See also Katherine Hayles, *How We Became Post-Human*, op.cit.

2 Mark Johnson, *Embodied Mind, Meaning and Reason, How Our Bodies Give Rise to Understanding* (Chicago: University of Chicago Press, 2017), 16.

counter-currents to it in the English-speaking world become evident, but they have by no means prevailed for this approach is still firmly entrenched.

One significant move against it has been launched by Lackoff and Johnson in terms of what they call "second generation cognitive science". Their credentials are eminently appropriate for this task. Lackoff is an ex-student of Chomsky who was expelled by the master for daring to question the dogmatic assumptions of Chomsky's Cartesian linguistics; and, indeed, we have previously quoted from his devastating critique of these dogmas. Johnson is an exponent of Pragmatism, particularly that of Dewey, but he seeks to go beyond Pragmatism into an exploration of meaning that is deeper even than language. He bases himself on the teachings of Paul Ricoeur, of whom he writes admiringly:

> It was not till I took courses from Paul Ricoeur – one on metaphor, a second on hermeneutics, and a third on imagination – that I began to see that there was an experience of meaning and value that went deeper than language. I learned to see the entire hermeneutic (interpretative) process of understanding not merely as an intellectual and linguistic act, but as constituting our whole embodied way of being in, of making sense of our world.[3]

Metaphor, hermeneutics and imagination are key themes in this book for which metaphorically speaking Hamlet stands; and Hamlet in turn stands in for Shakespeare whose hidden presence is invisibly there throughout, a so-called absent-presence in the modish jargon. Shakespeare represents the highest excellence of art; and art or aesthetics even at its very lowest is the attainment that most distinguishes the human from the animal mind, as we saw by reference to Gibson on perception and Donald in respect of mimesis, play and acting in general. Metaphor, hermeneutics and imagination are also key themes for Thompson whose work is close to that of Johnson. Johnson invokes them to serve as a critique of the current approaches in the humanities and social sciences. As he puts it, "metaphor, meaning and thought… could not be adequately accounted for by the reigning philosophical and linguistic traditions of the day".[4]

In philosophy, Johnson is decidedly averse to the whole "turning to language" in the British and American traditions of Analytic philosophy. He is critical of the both the Fregean formalists and the neo-Wittgensteinians, such

3 Ibid, 13.
4 Ibid, 8.

as Davidson, Searle and Rorty, "who conceives all philosophy as linguistic analysis and, where appropriate, criticism of our language games and linguistic practices."[5] In opposition to this linguistic establishment, he advocates four major alternative directions in which philosophy and the social sciences might go instead:

1. the rebirth of interest in pragmatic views of experience, meaning and value (i.e. present-day pragmatics);
2. the phenomenology of the embodied mind, especially in the style of Merleau-Ponty and, to a lesser extent, parts of Heidegger and Husserl that focus on the lifeworld;
3. the second-generation cognitive science, which pursues empirical studies of embodied cognition (in psychology, neuroscience, linguistics and anthropology);
4. ecological philosophies that emphasize organism-environment processes of meaning-making and that acknowledge the human connection to other animal species and the more than human world.[6]

Johnson accompanies his four point declaration of principle with the names of mainly American exponents of these tendencies which we have omitted so as not to incur charges of making invidious comparisons. He concludes with the statement that "currently we are just at the dawn of what might someday become a serious reconstruction of philosophy, but nobody can be sure where it will ultimately lead". We agree with this, without necessarily having to subscribe to all of his four points or the way he develops them. But there is enough commonality between what he is proposing in his work in general and what we have outlined in this and other books to make it evident that we are embarking on the same quest of a "serious reconstruction of philosophy". And as Johnson and Lackoff are well aware, this will also involve a thorough going attempt at a reconstruction of what is collectively known in German as *Geisteswissenschaft*, including what we call the humanities and social sciences or the sciences of Man.

Any attempt to understand "what a piece of work is Man" involves a consideration of the sciences of Man. But even cursory review of these, which we have so far not attempted, will quickly reveal that these are at present in a most perilous predicament. The threat to their continued existence as anything other than humdrum utilitarian and pragmatic disciplines with no serious intellectual content, sheer plodding scholarship, or mere abstract and abstruse jargon, comes from many directions. We can categorize these as two types:

5 Ibid, 8.
6 Mark Johnson, *The Meaning of the Body,* op. cit, 264.

outside pressures stemming from the natural sciences and the tendencies we have identified with the MIT-Mind; inside pressures stemming from avant-gardist deconstructive interpretative practices deriving from nihilistic Continental philosophies. Caught between these two forces, *Geisteswissenschaft* is squeezed as in a vice and is slowly being crushed to death.

The outside pressures are those which we have studied in this work, they are what Lackoff and Johnson have called "first generation cognitive science". This is the now long-standing attempt to formulate *Geisteswissenschaft* in scientistic terms that valorise formalization, quantification, information processing, technification, data analysis and other forms of blinkered empiricism. These latest attempts to turn the social sciences into strict sciences that can take their place besides the natural sciences only build on previous such designs, stemming from Positivism and other method driven philosophies of science.

If such outside pressures can be characterized as madness in method, their contrary inside complements might be termed as method in madness. For the madness of the pseudo-philosophy that now prevails within the social sciences and humanities is not without its method. It is known as "cultural theory" in which there is very little real culture and even less real theory: mass media production passes for culture and so-called French Theory for theory. The latter is generally an abstruse and arcane interpretative quasi-metaphysical jargon whose main function is to mystify and purvey an aura of exclusivity around those who pretend to understand it so as to overawe all the others. Cultural studies are usually replete with "cultural theory" in this sense.

When such "studies", of which there are very many varieties, are coupled with "political correctness" movements that are neither political nor correct in any serious sense, but merely stand for identity politics, then it becomes evident that academic interests are at play and no real intellectual concerns are any longer involved. We enter the realm of academic politics where the securing of publications and positions is all that matters and genuine work can no longer be distinguished from that which is fraudulent. Numerous tests carried out in placing bogus papers in "theory" journals have by now confirmed this conclusion "experimentally", so to speak.

What happens in academia is only a reflection of what takes place in society in general. Culture is now what can be called "global culture" which no longer has anything to do with *Geisteswissenschaft* or with any kind of truth or value. Books with titles such as "the death of truth" are now appearing which amply document these developments even at the popular level.[7] Global culture is

7 Michiko Kakutani, *The Death of Truth* (London: Collins, 2018).

largely a commercial product of a Culture Industry that subserves the interests of global capitalism. It is a mode of production of aesthetic wares based on advanced technologies, usually employing computers and other information media. These interlinked technologies now span the globe and enmesh almost everyone in their networks, for there is no way of escaping them. Not that there are very many left who wish to do so for by far the greatest proportion of mankind are only too happy to participate and be ensnared. The phenomenal success of Facebook and other social media attests to this fact.

We are now confronted with a wholesale change in human mentality that in our time has been promoted by the MIT-Mind and the "first generation cognitive sciences". Metaphorically expressed, this amounts to something akin to a cultural lobotomy in the brain of Man, where the faculties conventionally associated with the left hemisphere of the brain have been severed from those located in the right hemisphere. As a result, those of the left have grown unchecked by those of the right and assumed hypertrophic proportions. Those of the right have been repressed and stunted; they display features of dwarfism and arrested development.

In this work we have sought to heal this split and restore a better balance to these purported opposed halves of the brain, so to speak, and thereby we have aimed to recover something of the holistic nature of the mind. To do so we have stressed those aspects that run counter to those which the MIT-Mind emphasizes, and Hamlet has been our guiding light and prime symbol of these. We are not alone in this endeavour as the list of authors we have relied on makes evident.

Hence, as in the final credits screened at the end of a film, we proceed to list the cast of actors who have played the crucial roles in our play. Perhaps pride of place should be given to the scientists; first the neuroscientists and anthropologists, Damasio, Edelman, Donald and Deacon; then the biochemists and biologists, Varela, Atlan, Cohen and Stewart; and finally the psychologists, Gibson and Bruner. We have also referred to a number of physicists sympathetic to our enterprise, among them Ellis, Anderson, Darrigol, Laughlin and Pines. Last but not least, though we have disagreed with the prognostications of the future of science presented by Stent and Horgan, we have learned much from their exposition of where science stands at present; and analogously, due credit must also be given to Boden whose voluminous compilation made it so much easier to see what one was up against. On the philosophical side, there have been fewer to call on, but Thompson, Johnson and Humphreys stand out. Among the philosophers of the great tradition, Spinoza has been our cynosure, though we have endeavoured to give due credit to Descartes as the founder of modern science and philosophy.

However, the side of the mind we have sought to highlight is represented neither by Spinoza nor Descartes, but by Hamlet. References to *Hamlet* enable us to maintain continuities with our Western tradition as a whole, with Renaissance humanism, with the humanism of the Roman authors and Greek philosophers, above all Aristotle but going right back to Protagoras' dictum that Man is the measure of things. Hence, it is only fitting that we should conclude as we began with Hamlet's peroration on Man as the quintessence of dust:

> What a piece of work is a man, how infinite in faculties, in form and moving, how express and admirable in action, how like an angel in apprehension, how like a god! The beauty of the world; the paragon of animals; yet to me what is this quintessence of dust?

Bibliography

Anderson, Philip Warren. "More is different: Broken Symmetry in Nature and the Hierarchical Structure of Science". *Science* vol. 177, no. 4047 (1972): 393–6.

Aristotle. *De Anima*, Book II and Book III. In *Introduction to Aristotle*, edited by Richard Mc Keon and translated by J.A. Smith. New York: The Modern Library, 1947.

Aristotle. *Metaphysics*, Book I, Chapter 4. In *Introduction to Aristotle*, edited by Richard Mc Keon and translated by J.A. Smith. New York: The Modern Library, 1947.

Atlan, Henri. *Enlightenment to Enlightenment: Intercritique of Science and Myth*. Translated by Lenn J. Schramm. Albany: State University of New York Press, 1993.

Atlan, Henri. *The Sparks of Randomness*, vol. 1, *Spermatic Knowledge*. Translated by Lenn J. Schramm. Stanford: Stanford University Press, 2011.

Bargh, John A. "Bypassing the Will: Toward Demystifying the Non-conscious Control of Social Behaviour". In *The New Unconscious*, edited by Ran R. Hassin, James S. Uleman and John A. Bargh. New York: Oxford University Press, 2006: Chapter 2.

Bar-Yosef, Ofer. "PPNB Interaction Sphere". Review of J. Cauvin, In *Cambridge Archaeological Journal*, 11: (2001): 114–7.

Barkow, Jerome, Leda Cosmides and John Tooby (eds.). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press, 1982.

Bateson, Gregory. *Steps to an Ecology of Mind*: *A Revolutionary Approach to Man's Understanding of Himself.* New York: Ballantine Books, 1972.

Bedau, Mark A. and Paul Humphreys, eds. *Emergence: Contemporary Readings in Philosophy and Science*. Cambridge, Mass.: MIT Press, 2008.

Boden, Margaret. *Mind as Machine*: *A History of Cognitive Science*. Oxford: Oxford University Press, 2006.

Boden, Margaret. "Promise and Achievement in Cognitive Science". In *The Future of the Cognitive Revolution*, edited by David Martel Jackson and Christine E. Erneling. New York: Oxford University Press, 1997: 55–67.

Bruner, J.S. *In Search of Mind: Essays in Autobiography*. New York: Harper and Row, 1988.

Bruner, J.S. *Acts of Meaning*. Cambridge, Mass.: Harvard University Press, 1990.

Bruner, Jerome. "Will Cognitive Revolutions ever stop?" In *The Future of the Cognitive Revolution*, edited by David Martel Jackson and Christine E. Erneling. New York: Oxford University Press, 1997: 279–92.

Buchdahl, Gerd. *The Image of Newton and Locke in the Age of Reason*. London: Sheed and Ward, 1961.

Buchdahl, Gerd. *Metaphysics and The Philosophy of Science: The Classical Origins: Descartes to Kant*. Cambridge, Mass.: The MIT Press, 1969.

Caton, Hiram. *The Origin of Subjectivity: An Essay on The Discourses*. New Haven: Yale University Press, 1973.

Cauvin, Jacques. *The Birth of Gods and the Origins of Agriculture*. Cambridge: Cambridge University Press, 2001.

Chalmers, David. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press, 1996.

Changeux, Jean-Pierre. *Neuronal Man: A Biology of Mind*. Translated by Lawrence Garey. New York: Pantheon Books, 1985.

Churchland, P.S. *Brain-Wise: Studies in Neurophilosophy*. Cambridge, Mass.: MIT Press, 2002.

Chomsky, Noam. *Cartesian Linguistics.* New York: Harper and Row, 1966.

Chomsky, Noam. *The Logical Structure of Linguistic Theory*. New York: Plenum Press, 1975.

Chomsky, Noam. *Language and Responsibility*. Based on Conversations with Mitsou Ronat. Translated by John Viertel. New York: Pantheon Books, 1979.

Clayton, Philip and Paul Davies, eds. *The Re-emergence of Emergence*. Oxford: Oxford University Press, 2006.

Cohen, Jack and Ian Stewart. *The Collapse of Chaos: Discovering Simplicity in a Complex World.* New York: Penguin, 1940.

Crutchfield, James P. "Is Anything ever New? Considering Emergence". In *Emergence: Contemporary Readings in Philosophy and Science*, edited by Mark Bedau and Paul Humphreys. Cambridge, Mass.: MIT Press, 2008, Chapter 15.

Damasio, Antonio. *Descartes' Error: Emotion, Reason and the Human Brain.* New York: Harper Collins, 2000.

Damasio, Antonio. *Looking for Spinoza*: *Joy, Sorrow and the Feeling Brain*. Orlando, Florida: Harcourt, 2003.

Darrigol, Olivier. *The Worlds of Flow: A History of Hydrodynamics from the Bernoullis to Prandtl.* Oxford: Oxford University Press, 2005.

Davis, Philip J. and Reuben Hersh. *Descartes' Dream*: *The World According to Mathematics*. Boston: Houghton Mifflin, 1973.

Dawkins, Richard. *The Ancestor's Tale: A Pilgrimage to the Dawn of Evolution.* Boston: Houghton Mifflin, 2004.

Deacon, Terence W. *The Symbolic Species: the Co-Evolution of Language and the Brain*. New York: Norton, 1997.

Deacon, Terence W. "Emergence; The Hole at the Wheel's Hub". In *The Re-emergence of Emergence*, edited by Philip Clayton and Paul Davies. Oxford: Oxford University Press, 2006: 111–50.

Deacon, Terence. *Incomplete Nature: How the Mind Emerged from Matter*. New York: Norton, 2012.

Dennett, D.C. and D.R. Hofstadter (eds.). *The Mind's I.* New York: Basic Books, 1981.

Dennett, Daniel. *Consciousness Explained*. Boston: Little Brown, 1991.

Donald, Merlin. *The Origins of the Modern Mind*: *Three Stages in the Evolution of Culture and Cognition*. Cambridge, Mass.: Harvard University Press, 1991.

Donald, Merlin. "Human Cognitive Evolution: What we Were, What we are Becoming." Social Research, vol.60, no.1 (Spring 1993): 143–70.

Donald, Merlin. "Hominid Enculturation and Cognitive Evolution." In *Cognition and Material Culture*: *The Archaeology of Symbols*, edited by Colin Renfrew and Chris Scarre. Cambridge: McDonald Institute Monographs, 1998: 7–18.

Donald, Merlin. "Preconditions for the Evolution of Protolanguages". In *The Descent of Man: Psychological Perspectives on Human Evolution* by Michael Corballis and Stephen E.G. Lea. Oxford: Oxford University Press, 1999, Chapter 8. DOI:10.1093/acprof:oso/9780192632593.003.0008.

Donald, Merlin. *A Mind So Rare: The Evolution of Human Consciousness*. New York: Norton, 2001.

Donald, Merlin. "Art and Cognitive Evolution". In *The Artful Mind and the Riddle of Human Creativity*, edited by Mark Turner. New York: Oxford University Press, 2006: 3–20.

Donald, Merlin. "The Roots of Art and Religion in Ancient Material Culture". In *Becoming Human, Innovation in Prehistoric Material and Spiritual Culture*, edited by Colin Renfrew and Iain Morley. Cambridge: Cambridge University Press, 2009: 95–103.

Donald, Merlin. "Consciousness and the Freedom to Act". In *Free Will and Consciousness*: *How Might they Work?*, edited by Roy F. Baumeister, Alfred R. Mele and Kathleen D. Vohs. Oxford: Oxford University Press, 2010: 8–23.

Donald, Merlin. "Evolutionary Origins of Autobiographical Memory: a Retrieval Hypothesis". In *Understanding Autobiographical Memory: Theories and Approaches*, edited by Dorthe Berntsen and David C. Rubin. Cambridge: Cambridge University Press, 2012: 269–88.

Donald, Merlin. "Mimesis Theory Re-examined, Twenty Years after the Fact". In *Evolution of Mind, Brain and Culture*, edited by Gary Hatfield and Holly Pittman. Chicago: Chicago University Press, 2013: 169–92.

Donald, Merlin. "Key Cognitive Preconditions for the Evolution of Language". *Psychonomic Bulletin and Review*, Vol. 24, No. 1 (February 2017): 204–8.

Donald, Merlin. "Self-Programming and Self-Domestication of the Human Species: Are we approaching a Fourth Transition?" In *Evolution, Cognition and the History of Religion: A New Synthesis,* edited by Anders Klostergaard Petersen, I.S. Gilhus, L.A. Martin, J.S. Jensen and J. Sorensen. Leiden: Brill, 2018: 159–70.

Dupuy, Jean-Pierre, *The Mechanization of the Mind: The Origins of Cognitive Science*. Translated by M.B. de Bevoise, Princeton: Princeton University Press, 1999.

Eccles, J.C. and K.R. Popper. *The Self and its Brain*. Berlin: Springer, 1977.

Edelman, Gerald M. *Neural Darwinism*: *The Theory of Neuronal Group Selection*. New York: Basic Books, 1987.

Edelman, Gerald M. *The Remembered Present: A Biological Theory of Consciousness*, New York: Basic Books, 1989.

Edelman, Gerald M. *Bright Air, Brilliant Fire.* New York: Basic Books, 1992.

Edelman, Gerald and Giulio Tononi. *A Universe of Consciousness: How Matter Becomes Imagination.* New York: Basic Books, 2000.

Edelman, Gerald. *Wider than the Sky: A Revolutionary View of Consciousness*. New York: Penguin, 2004.

Ellis, George F.R. "On the Nature of Emergent Reality". In *The Re-emergence of Emergence*, edited by Philip Clayton and Paul Davies. Oxford: Oxford University Press, 2006: 79–110.

Ellis, George F.R. "Issues in the Philosophy of Cosmology". In *Handbook of the Philosophy of Physics*, edited by J. Butterfield and J. Earman. Amsterdam: Elsevier, 2007: 1183–286.

Einstein, Albert. *Ideas and Opinions*. New York: Crown, 1954.

Erneling, Christina. "Challenges to Cognitive Science: The Cultural Approach". In *The Future of the Cognitive Revolution*, edited by D.M. Johnson and C.E. Erneling. Oxford: Oxford University Press, 1997: 275–8.

Erneling, Christina. "Cognitive Science and the Study of Languages". In *The Future of the Cognitive Revolution*, edited by D.M. Johnson and C.E. Erneling. Oxford: Oxford University Press, 1997: 115–7.

Fodor, J.A. *The Modularity of Mind*: *An Essay in Faculty Psychology*. Cambridge, Mass.: MIT Press, 1983.

Freud, Sigmund. *Beyond the Pleasure Principle*. Translated by James Strachey. New York: Norton, 1961.

Frith, Chris D. *Making up the Mind: How the Brain Creates our Mental World.* Oxford: Blackwell, 2007.

Frith, Chris D. "Social Cognition". In *The Sapient Mind: Archaeology meets Neuroscience*, edited by Colin Renfrew, Chris Frith and Lambros Malafouris. Oxford: Oxford University Press, 2009: 177–86.

Gaukroger, Stephen. *Descartes: An Intellectual Biography*. Oxford: Clarendon Press, 1995.

Gardner, Howard. *The Mind's New Science*: *A History of the Cognitive Revolution*. New York: Harper Collins, 1985.

Gelder, Tim van. "Wooden Iron? Husserlian Phenomenology Meets Cognitive Science". In *Naturalizing Phenomenology*, edited by Jean Petitot, Francisco J. Varela and Bernard Pachoud. Stanford: Stanford University Press, 1999: 245–65.

Gibson, J.J. “What is Form?” *Psychological Review* 58 (1951): 405–13. Also in *Reasons for Realism: Selected essays of J.J. Gibson,* edited by Edward Reed and Rebecca Jones. Hillsdale, N.J.: Lawrence Erlbaum, 1982: Chapter 4.1.

Gibson, J.J. “Information Contained in Light”; “Ecological Optics”; “On Theories for Visual Space Perception”; “Direct Perception and Indirect Apprehension”; “Affordances”. All in *Reasons for Realism: Selected Essays of J.J. Gibson*, edited by Edward Reed and Rebecca Jones. Hillsdale, N.J.: Lawrence Erlbaum, 1982.

Glaser, Jack and John F. Kihlstrom. “Compensatory, Automaticity: Unconscious Volition is not an Oxymoron”. In *The New Unconscious*, edited by Ran R. Hassin, James S. Uleman and John A. Bargh. New York: Oxford University Press, 2005: Chapter 7.

Glick, James. *Chaos: Making a New Science*. New York: Viking, 1987.

Gorman, James. “The Brain’s Inner Language”. *The New York Times*, Feb. 24, 2014: 86.

Gould, S.J. *Ontogeny and Phylogeny*. Cambridge: Cambridge University Press, 1977.

Gregory, Richard L. “Knowledge in Perception and Illusion”, *Philosophical Transactions of the Royal Society* 352 (1997): 1121–8.

Griffin, Donald. *Animal Minds*. Chicago: Chicago University Press, 2012.

Haken, Hermann. *Information and Self-Organization: A Macroscopic Approach to Complex Systems*. New York: Springer, 2000.

Hayles, Katherine. *How we Became Post-Human*. Chicago: Chicago University Press, 1999.

Hassin, Ran R., James S. Uleman and John A. Bargh (eds.): *The New Unconscious*. New York: Oxford University Press, 2005.

Hegel, G.W.F. *Philosophy of Mind: Part Three of the Encyclopedia of Philosophical Sciences*. Translated by A.V. Miller. Oxford: The Clarendon Press, 1971.

Heims, Steve Joshua. *The Cybernetic Group*. Cambridge, Mass.: MIT Press, 1991.

Hobbes, Thomas. *Leviathan; or the Matter, Forme of a Commonwealth Ecclesiastical and Civil*, edited by A.D. Lindsay. London: Dent, 1914.

Horgan, John. *The End of Science*: *Facing the Limits of Knowledge in the Twilight of the Scientific Age.* New York: Basic Books, 1996.

Horgan, John. *The Undiscovered Mind.* New York: Broadway Books, 1999.

Horgan, John. “The Final Frontier? Are we Reaching the Limits to Science?” *Discover Magazine*, Sep.22, 2006.

Horgan, John. “Interview with George F.R. Ellis” in *Scientific American*, July 22, 2014.

Huizinga, John. *Homo Ludens*: *A Study of the Play Element in Culture*. Anonymous translation. London: Routledge and Kegan Paul, 1949.

Humphreys, Paul. *Emergence: A Philosophical Account*. New York: Oxford University Press, 2016.

Husserl, Edmund. *Cartesian Meditations: An Introduction to Phenomenology*. Translated by Darion Cairns. The Hague: Nijhoff, 1969.

Impey, Chris. *How it began: A Time Traveller's Guide to the Universe*. New York: Norton, 2012.

Johansson, Gunnar. "A Letter to Gibson". *Scandinavian Journal of Psychology*, vol. 2 (Sep. 1970), 67–77. https://doi.org/10.1111/j.1467-9450.1970.tb00719.x.

Johnson, Mark. *The Meaning of the Body: Aesthetics of Human Understanding*. Chicago: Chicago University Press, 2007.

Johnson, Mark. *Embodied Mind, Meaning and Reason: How our Bodies Give Rise to Understanding*. Chicago: University of Chicago Press, 2017.

Joos, Erich. "The Emergence of Classicality from Quantum Theory". In *The Re-emergence of Emergence*, edited by Philip Clayton and Paul Davies. Oxford: Oxford University Press, 2006: 53–78.

Kaplan, Sarah. *The Washington Post*. Reprinted in *The Age*, Melbourne, 26 June, 2018.

Kakutani, Michiko. *The Death of Truth*. London: Collins, 2018.

Karmiloff-Smith, Annette. *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, Mass.: MIT Press, 1992.

Kauffman, Stuart. *At Home in the Universe: A Search for the Laws of Self-organization and Complexity*. New York and Oxford: Oxford University Press, 1996.

Kienker, P.K., T.J. Sejnowski, G.E. Hinton and L.E. Schumacher. "Separating Figure from Ground with a Parallel Network". *Perception* 15 (1986): 197–216.

Koch, Christof. "Free Will, Physics, Biology and the Brain". In *Downward Causation and the Neurobiology of Free Will*, edited by Nancey Murphy, George F. R. Ellis and Timothy O'Connor. Springer, 2009: 31–52.

Krohn, Wolfgang, Günther Küppers, Helga Nowotny (eds.). *Self-Organization – Portrait of a Scientific Revolution,* vol. XIV, *Sociology of the Sciences Yearbook*. Dordrecht: Kluwer, 1990.

Kuhn, Thomas. *The Structure of Scientific Revolutions*. Chicago: Chicago University Press, 1962.

Lackoff, George and Mark Johnson. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought.* New York: Basic Books, 1999.

Laplanche, J. and J.B. Pontalis. *The Language of Psychoanalysis*. Translated by Donald Nicholson-Smith. New York: Norton, 1973.

Laughlin, Robert B. and David Pines. "The Theory of Everything". *Proceedings of The National Academy of Science* 97 (2000). Also in *Emergence: Contemporary Readings in Philosophy and Science*, edited by Mark Bedau and Paul Humphreys. Cambridge, Mass.: MIT Press, 2008, Chapter 14.

Leroi-Gourhan, André: *Treasures of Prehistoric Art*. Translated by Norbert Guterman. New York: Harry N. Abrams, 1966.

Lewin, Roger. *Complexity: Life at the Edge of Chaos*. Chicago: University of Chicago Press, 1992.

Lewis, Rhodri. *Hamlet and the Vision of Darkness*. Princeton: Princeton University Press, 2017.

Linley, David. *The End of Physics*: *The Myth of a Unified Theory*. New York: Basic Books, 1993.

Locke, John. "Epistle to the Reader". In *Essay Concerning Human Understanding*, edited by Raymond Wilburn. London: Dent, 1947: xxiii.

Lowe, E.J. "Personal Experience and Belief". In *Cognition and Material Culture*: *The Archaeology of Symbols*, edited by Colin Renfrew and Chris Scarre. Cambridge: McDonald Institute Monographs, 1998: 89–96.

Lucretius. *On the Nature of Things*. Translated by Martin Ferguson Smith. London: Sphere Books, 1969.

Machamer, P.K. and R.G. Turnbull (eds.). *Studies in Perception*. Columbus: Ohio State University, 1978.

MacLean, Paul D. *A Triune Concept of Brain and Behaviour: The Triune Brain in Evolution*. Toronto: University of Toronto Press, 1973.

Mainzer, Klaus. *Thinking in Complexity: The Complex Dynamics of Matter, Mind and Mankind*. New York and Berlin: Springer, 1994 and 1997.

Maturana, H.R. and F.J. Varela. *Autopoiesis and Cognition: The Realization of the Living*. Boston Studies in the Philosophy of Science. Dordrecht: Reidel, 1980.

McCrae, Norman. *John von Neumann, the scientific genius who pioneered the modern computer, game theory, nuclear deterrence and much more.* New York: Pantheon Books, 1992: 283.

McCulloch, Warren S. and Walter Pitts. "A Logical Calculus of the Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics* 5 (1943): 115–33.

McKie, Robin. "Malaria", *The Observer*, April 8, 2018.

Merleau-Ponty, Maurice. "The Structure of Behaviour". Translated by Alden L. Fisher. London: Methuen, 1963.

Morin, Edgar. *The Nature of Nature*, vol. I, *Method*: *Towards a Study of Humankind*. Translated by J.L. Bélanger. New York: Peter Lang, 1992.

Morin, Edgar. *La Méthode,* vol. III*, La Connaissance de la Connaissance.* Paris: Edition du Seuill, 1986.

Morowitz, Harold. *The Emergence of Everything*. New York: Oxford University Press, 2004.

Mukherjee, Siddhartha. *The Gene: An Intimate History.* London: Bodley Head, 2016.

Petitot, Jean, Francisco J. Varela and Bernard Pachoud. *Naturalizing Phenomenology*. Stanford: Stanford University Press, 2000.

Philipse, Herman. "Transcendental Idealism." In *The Cambridge Companion to Husserl*, edited by Barry Smith and David Woodruff Smith. New York: Cambridge University Press, 1995: 239–322.

Pinker, Steven. *The Language Instinct: The New Science of Language and the Mind.* New York: Penguin Books, 1994.

Posner, M.I. and G.D. Girolamo. "Cognitive Neuroscience". *Psychological Bulletin* 126: 873–88.

Poston, T. and I. Stewart. "Nonlinear Modelling of Multistable Perception". *Behavioural Sciences*, 23, Issue 4 (1978): 318–34.

Potter, David. *Computational Physics.* London: Wiley, 1973.

Redner, Harry. *The Ends of Philosophy: An Essay in the Sociology of Philosophy and Rationality.* London: Croom Helm, 1986.

Redner, Harry. *The Ends of Science: An Essay in Scientific Authority*. Boulder, Colorado: Westview, 1987.

Redner, Harry. *A New Science of Representation: Toward a Unified Theory of Representation in Science, Politics and Art*. Boulder, Co.: Westview Press, 1994.

Redner, Harry. *Malign Masters: Philosophy and Politics in the Twentieth Century*. Basingstoke: Macmillan, 1997.

Redner, Harry. *Beyond Civilization: Society, Culture and the Individual in the Age of Globalization.* New Brunswick, N.J.: Transaction Publishers, 2013.

Redner, Harry. *The Tragedy of European Civilization: Toward an Intellectual History of the Twentieth Century*. New Brunswick, N.J.: Transaction Publishers, 2015.

Reed, Edward. "The Cognitive Revolution from an Ecological Point of View". In *The Future of the Cognitive Revolution*, edited by D.M. Johnson and C.E. Erneling. Oxford: Oxford University Press, 1997: 261–74.

Reed, Edward and Rebecca Jones (eds.). *Reasons for Realism: Selected Essays of J.J. Gibson*. Hillsdale, N.J.: Lawrence Erlbaum, 1982.

Renfrew, Colin, "Mind and Matter: Cognitive Archaeology and External Symbolic Storage." In *Cognition and Material Culture: The Archaeology of Symbols*, edited by Colin Renfrew and Chris Scarre. Cambridge: McDonald Institute Monographs, 1998:1–6.

Roth, Gerhard and Helmut Schwegler. "Self-organization, Emergent Properties and the Unity of the World". In Wolfgang Krohn, Günther Küppers, Helga Nowotny (eds.). *Self-Organization – Portrait of a Scientific Revolution,* vol. XIV, *Sociology of the Sciences Yearbook*. Dordrecht: Springer, 1990: 36–50.

Rothschild, Lynn J. "The Role of Emergence in Biology". In *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*, edited by Philip Clayton and Paul Davies. Oxford: Oxford University Press, 2008: 151–65.

Schiller, Friedrich. *On the Aesthetic Education of Mankind.* Translated and edited by Elizabeth M. Wilkinson and L.A Willoughby. Oxford: Clarendon Press, 1967.

Shakespeare, William. *Macbeth*, Act I, scene v.

Shakespeare, William. *Hamlet*, Act II, scene ii.

Shannon, Claude and Warren Weaver. *The Mathematical Theory of Communication.* Urbans: University of Illinois Press, 1949.

Smith, Kiona N. "Neanderthals were artists and thought symbolically". In *Ars Technica*, 23 Feb. 2018.

Smolin, Lee. *The Trouble with Physics: The Rise of String Theory, the Fall of Science, and What Comes Next.* New York: Houghton Mifflin, 2006.

Spinoza, Baruch. *Ethics*, Part III, proposition 2. In *The Collected Works of Spinoza*, vol, I. Edited and translated by Edwin Curley. Princeton: Princeton University Press, 1985.

Stadler, Michael and Peter Kruze. "Cognitive Systems as Self-Organizing Systems". In *Self-Organization – Portrait of a Scientific Revolution,* vol. XIV, *Sociology of the Sciences Yearbook*, edited by Wolfgang Krohn, , Günther Küppers and Helga Nowotny. Dordrecht: Springer, 1990: 181–93.

Stent, Gunther S. *The Paradoxes of Progress*. San Francisco: Freeman, 1978.

Stewart, Ian and Jack Cohen. *Figments of Reality: The Evolution of the Curious Mind.* Cambridge: Cambridge University Press, 1997.

Tallis, Raymond and Howard Robinson (eds.). *The Pursuit of Mind.* Manchester: Carcanet Press, 1991.

Thompson, D.W. *On Growth and Form.* Cambridge: Cambridge University Press, 1942.

Thompson, Evan. *Mind in Life*: *Biology, Phenomenology and the Sciences of Mind.* Cambridge, Mass.: Harvard University Press, 2007.

Thompson, Evan. "Contemplative Neuroscience as an Approach to Volitional Consciousness". In *Downward Causation and the Neurobiology of Free Will*, edited by Nancey Murphy, George F.R. Ellis and Timothy O'Connor. New York: Springer, 2009: 187–97.

Turing, Alan. "On Computable Numbers with an Application to the *Entscheidungsproblem*". *Proceedings of the London Mathematical Society*, vol. 2, 42/3 (Nov.1936): 230–65.

Turkle, Sherry. *The Second Self: Computers and the Human Spirit*. New York: Simon and Schuster, 1984.

Uleman, James S. "Becoming Aware of the New Unconscious". In *The New Unconscious,* edited by Ran R. Hassin, James S. Uleman and John A. Bargh. New York: Oxford University Press, 2006: Introduction.

Varela, Jean, Petitot Francisco J., Bernard Pachaud and Jean-Michel Roy (eds.). *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*. Stanford: Stanford University Press, 1999.

Waldrop, W. Mitchell. *Complexity: The Emerging Science at the Edge of Order and Chaos*. New York: Viking, 1992.

Wegner, Daniel M. "Who is the Controller of Controlled Processes?" In *The New Unconscious*, edited by Ran R. Hassin, James S. Uleman and John A. Bargh. New York: Oxford University Press, 2006: Chapter 1.

Weinberg, Steven. *The First Three Minutes*. New York: Basic Books, 1993.

Weinberg, Steven. *Dreams of a Final Theory: The Search for the Fundamental Laws of Nature*. London: Random House, 1993.

Weinberg, Steven. "Newtonianism, Reductionism and the Art of Congressional Testimony". In *Nature* 330 (1987): 433–7. Also in *Emergence: Contemporary Readings in Philosophy and Science*, edited by Mark Bedau and Paul Humphreys. Cambridge, Mass.: MIT Press, 2008: Chapter 18.

Woit, Peter. *Not even Wrong: the Failure of String Theory and the Search for Unity in Physical Law.* New York: Basic Books, 2006.

Yakira, Elhanan. *Spinoza and the Case for Philosophy.* Cambridge: Cambridge University Press, 2015.

Index