In the Theater of Consciousness

The Workspace of the Mind.

Bernard J. Baars

Oxford University Press

1997

The Neurosciences Institute
www.nsi.edu/users/baars
San Diego, Calif. 92121

HOME: 3616 Chestnut St., Apt. 3
Lafayette, Calif. 94549

baars@nsi.edu
bbaars@comcast.net

This book is dedicated to Francis H.C. Crick,

--- who knows a fundamental scientific problem when he sees it ---

for his leadership in bringing scientists back to consciousness

and

to those who advance freedom of speech and thought

wherever it may be under siege.

Table of Contents

Preface.

Ever since the *nova scientia* of Galileo and Copernicus began the revolutionary rise of modern physics, new sciences have been proclaimed with some regularity. Most of these announcements turn out to be false alarms. But today we actually find ourselves at one of those rare nodal points in the evolution of human understanding: For the first time in the hundred years since William James' *Principles of Psychology,* serious brain and psychological scientists are exploring conscious experience --- often under obscure labels --- but now with far better evidence and theory than ever before.

Only distant hints of the current ferment in scientific consciousness research have reached the public; some of the most fascinating findings and ideas have simply gone unnoticed. Yet our own consciousness is in many ways the most significant topic imaginable to us as human beings; nothing else is as close to us, and nothing has been as consistently baffling and mysterious to untold generations gone before. Urgent ethical questions depend on a better understanding of human consciousness, and in a world where science is often seen as a double-edged sword, the public has a natural interest in understanding developments at the frontier. As a cognitive neuroscientist with almost two decades of experience grappling with these questions, I thought it was high time to tell the story of some of the best scientific work available today. This book is the result.

The new consciousness science.

The scientific race for consciousness is now on. Dozens of laboratories are focusing their efforts to be first with major new findings. Widely respected figures in neurobiology like Francis Crick, Gerald Edelman, Rodolfo Llinás, Michael Gazzaniga and Joseph Bogen have already devoted many years to the enterprise. Psychologists Endel Tulving, Daniel Schacter, Hanna and Antonio Damasio, Morris Moscovitch, and others have studed memory after brain lesions with some remarkable results. Cognitive scientists who have made strong claims include Michael Posner, Tim Shallice, Daniel Schacter, and even myself. Philosophers are having a field day.

Human experience is a humbling problem, but one on which we can make step-by-step progress. Whole bodies of research in perception, selective attention, and immediate memory form a solid and reliable basis for this effort. In the last several years psychologists and brain imaging researchers have begun close and successful collaborations, so that we are beginning to see what the brain is actually doing when it is thinking and seeing and remembering, and much of the evidence fits our expectations. For the first time we can try to bring all these sources together to see if they make a unified story. Even as we do so, new findings appear each week. Today a sizable body of evidence points to the conclusion that consciousness is a key biological adaptation that makes it possible for the brain to interpret, learn about, interact with, and act upon the world.

--- o ---

This book is based on a framework for understanding the large domain of evidence bearing on our personal experience called Global Workspace theory. GW theory presents a "theater model" in which consciousness requires a central workspace, much like the stage of a theater. It is set out in detail in my book *A cognitive theory of consciousness* (Cambridge University Press, 1988) and a set of technical publications. (1)

Global Workspace theory is based on the belief that, like the cells of the human body, the detailed workings of the brain are widely distributed. There is no centralized command that tells neurons what to do. Just as each cell in the body is controlled by its own molecular code, the adaptive networks of the brain are controlled by their own aims and contexts. To organize this vast distributed domain there is a network of neural patches that work together to display conscious events. Today the best candidates for these loci of conscious experience may be the sensory projection areas of the cortex, where the great neural radiations coming from the eyes, the ears, and the body first reach the surface of the brain. A few small structures of the core brain stem and midbrain are essential to consciousness, but great quantities of tissue elsewhere in the brain can be lost without causing a loss of conscious experience. Conscious contents appear to be disseminated globally to a great multitude of networks throughout the brain that are unconscious, but that have observable conscious consequences downstream.

As it happens, all unified theories of cognition today are theater models. Global Workspace theory derives from the integrative modeling tradition of Allan Newell, Herbert A. Simon, John Anderson and others in cognitive science. It is consistent with models of working memory by Alan Baddeley, of the mind's eye by Stephen Kosslyn, of explicit knowledge after brain damage by Daniel Schacter and Morris Moscovitch, the thalamocortical searchlight elaborated by Francis Crick, and society models outlined by Michael Gazzaniga and Marvin Minsky. The brain implications of Global Workspace theory have been explored by James Newman and myself. British mathematician John G. Taylor and others are working to apply modern "neural net" models to the problem. The convergence of ideas today is simply astonishing.

--- o ---

I have sprinkled this book with demonstrations that I hope will appeal to your personal experience, but that *also* yield reliable public reports of the kind we use in science. From a scientific point of view we cannot share your personal experience, so that we can deal only with your *descriptions* of your experiences. However, for many well-studied phenomena the subjective and the objective evidence converges so well that the distinction has little practical significance. You could, if you wish, understand all the demonstrations in this book from a subjective point of view. Or conversely, you could pretend that none of the demonstrations apply to your own experience, and that we are exploring the objective behavioral and brain processes of an utterly unknown species infesting the surface of the fourth planet of Sol. Neither pretense is necessary, because the inside and outside point of view on the evidence tends to dovetail so well. The many convergent pieces of evidence persuade me at least that in practice, the famous gap between mind and body is a bit of a myth.

The really daring idea in contemporary science is that human conscious experience can be understood *without* miracles --- just as it was Darwin's radical idea that the origin of species could be understood without divine intervention. We are beginning to see consciousness as a key biological adaptation with multiple functions. Conscious contents trigger a host of unconscious processes and are shaped in turn by unconscious *contexts*. Consciousness appears to be essential in integrating perception, thought, and action, in adapting to novel circumstances, and in providing information to a self-system. By comparison, the vast audience of *un*conscious resources seems to be much more isolated and autonomous.

--- o ---

Readers with access to the Internet can use the names of researchers cited throughout this book to find original books and articles. Those interested in the history of behaviorism, introspectionism and the emergence of cognitive psychology may wish to peruse my earlier book, *The cognitive revolution in psychology*.

I owe an immense debt to scientists and philosophers whose dedicated work has helped us return to consciousness. In psychology and cognitive science they include Michael Posner, Don Dulany, Ernest R. Hilgard, George A. Miller, Herbert Simon, Allan Newell, Endel Tulving, George Lakoff, Steve Palmer, Daniel Schacter, John Kihlstrom, Arthur Reber, Bruce Mangan, Tom Natsoulas, Ellen Langer, Donald Norman, George Mandler, John G. Taylor, Peter H. Greene, Donald G. MacKay, William P. Banks, Donald Broadbent, Alan Baddeley, Roger Shepard, Daniel Wegner, John Bargh, Geoffrey Underwood, Stephen LaBerge, David LaBerge, Antti Revonsuo, Terry Sejnowski, and many others. In philosophy the indispensables include Daniel C. Dennett, John R. Searle, Owen Flanagan, Ned Block, David Chalmers, and Patricia and Paul Churchland. In neurobiology I have learned much from the work of Francis Crick, Roger

Sperry, Gerald Edelman, Joseph Bogen, Michael Gazzaniga, Steve Hillyard, Ron Mangun, Semir Zeki, Hannah and Antonio Damasio, V.S. Ramachandran, Morris Moscovitch, Allan Hobson, Nikos Logothetis, Risto Näatanen, James Newman, Oliver Sacks, David Galin, and Charles Yingling. In clinical psychology and psychiatry I want to gratefully acknowledge Mardi J. Horowitz, Jerome S. Singer, David Spiegel, Joseph Weiss, and Lester Luborsky. Many others have my sincere gratitude.

Special thanks are owed for friendship and intellectual stimulation to William P. Banks, fellow editor and founder of Consciousness and Cognition: An international journal, published by Academic Press, which has been a joy and an intellectual lifeline for five years; to Bruce Mangan for his valuable insights into "fringe" consciousness; David Galin, Joseph Bogen, Jim Newman, and Michael Wapner for their rigorous thinking and friendship; Francis Crick, Christof Koch, and Stanley Klein for setting me straight on various aspects of neurobiology; Arthur Blumenthal for help in understanding the history of experimental psychology; Scott Slotnick for his competence, intelligence, and enthusiasm, and for reminding me what a remarkable privilege it is to be alive and working on these historic issues. Katie McGovern has given much intellectual feedback as well as the gift of her love and friendship; Megan and Chris McGovern helped keep me down to earth; Andrew McGovern helped correct the manuscript; and my parents, Louis and Lynn Baars, made it all possible. I am forever in their debt.

Bernard J. Baars

Footnotes.

(1,) Daniel Dennett and Marcel Kinsbourne have criticized a concept called a "Cartesian Theater, " which should probably be called the Cartesian Theater *Fallacy*. Three centuries ago René Descartes proposed that conscious experience comes together in a single point in the brain, in the tiny pineal gland. Dennett and Kinsbourne maintain that the Cartesian Theater reflects a widely shared intuition that consciousness involves a single point center. The Cartesian Theater is a fallacy, a *reductio ad absurdum*, not the sort of thing anyone today would care to use in reality. There is of course no single point in the brain where "everything comes together." We do know of numerous well-established brain maps of the world and of the body, and of convergence zones which integrate many different sources of information into some coherent account of current reality. There is solid evidence that some of these brain maps are conscious, and that others help shape conscious experience. But no scientific model today commits the Cartesian fallacy. Certainly none of the scientific theater models that have been proposed since the 1950s suggest that all conscious experiences come together in a single point. Real theaters are not constructed that way; they work just fine, and provide productive tools for thinking.

Prologue

You are conscious and so am I. This much we can tell pretty easily, since when we are not conscious our bodies wilt, our eyes roll up in their orbits, our brain waves become large , slow, and regular, and we cannot read a sentence like this one.

While the outer signs of consciousness are pretty clear, it is our inner life that counts for most of us. The contents of consciousness include the immediate perceptual world; inner speech and visual imagery; the fleeting present and its fading traces in immediate memory; bodily feelings like pleasure, pain, and excitement; surges of feeling; autobiographical memories as they are recalled; clear and immediate intentions, expectations, and actions; explicit beliefs about oneself and the world; and concepts that are abstract but focal. In spite of decades of behavioristic denial, few would quarrel with this list today.

At this instant you and I are conscious of some aspects of the act of reading --- the shape of *these letters* against the white texture of the page, and the inner sound of *these words.* But we are probably not aware of the touch of the chair, of a certain background taste, the subtle balancing of our body against gravity, a flow of conversation in the background, or the delicately guided eye fixations needed to see *this phrase;* nor are we now aware of the fleeting present of only a few seconds ago, of our affection for a friend, and some of our major life goals. These unconscious elements are as important as the conscious ones, because they give us natural comparison conditions.

For example: While you are conscious of words in your visual *focus*, you surely did not consciously label the word “focus” just now as a noun; yet this sentence would be incomprehensible if highly specialized language analyzers --- located in the cortex of the brain, just above each ear --- did not label “focus” as a noun *unconsciously*. The meaning would change significantly if you understood it to be a verb or an adjective.

On reading "f o c u s,” you were surely unaware of its nine alternative meanings, though in a different sentence you would instantly bring a different meaning to mind. What happened to the others? A wealth of evidence supports the notion that some of those meanings existed unconsciously for a few tenths of a second before your brain decided on the right one. Most words have multiple meanings, but only one can become conscious at a time. This seems to be a fundamental fact about consciousness.

These examples illustrate the sense of the word "consciousness" we wish to understand --- that is, *focal* consciousness of easily described events, like "I see a printed page," or "He imagined his mother's face." A great body of evidence shows that conscious contents like this can be reported *as conscious* with great accuracy under the right conditions. These conditions include immediate report, freedom from distraction, and some way for the outside observer to verify the report. These are standard laboratory conditions that apply to thousands of experiments in perception, memory, attention, and mental imagery. They also fit the demonstrations presented throughout this book.

Whenever a question about the meaning of consciousness arises in these pages, I would invite you to revisit the paragraphs above. The meaning of “consciousness” intended here is best illustrated by your own experience.

Verifiable public report is the key to scientific evidence, but *your experience here and now* is quite a good index to the evidence. All the subjective demonstrations used in this book can be tested objectively, and all the objective facts can be experienced by you and me. That is why we believe we can talk about consciousness *as such.*

The metaphor.

In Book VII of Plato's Republic we find the following allegory:

> Imagine mankind as dwelling in an underground cave ... in this they have been since childhood, with necks and legs fettered, so they have to stay where they are. They cannot move their heads round because of the fetters, and they can only look forward, but light comes to them from fire burning behind them ... . Between the fire and the prisoners ... imagine a low wall has been built, as puppet showmen have screens ... then bearers carrying along this wall all sorts of articles which they hold .... statues of men and other living things, made of stone or wood ... What do you think such people would have seen of themselves and each other except their shadows, which the fire cast on the opposite side of the cave? ... such persons would certainly believe that there were no realities except those of those shadows .... (p. 312)

Plato's famous Allegory of the Cave has odd and unexpected resonances down the ages, and indeed in Asian philosophy as well. Indeed, two and a half millennia later another observer wrote,

> A common metaphor is that of a 'spotlight' of visual attention. Inside the spotlight the information is processed in some special way. This makes us see the attended object or event more accurately and more quickly and also makes it easier to remember. Outside the 'spotlight' the visual information is processed less, or differently, or not at all. The attentional system of the brain moves this hypothetical spotlight rapidly from one place in the visual field to another, just as, on a slower time scale, you move your eyes. (Crick, 1993, p. 62)

Plato's point was the fallibility of conscious perception compared to the eternal verities of philosophy, while Francis Crick was aiming to understand the relationship between the brain structure called the thalamus and the great cerebral cortex, both necessary for conscious experience. What is the difference between Plato's fire-cast shadows and Crick's thalamic spotlight? I feel moved as much by the similarities as the differences: Both are unifying conceptions of human consciousness. In fact, both seem to reflect the same underlying metaphor of our personal experience, the *theater metaphor*. (2) Plato’s Allegory of the Cave may be more elaborate, but Crick’s spotlight has a more solid basis in brain anatomy and physiology. A number of cognitive and brain scientists have suggested versions of the theater metaphor, and in ancient times the same theme was sounded in Vedanta philosophy, at various points in Western thought, and surely by poets and philosophers in many other times and places.

Theater models have become attractive again today as scientists have realized how much of the brain’s work is done unconsciously, by innumerable small bits of specialized brain tissue. This may be hard to realize from our own experience, but just look at the brain:

1. The brain has on the order of 100,000,000,000 nerve cells.

2. Each nerve cell has about 10,000 inputs (dendrites) from other neurons, and only a few outputs (axon terminals).

3. Neurons are so interconnected that one can get from one cell to any other cell in the brain in seven steps or less.

4. Most of the brain consists of small assemblies of brain cells, arrays, columns, maps, clusters, networks, functional routines, and great swaths of cable connections, all with highly specific functions. Some nerve cells pick up point sources of red light in a single location in the visual field. Others specialize in short lines oriented 45 degrees above the horizontal. Still others recognize faces or coordinate sights with sounds. Specialization is the name of the game for most neurons. But the great bulk of these tissues are unconscious.

The brain seems to show a *distributed style of functioning,* in which the real work is done by millions of specialized, sophisticated systems without detailed instructions from some command center. By analogy, the human body also works cell by cell; unlike an automobile, it has no central engine that does all the work. Each cell is specialized for a particular function according to instructions encoded in its DNA, its history, and chemical influences from other tissues. And the cell is of course the body's basic unit of organization. In its own way the human brain shows the same distributed style of organization.

The theater metaphor is useful because a great array of evidence indicates that *consciousness creates access* to many knowledge sources in the brain. And yet only a fraction of the brain seems to directly support conscious experience. This consciousness network seems to include the sensory areas of the cortex, perhaps some surrounding areas, and a few subcortical structures; together they provide the stage for the unconscious audience in the rest of the brain. Consciousness seems to be the publicity organ of the brain. It is a facility for *accessing, disseminating and exchanging information*, and for *exercising global coordination and control.*

As it happens, all of our unified models of mental functioning today are theater metaphors; it is essentially all we have. (1, 2) They are called by a variety of names and have been developed over the last forty years based on a vast range of evidence, from studies of chess players to work on inner rehearsal, from mental rotation of visual images to the subtle effects of brain damage. A remarkable group of distinguished scientists have devoted careers to these integrative conceptions of human cognition.

What has not been done so far is to forge a working link between these great bodies of thought and the core issue of human experience.

That is the aim of this book.

(1) Neural net models, developed over the last ten years, have led to important theoretical insights. But they have not yet been scaled up to large-scale integrative theories of cognition. Neural nets do a superior job modeling some phenomena of consciousnesss, but they do not easily reflect

its large-scale architectural properties. One promising approach is to build hybrid architectures combining neural nets with theater models.

(2) Readers wishing to compare cognitive theater models to the "Cartesian Theater" of Daniel P. Dennett and Marcel Kinsbourne, please see Footnote (1) on page (Preface).

Part I.

Carving nature at the joints.

What evidence is relevant to consciousness *as such?* This has been one of the most difficult questions for psychologists and brain scientists to contend with. It has led to great philosophical controversy; but there is a growing scientific consensus today, so that in research circles the debates have faded.

As everywhere else, evidence for consciousness is that which "carves nature at the joints." In this case, it involves experimental comparisons between very similar conscious and unconscious mental processes; that is, comparisons that allow us to treat consciousness as a variable. The closer the comparison, the more it tells us about the differential effects of consciousness. Waking up this morning compared to being in deep sleep a few minutes before, provides one example. Another is the things you pay attention to, compared to brain processing of similar things you do not. A third is your memory of this morning's breakfast, before and after conscious recall. There are dozens of such cases, and together they place strong empirical bounds on the concept of conscious experience.

Chapter 1

Treating consciousness as a variable.

There is no more important quest in the whole of science probably than the attempt to understand those very particular events in evolution by which brains worked out that special trick that enabled them to add to the cosmic scheme of things: colour, sound, pain, pleasure, and all the other facets of mental experience.

--- Roger Sperry

We can only study something if we can treat it as a variable, to compare its presence to its absence. A number of historic breakthroughs in science emerged from the realization that some previously assumed constant, like atmospheric pressure or gravity, was actually a variable. It required a great leap of imagination for natural philosophers to understand that all objects in the uninverse need not fall toward the earth; that gravity could be different elsewhere. Newton's ability to make that leap led to the solution of the ancient puzzle of planetary motion.

Consciousness has seemed to be different from all other scientific concepts; it has been extraordinarily difficult to treat it as a variable. The persistent pattern over centuries has been to see our own experience as the only psychological domain that can be conceived, one that has no kinship to any conceivable comparison condition. And our own consciousness is hard to study because we cannot vary it; as soon as we lose consciousness, we can no longer observe. And how can we study someone else's conscious experience?

It is actually possible to compare conscious events that people can report accurately to unconscious ones that can be inferred and studied indirectly. Scientists have now tested a number of cases in which we believe conscious and unconscious conditions to be quite similar. We know a great deal about normal versus subliminal perception; attended versus nonattended speech; novel versus routine and automatic processes; explicit compared to implicit memory; and much more.

We can call this method contrastive phenomenology, *to emphasize the involvement of private experience. Phenomenology is the study of consciousness based on subjective experience; In cognitive science we always supplement subjective reports with objectively verifiable methods. Nevertheless the reports we routinely collect in studies of perception, imagery, and memory, are generally reports about conscious experiences. Thus in modern science we are practicing a kind of verifiable phenomenology.*

For scientific purposes we prefer to talk about public reports of conscious experiences compared to public, inferential evidence for unconscious processes. But there is such a close mapping between objective reports and subjective experience in this book, that for all intents *and purposes we can talk of* your phenomenology, *your consciousness as experienced. You can demonstrate to yourself all the basic properties that have begun to emerge.*

Mozart was a master of consciousness. Only a composer of the highest musical order could open the delectable *Sinfonia Concertante* on a *single note,* played by the solo violin and viola in perfect unison an octave apart, two indistinguishable sounds except for the unearthly depth of the added harmonics. *One note* --- think about it. Only a master could bounce madly from Leporello's low buffoonery in the first scene of *Don Giovanni* to the Don's botched attempt to seduce Don'Anna, a fight to the death, more clowning, and finally Don'Anna's furious cry of *vendetta!* --- changing mood in seconds, while maintaining a seamless flow of light and tender commentary in the orchestra --- and somehow combine all these extreme and incompatible elements into a sparkling conscious unity.

While writing this chapter in Denmark I had a chance to visit the great domed cathedral in the old capital city of Roskilde, where seven centuries of Danish royalty lie buried. Standing in the ancient church it became suddenly clear that it is not a *building* in the modern sense at all. The Domkirke is a many-layered story-telling domain. Every niche is filled with symbols, grave stones, allegories, reminders and warnings about the pride and power of kings and queens, and the evocative stories of the faith. No one could take it in at once; one would have to go back many times, exploring its meanings layer by layer, as did the people who worshipped in it for generations. The cathedral was a massive effort to shape the beliefs and fantasies of its people, even as a television program shapes the beliefs and fantasies of children today. But the cathedral was not only for children; it has deeper meaning, more profound attachment to life's joys and terrors, more reminders of the certainty of death, more promises of life after death if one obeyed, more propaganda (in the original meaning of the word) designed to propagate the faith.

In a way scientific exploration of human consciousness is just an extension of the arts, theater, literature and even religion. We are returning to a project that has moved human beings for centuries: to apply the mind to its own understanding. I can imagine nothing more thrilling and ultimately practical.

The challenge.

For years it was common to hear scientists say that human consciousness was unlike any other scientific problem in that it was not at all clear what evidence was relevant to it; as for theory, it seemed so far beyond our comprehension that it was hardly worth talking about. The trouble is that predictions of failure are self-fulfilling. If we are convinced that we will never learn about our own experience we will not even try to understand it, and then we will indeed remain as uninformed as ever. Such all-devouring skepticism seems unreasonable.

Consciousness may not be a solvable problem. In physics the exact orbital dance of three bodies in space is believed to be unsolvable. In biology, recapturing the precise genetic pressures that led to a given mutation is not possible. There are other questions that are simply not possible to answer, perhaps ever. Nature does not come with guarantees. The question remains, is consciousness something that can be known?

Consciousness is a big phenomenon.

A useful rule of thumb in science is to go for the big phenomena first --- those that are stand out like great craggy mountains from the plain --- things like gravity, color and the astonishing family resemblances between animals dwelling continents apart. The big phenomena are the ones that are so obvious that you have to *try* to overlook them. That is where the early payoffs can be found for research; subtleties come later.

Consciousness often seems to be the biggest, loudest phenomenon we could possibly study. No alien space visitor could fail to observe that vertebrates, including humans, engage in purposeful motion only two-thirds of the earthly day. In the remaining third, we hibernate. Coming back to consciousness in the morning, we humans report a rich and varied array of experiences: colors and sounds, feelings and smells, images and dreams, the rich pageant of everyday reality. We rise and begin to engage in purposeful action. Our brains begin a whole new mode of functioning. In this broad sense, the centrality of consciousness is pretty much beyond doubt.

It is sometimes said that there is no known brain index of consciousness, but that is not true. Place two electrodes on any spot on a person's scalp (silver dimes will do), connect wires to the coins, and hook the wires in turn to your stereo amplifier. Show the stereo output on a video screen, and you will see large, slow, regular electrical waves while your subject is in deep sleep, followed by small, rapid, irregular wavelets when he or she wakes up. Even such gross electrical signals, seeping through layers of skin and bone, show the conscious state to be utterly different from deep sleep or coma. It is not a subtle sort of thing.

Now imagine science as a great community effort to fit together an enormous cutout puzzle, too large for any single mind to solve, with big and small pieces. Which do you start with? The advantage of starting with the big pieces is that they constrain many others; the smaller ones don't tell us nearly as much. It is for that reason, if for no other, that conscious experience is such a vital scientific issue. It is a great central piece that locks many others into place.

Treating consciousness as a variable.

We live in the middle of a gravity well, a local maximum of gravitational attraction that makes it expensive in energy to escape into space. But the effects of the earth's gravity on our bodies has become so predictable to each of us that its very existence goes unnoticed. It required a leap of imagination for natural philosophers to understand that gravity could be different elsewhere in the universe than here on earth. Newton's ability to make that leap made it possible to solve the ancient puzzle of planetary motion. Variable gravity was the intellectual breakthrough needed to understand bodies in motion, one that solved many problems as soon as it clicked into place. Overcoming our earthbound perspective was an essential precondition for classical physics.

A hundred years ago naturalists had to *learn* to think of species as variable. Animal and plant species are stable over the human lifespan and our intuitive notion of immutable species is by no means irrational or dumb. If few naturalists before Charles Darwin believed in evolution it was in part because of the unbelievable amount of time needed for species to evolve, and in part because evolution broke the boundaries of the comforting human-sized universe as told in Genesis. Whatever the reason, seeing species as variable over time made it possible to understand how house cats and sabre-tooth tigers could be close relatives, and how farmyard chickens could be evolutionary cousins to Tyrannosaurus Rex. Seeing a presumed constant as a variable was the key.

The same perspective trick has worked at other turning points in history. Relativity theory turned space-time into a variable. Plate tectonics changed the continents into floating slabs of earth crust, and Riemannian geometry bent Euclid's parallel lines into converging arcs. Again and again we have been obliged to go beyond our local world with its predictable gravity, immutable plants and animals, flat and immovable surfaces, stable atmospheric pressure and well-defined accepted beliefs. As life's constants become relativized all the certainties we live by become frighteningly unpredictable for a while. The payoff, of course, is insight.

Consciousness has seemed to be different from all these scientific concepts; it has been almost impossible to treat as a variable. The persistent pattern over written history has been to see conscious experience as the *only* psychological domain that can be conceived, one that has no kinship to any conceivable comparison condition.

In the year 1637, Descartes expressed the view that everything in the mind must be conscious:

> ... there can be nothing in the mind... of which it is not aware, this seems to me self-evident. For there is nothing that we can understand to be in the mind ... that is not a thought or dependent on a thought.

In the early Enlightenment this suggests an image of the mind as a sunny little room, in which one can look around and see the daylight shining through the window, the bright flowers on the table, the portrait on the wall. We can be conscious of everything in this little room, even ourselves. There is no unconscious in this sunlit early Enlightenment mind, just a turning of attention from one thing to another. Of course if there is no unconscious knowledge, consciousness has no comparison condition.

--- o---

By wide consent the foremost work on human mental processes, even today, is William James's *Principles of Psychology,* which appeared in 1890. The *Principles* offers 1,300 pages of inspired dialogue on the major topics of psychology. Building on fifty years of European studies, it has given us classic descriptions of selective attention, mental imagery, hypnosis, habit and effortful concentration, the stream of consciousness, the basic arguments for and against unconscious processes, a theory of voluntary control and impulsiveness, the crucial distinction between self-as-subject and self-as-object, and much more. On many of these topics James' thinking is fully up to date, and it is embarassing but true that much of the time he is still ahead of the scientific curve.

Entire research domains have been inspired by single passages in the *Principles.* Attention researchers routinely cite James' passage on selective attention as the definition of their topic. (Chapter 4.) Other seminal passages describe the limits of immediate memory, the "tip-of-the-tongue" phenomenon, the use of conscious images to control voluntary action, the central role of habit and automaticity, and more.

And yet *The Principles of Psychology* contains a crucial flaw. Along with sparkling insights, stated with unequalled clarity and verve, James was convinced the mind was limited to conscious processes alone. James believed that consciousness was the sole instance of mentality, while *un*conscious events were "only physical." The two were different metaphysical substances. It was impossible to treat consciousness as a variable, and as a result it was impossible to study its effects.

Most people in the nineteenth century agreed with James: "unconscious intelligence" was a bizarre oxymoron to our great-grandparents. Consciousness was the crown of human reason; unconsciousness was merely a bodily function. Only toward the end of the nineteenth century did scientific thinkers --- notably Pierre Janet in Paris and Sigmund Freud in Vienna --- begin to infer unconscious processes quite freely, based on post-hypnotic suggestion, jokes, neuroses, mental disorders, slips of the tongue, thought avoidance and the like. Freud's ideas have achieved so much influence among the educated public that the art and literature of our time is incomprehensible without them. But Freud had curiously little impact in scientific psychology and brain science, because his claims about unconscious influences could not be tested in a persuasive way.

Behavioristic scientists after James's death in 1910 rejected the whole business of conscious experience because it seemed rife with endless, useless perplexities. Naturally they made little progress on these questions during their seven decades of dominance, because they had no way of thinking about *either* conscious or unconscious processes. They, too, were unable to view consciousness as a variable. Neither William James nor B. F. Skinner could apply the experimental method to the most humanly important topic of all.

The cognitive unconscious.

There is now solid evidence that for most conscious events we can find unconscious ones of comparable complexity. Unconscious routines are believed to be involved in all mental tasks, though they seem to lack the unity, coherence, and accessibility of conscious experiences.

The simplest pattern of findings comes from subliminal presentation of words. After years of controversy psychologists Norman Dixon and Anthony Marcel finally provided persuasive results that words presented too fast to identify consciously are still processed in the brain, at least to the level of meaning. A subliminal word like /dog/ makes it easier to identify /puppy/ as a real word a few seconds later. Hundreds of careful subliminal studies have been published over the last ten years, and this "unconscious priming effect" now looks very reliable.

Subliminal processes seem quite limited, however. Anthony Greenwald has recently shown that two-word compounds like *honey cake, rat hole,* and *potato salad* cannot be combined into single concepts subliminally. Instead, they are treated as single words, *honey, cake, rat,* and *hole,*. Combining two words into a single concept seems to require consciousness. And there is little evidence that subliminal stimuli can influence our actions or attitudes. One cannot flash *buy coke* on a movie screen and expect to increase Coca Cola sales during intermission. But unconscious word identification does exist.

Even more persuasive has been research on automaticity of highly practiced skills, especially automatic language processing. When you and I read a sentence such as this one, we do much more work unconsciously than consciously. Learning a language is largely a process of establishing automatic routines for word recognition, syntactic inference, and semantic interpretation. All those hours very young children spend listening to older children and adults talking, with little or no understanding, do pay off in the most intellectually remarkable feat of our lives, the conquest of language. Becoming a skilled speaker and listener means paying attention (*being* *conscious* of sounds and meanings), but over time each new discovery becomes automatic and unconscious, as it must, because the vast complexity of language analysis would otherwise overwhelm our limited conscious capacity.

All that has become clear only recently. Over the centuries our own experience has been so compelling that most people have been simply unable to believe that unconscious processes could be anything like conscious ones. Consciousness was literally incomparable.

If we live at a historic time today for the study of human experience, it is not just that we have more facts, but that we can treat consciousness as a variable like so many others. The evidence discussed in this book comes from a large set of comparisons between conscious and unconscious processes. More findings come in by the week and month.

Today, remarkable new brain imaging techniques are showing us the heartbeat of the living brain. Psychological methods have been honed to a sharp edge, so that we now routinely measure the time course of mental processes down to a few tens of milliseconds. We have much better ways of modeling mental process than ever before. But the best technical tools are of no help if we cannot *think* with clarity. That is one of our aims.

Contrastive phenomenology: The experimental method.

Figure 1-1 presents visual image you may want to keep in mind to symbolize experimental comparisons in which we try to keep everything constant except the degree of consciousness of whatever is going on. It shows two PET scans of a single horizontal slice of a single person's brain, while the subject was performing the same activity, four weeks apart. Everything is held as constant as possible, but there is a difference. In the left scan the subject was just beginning to learn to play the computer game Tetris; on the right, he or she had acquired considerable skill after a month of practice. The dramatic contrast in brightness between the left and right scans may reflect the fact that people are conscious of novel tasks in much greater detail than they are of the same task when it has become automatic. In other words, we may be looking in these brain images at the effects of different degrees of consciousness in playing Tetris. The two brain scans, side by side in close experimental comparison, provides a helpful reminder of the method of *contrastive phenomenology.*

Figure 1-1 is in the color photo section.

If we could zoom in on one individual neuron, located in a single bright spot on the left scan, we would see the nerve cell communicating frantically to its neighbors about 1000 times per second, and taking in a continuous flow of input from others along its 10,000 dendrites. Local blood capillaries are expanding to service the increased energy needs of the cell, with oxygen and glucose being ferried rapidly to the spot and metabolic waste products flowing away in a ceaseless stream. Star-shaped glial support cells actually walk to the sites of neural activity to ~~render support.~~ dispose of damaged cells.

PET scans reflect the blood flow carrying nutrients to nerve cells, and the more active the cells, the greater the local blood flow tends to be. In this case, the subject was injected just before the PET scan with a solution of a glucose isotope, which emits positrons over a period of 30 to 40 minutes.

Active nerve cells require more glucose, and greater concentrations of positrons signal greater metabolic activity over time.

The subject's head is surrounded by a gigantic steel drum containing positron detectors. A positron hit on a detector triggers a small voltage whose location is noted and sent to a powerful computer, which summarizes the concentration of emissions on a numerical map, and then constructs a precise three-dimensional graphic model of the living brain in arbitrary pseudocolors --- the colors you see on the brain scan --- so that we can later select any desired slice of the brain during the critical minute of recording and study it in detail. The bright colors symbolize high metabolic activity and thus high neural activity; the dull colors stand for lower neural activity.

One hypothesis then is that the left scan is brighter because the subject is just learning to play Tetris, that he or she is therefore conscious of many more details of the game; increased consciousness may mean greater neural activity. While we always need to test alternative hypotheses, one reasonable hypothesis about the differences between the two sides is that we are seeing a drop in conscious access to the details of Tetris, as a consequence of learning. The right hand brain scan is so much duller in color because there is less activity.

For a single piece of evidence like this there are of course other explanations. Perhaps the left scan reflects greater *effort* on the part of the subject, rather than more conscious involvement. Maybe new experiences involve more widespread neural activity, independently of consciousness. Maybe the neural code changes between new and old skills. Perhaps there is a change in the location of neural activity between the left and right scans. In fact, there is evidence for several of these ideas.

The most important point for us here is that *the consciousness* *hypothesis is clearly testable.* That is the critical step. Now suppose we find, as we will see to an extent later in this book, that:

1. Sensory areas that are conscious during some task always show more metabolic activity; that

2. Paying attention to one thing rather than another involves greater neural firing; and that

3. Supraliminal compared to subliminal stimulation causes more brain activity as well.

We will then have several converging sources of information to support the hypothesis that *consciousness generally involves increased neural activity, compared to unconscious control conditions.* We can then tentatively conclude that consciousness appears to be associated with increased brain activity, and look for more evidence:

4. In cases of brain damage in which people lose conscious access to some event that is still represented in the brain;

5. In situations in which there are two identical streams of processed input, but people are conscious of only one;

6. In situations in which there is hypnotically suggested perceptual blocking.

Do you see where we are going? The basic strategy is to test the difference between very similar conscious and unconscious brain processes, so that over more and more cases --- assuming we obtain a strong pattern of empirical results --- the consciousness hypothesis can become weaker or stronger. It is the notion of contrastive phenomenology that is key here. It does not *guarantee* an answer to the historic questions about consciousness, because science does not come with guarantees. It does give us an approach to testing the real world for a real answer.

Once we can treat consciousness as a variable like all the others, teasing out the most plausible story becomes only a matter of normal science. It is never easy, and it is certainly not a foregone conclusion that consciousness will turn out to be the critical difference between the left and right scan in Figure 1-1. But even today, we can say that it is most unlikely that consciousness is *unrelated* to learning and automaticity, because of the consistent phenomenological evidence, as we shall see next.

Demonstration: The "phenomenology" in contrastive phenomenology.

To remind ourselves of the different *subjective experiences* between skilled and unskilled performance of the same task,we need only observe ourselves in the act of reading. At this instant your eyes are jumping and fixating, jumping and fixating in a purposeful fashion from one informative phrase on this line to the next, and you are effortlessly identifying letters and words, probably carrying on some inner speech, and encoding the letters on the page into abstract meanings within a second or so. To compare that virtuoso performance to the same experience when it is novel, we need only turn this book upside-down and read the next paragraph. Go ahead and do it.

What do you notice? Well, if you are like me you will become conscious of many details of reading that are normally quite unconscious. You will have trouble telling the difference between letters like n and u, m and w, and d, p, q, and b. You will soon learn to use the surrounding context of each letter to decide if the word is "deer," "peer," or "beer." You will consciously wonder, is that word "quite" or "paint"? Well, it can't be a "q" if it is not followed by a "u," so it must be a "p". All this conscious thinking will get in the way of understanding the *meaning* of this paragraph, so you will have slowed down quite a lot from your ordinary reading speed. You may be vocalizing more than usual to stay focused on the letters and words. If we could take a PET scan of your brain at this moment, we should be able to see much more intense activity than if you were to read this paragraph as usual, right side up.

This is what we mean by contrastive phenomenology. The key is to compare two active brain processes that are similar in most ways, but differ mainly in respect to consciousness. When the game Tetris has become nearly automatic it is still Tetris; when you read the paragraph above upside down you are still reading the same material. Granted that there are some other differences, such as slower reading speed, and the like, these differences are not likely to explain the robust fact that automatic reading is unconscious in many respects, but that it becomes conscious when it is turned into a novel task.

Contrastive analysis does not give an instantaneous answer to the questions about consciousness. It does allow us to ask that question in an empirically sensible way, just as we do anywhere else in science. It is not the last step on the path to an answer, but it could be the first.

Comparing conscious and unconscious qualia.

Many philosophers claim that "qualia," the qualitative aspects of perception and imagery, are the most philosophically puzzling aspects of human experience. That includes colors, textures, shapes, the warm intensity of a musical chord, or the compact purity of a trumpet call. It is hard to imagine how one could explain those conscious qualities, or even what it would mean to engage in such explanation. Surely the firing patterns of nerve cells do not explain the experience of royal blue velvet ?

The cognitive linguist George Lakoff has given a counterargument to the idea that sensory qualities are impossible to explain. He suggests that we can carry out phenomenological contrasts even with qualitative experiences like color. Surely an artist working feverishly on an oil painting is conscious of colors and shapes, but not of that moment a week ago when an odd contrast between a light purple swath of cloth and its pale yellow surround captured his or her imagination.

Yet those moments of past contemplation shape our consciousness in the present --- unconsciously, of course. There must therefore be "unconscious qualia"; the artist must be able to recall that particular shade of purple from memory, where it rested unconsciously for weeks or months; and the process of mixing paints to get just the right hue must be partly automatic. Otherwise how could that particular color be remembered and mixed just at the moment it is needed? And if there are unconscious representations of qualia, it follows that we can, in principle, compare matched conscious and unconscious qualia.

The existence of unconscious sensory qualities does not mean that we have *explained* the experience of a musical chord played on a piano, but rather that we can put the question in a testable way. If we can understand the *difference* between a conscious experience and unconscious knowledge of a C major chord, we will have achieved something like an explanation of that unique and inimitable resonance.

The two-channel experiment: A primary source of evidence about consciousness.

In *two-channel experiments* people receive two dense streams of information that are different enough that they cannot be fused into a single conscious flow. Two narratives will do, one in the left ear and the other in the right, or two rapid ballgames projected on a single video screen, say soccer and basketball. Under these circumstances the brain is compelled to select just one stream of information to come to consciousness. If the stream we are conscious of to makes sense, the other one will be unconscious. The two-channel experiment provides dozens of revealing opportunities to study the nature of the conscious and unconscious flow and the interaction between them. It is one of our most productive source of evidence about conscious *contents*.

Two channel experiments began decades ago with experiments by British psychologist Donald Broadbent and others, in which subjects were asked to listen to two spoken stories, one in each ear, and to <u>shadow</u> one of them. Shadowing simply means repeating each word immediately as it is heard, like a shadow following a strolling pedestrian on a sunny day. Most dual-channel research has been done with auditory streams of information, but to give you with actual experience of these phenomena we will use visual examples. Ideally the words below would be presented by computer, one word at a time, at a fixed point in the middle of a screen. We will be a little more primitive technically and ask your brain to do the work for us.

Here is a visual version of a standard two-channel experiment. Let's start with a familiar poem for the conscious stream. Please read the capitalized words below, out loud if you can, at a fairly brisk pace. Ignore the words in small letters:

Example A.

MARY paper HAD A brick cream LITTLE LAMB morning ITS FLEECE day brick WAS dot WHITE flower fork AS SNOW, brick gem AND frog EVERYWHERE front brick THAT MARY home house ink WENT knife brick THE LAMB milk WAS SURE lilly car TO GO.

For convenience we can call the capitalized words the *primary stream*, and the lowercase words the *secondary stream.*

This method works best if you can read the primary stream of words as quickly as you can, without losing track of its meaning. If you experienced some intruding lowercase words, try reading the primary stream more quickly, until you are clearly aware of one stream and not the other.

Whenever my students and I try this demonstration at an adequate pace we seem to lose access to the lowercase words quite rapidly. Did you realize that "brick" was repeated in the secondary stream four different times? If not, you were probably less conscious of the secondary words. In careful experiments using auditory streams, one to each ear, Donald Norman has found that the *same* word repeated 35 times in the unconscious ear could not be recalled immediately afterward.

Behaviorism was still dominant when two-channel listening was first explored by Donald Broadbent, and at the time, this very revealing technique was never thought of as a tool for studying *consciousness.* "Consciousness" was simply not a defined term. Many researchers still think of these powerful effects as if they only reflect *selection* not consciousness. From our point of view they show both; there is selection in deciding to listen to one ear or to read only capitalized words, and as a result the selected stream of information becomes conscious --- as you know from your own experience. We have a formal test: Does the "attended stream" allow you to make accurate reports about your experience? If so, by common practice it is considered to be conscious.

Now we can easily experiment with different properties of the two streams, and observe some of their brain effects. A unified experience depends on many factors, including some in the secondary stream. Try for instance to read only the *lower case words* in Example A. What is the difference between focusing on the primary and secondary stream? Which way do you get more intrusions? Because the secondary stream contains only disconnected words, people in such experiments commonly report more difficulty maintaining a single conscious focus. Consciousness has a great preference for predictable structure over time.

What if there is no visual difference between the two streams? Let's try to turn both streams into capitals. Try the following example, at the same brisk pace as before.

<u>Example B.</u>

MARY PAPER HAD A BRICK CREAM LITTLE LAMB MORNING ITS FLEECE DAY PARTY WAS DOT WHITE FLOWER FORK AS SNOW, GARDEN GEM AND FROG FRONT EVERYWHERE BOX THAT MARY HOME HOUSE INK WENT KNIFE LETTER THE LAMB MILK WAS SURE LILLY CAR TO GO.

It seems awfully hard to hang on to the primary stream without visible help from word size. But is it just a matter of the visible stimulus? What about keeping the visible distinction between the capitalized and lowercase flow, but scrambling the words of "Mary had a little lamb"? Let's try it.

Example C

WAS paper SNOW A brick AND cream SURE morning ITS day HAD brick WAS dot WHITE flower THE fork MARY, brick AS gem LITTLE frog GO front brick MARY home FLEECE house TO ink LAMB knife THAT brick LAMB milk EVERYWHERE lilly car WENT.

Evidently the *structure and cohesion* of the conscious stream is also an important factor. This is a basic finding about conscious experience. As a rule *anything* that helps us to maintain a sense of coherence about the conscious stream also tends to keep the two streams separate and distinct. When the primary stream coheres well, the secondary one is shut out quite effectively.

Given a convenient experimental technique we can explore all the variables we can think of. We can compare *saying* "Mary had a little lamb" to *singing* it. (Which do you think will work better?) You can change the incentives people have to separate the two channels, compare younger and older people, and look for the effects of specific themes and contents.

What about the meaning level of language? Let’s see if we change the wording a little bit without altering the meaning substantially ...

Example D.

MARY paper OWNED brick cream A BABY LAMB morning WITH PALE day WOOL, brick WHEN dot SHE flower brick

WALKED gem frog ANYWHERE front brick THE SMALL home house ink ANIMAL knife brick CAME milk AFTER lilly car HER.

That seems to slowing things quite a bit. But let's change the meaning even more, so that the conscious stream becomes less predictable.

Example E.

JOHN paper STOLE A brick cream GIANT LLAMA, morning ITS FUR hat ink room WAS SOILED brick AND SMELLY. brick WHEN dot THE ANIMAL part RAN block lilly ANYWHERE, jewel THE STRAPPING spoon YOUNG flower fork MAN, brick HAD TO gem frog PURSUE front brick THE SILLY home house ink BEAST HOME.

Obviously we could now experiment with Jabberwocky speech (no real words at all, just syntax and function words), and any number of other variables. Much evidence already exists along these lines. We will refer to it throughout this book.

Is the unconscious stream simply shut out of the nervous system? Apparently not. Although words from the unattended stream cannot be reported there are several indication that thay can still be processed *unconsciously.* Thirty years ago Neville Moray found that one’s own name in the unconscious stream will break through to consciousness. Other highly significant events, like a baby crying in the night, will do the same.

Here is an illustration.

Example F.

MARY paper HAD A cream, danger! LITTLE LAMB fire! help! ITS FLEECE day, fear! WAS alarm! WHITE bomb! fork A S SNOW, scare! gem AND scandal! EVERYWHERE (your name) brick THAT MARY home, house, ink WENT knife, brick THE LAMB milk WAS SURE lilly TO GO.

You get the idea. It is a simple effect, but it has important theoretical implications. It shows, for example, that unconscious word processing is not superficial; unconsciously we probably analyze sound, word identity and meaning, and personal significance. How else could the brain detect the significance of the unconscious information?

While this "breakthrough" effect for significant unconscious stimuli has been known for many years, there are still many unanswered questions. Will *new* combinations break through to consciousness? We might think not; but we can actually test the matter with a little bit of work. It is easy to find two-word combinations that are emotionally evocative when they are combined, but not separately. For example, "eat" followed by "dirt", "drink" by "acid," or "drop" by "dead." Do these new word combinations break through to consciousness? If single alarming words can break through, but novel pairs of words do not, we may have learned something significant about conscious and unconscious mental processes.

Or suppose you want to test a hypothesis about unconscious personality influences. Psychoanalysts maintain that certain unconscious thoughts are alarming for one kind of person, but not for others. Lloyd A. Silverman and Howard Shevrin have advanced claims about subliminal messages with highly personal content; for example, schizophrenics are said to feel more relaxed and calm after subliminal presentation of "Mommy and I are one." These words are said to lead to a temporary reduction in symptoms.

Such claims are controversial but not untestable. Suppose we ask someone who suffers from a fear of heights to listen to a conscious sentence like "I saw a tiny car in the street below," and simultaneous with the last word "below" present a word like "fall" in the secondary ear. Would "fall," which is consistent with the fear of falling from a height, tend to break through to consciousness, as a baby's cry in the night will wake up its parents? If we wanted to test a conflict hypothesis about personal guilt as a cause of phobia, would the unconscious word "guilt" tend to break through? These questions can be tested easily with very close experimental comparisons.

The cognitive psychologist Donald G. MacKay has discovered even subtler influences from the unconscious stream. Most words in English have more than one meaning, and a great part of understanding speech is deciding which meaning of an ambiguous word is the right one in some particular context. Can unconscious input influence conscious word interpretation? In a classic experiment MacKay presented a sentence like “John and Mary were walking by the bank” in the conscious channel. One set of subjects received a series of words in the unconscious channel, designed so that “money” would be presented just as the word “bank” occurred on the conscious side. A comparison group received “water” instead of “money.” Question: Would the conscious word “bank” be swayed by the simultaneous unconscious word? As it turns out, “money” shifted the conscious interpretations to “financial institution,” while “water” moved it toward “river bank.”

Notice that MacKay's finding is quite different from unconscious "breakthroughs" into the conscious stream. When a baby's cry breaks through to consciousness it interrupts other thoughts. But in MacKay's experiment the conscious flow is never interrupted; only the *meaning* of ambiguous words seems to shift. Evidently there is deep language processing going on on the unconscious side, since sound, word identity and meaning must be understood before the conscious meaning of "bank" can be swayed.

Notice that these interactions of conscious and unconscious goings-on suggest a mind that is not divided into two isolated boxes, one called "Conscious," and the other, "Unconscious." A naive reading of Freud might suggest that conscious and unconscious events are quitet separate and do not interact on a continuing basis. The cognitive unconscious that emerges from these experiments is quite different. Rather than showing two separate boxes, the two sides of the mind interact ceaselessly, like some Jamesian stream, whose course is continually shaped by unconscious rocks, banks, shoals and trenches, invisible but powerful.

In sum, the two-channel experiment is simple, convenient, and revealing. It allows us to vary both the conscious and the unconscious stream, and to study their interaction. Only a fraction of its possibilities have been explored to date. Far more remains to be done with this method, which may be the most practical way to study contrastive phenomenology today.

Consciousness as a state: *Wakefulness and coma*

A final source of phenomenal contrasts comes from comparing waking consciousness with deep sleep and coma. An astonishing amount of brain tissue can be lost from the great cerebral hemispheres without abolishing the state of consciousness, while tiny lesions in the slender axial core of the brain cause irreversible coma. We have known since the 1950s that a small area in the brainstem called the *reticular formation* is necessary for waking consciousness. (Figure 1-2) When people with head injuries go into coma, it is often because tissue damage at the front of the head causes widespread swelling, choking off the blood supply even to the brainstem. When oxygen supply is lost to the reticular formation the result is coma; and because the nerve centers that control breathing and heart action are located very nearby, death often follows.

It has now become clear that a second part of the core brain is required for waking consciousness; it is located an inch or so upward from the reticular formation, in the central transfer station of the brain, the thalamus.

The case of Karen Ann Quinlan caused headlines some years ago as a dramatic example of loss of consciousness *without* loss of vital functions like breathing and heart beat, a type of coma called *persistent vegetative state.* Karen Quinlan, a young woman in her twenties, lost consciousness after suffering cardiac arrest, followed by a stroke that blocked the flow of oxygen to her midbrain in a catastrophically precise fashion. Given the rather barbaric state of our medical ethics her family was confronted with the choice either to leave her on life support for an indefinite future, or to allow her to starve slowly to death. Newspapers headlined the story for years while the case dragged through the courts, until at last, mercifully perhaps, she was allowed to die.

Postmortem analysis of Karen Quinlan's brain has now been published in the *New England Journal of Medicine,* showing that the crucial damage involved the loss of quite a small part of each thalamus. Additional damage was found, but in regions that do not seem to affect consciousness directly. Consider Figure 1-2.

Insert Figure 1-2 about here.

Imagine a giant dome divided into two equal halves, the two cerebral hemispheres. Each half-dome has an egg-shaped body lying on one side of the midline. That is the thalamus, truly a minibrain within the brain. The major centers of the thalamus, the nuclei, make up a microcosm of the great hemispheres, each one intimately connected with a corresponding area of cortex. On the sensory side, each nucleus mirrors the sensory cortex, and on the output side, each one reflects the motor cortex. So intimately are thalamic centers bound up with corresponding areas of cortex that the prefrontal cortex is anatomically defined as those parts of the frontal cortex that are connected with the forward nuclei of each thalamus.

Now imagine great fiber cables flowing inward from the eyes, the ears and the skin, to major relay centers in the two egg-shaped thalami. On emerging, they radiate in elegant wings of cell fibers to the sensory cortex, located on the surface of each halfdome ---- vision at the rear, touch on top, and hearing on the side of each hemisphere. Motor tracts flow the other way, coming down from cortex, stopping off in each thalamus and then going out to the muscles.

To complete the picture, take a close look at each thalamic egg and notice that deep inside, sandwiched between the massive transfer stations for the sensory and motor tracts, is are a few small structures: they look like a few small pockets inside the walls that separate the main divisions of the thalamus. These small islands are called the intralaminar nuclei (ILN), because they fit inside the *laminae*, the layers of white tissue that separate the major thalamic nuclei. Neurons in each set of ILN send a fine, widely projecting spray of fibers to all parts of the great cortical dome above. Through these fibers the ILN cells trigger *cortical arousal,* the distinctive electrical signature of waking consciousness.

Neurosurgeon and researcher Joseph Bogen has pointed out that damage to both right and left ILN causes complete coma, indicating that the two sets of intralaminar nuclei constitute a *necessary condition* for waking consciousness. Aside from the reticular formation, this is the only part of the brain that seems indispensible for waking consciousness. (Figure 1-2).

Karen Quinlan's great misfortune was to fall victim to a stroke that blocked the blood supply to *both* ILN clusters on either side of the brain. Loss of the intralaminar nuclei on one side can be sustained, sometimes with temporary coma; we can be conscious with just one healthy side, just as we can breathe with only one lung. But damage to both sets of intralaminar nuclei does occur, because by some great flaw in the brain's design, a single artery in the center of the brain divides in two to supply life-giving oxygen to *both* ILNs on either side of the midline. If that single artery is interrupted before it divides in two, both sets of intralaminar nuclei may die for lack of oxygen. A stroke at this strategic junction can wreak havoc very quickly: several minutes of low oxygen may cause irreversible damage. Bogen estimates that the fatal damage to either ILN need to be no bigger than a pencil eraser.

A dramatic comparison can be made between two kinds of brain damage, one in the core brain (the ILN and the reticular core of the brain stem), and the other, in the massive cortical mantle that wholly covers the human brain. There is a fairly common operation called a *hemispherectomy,* which involves removal of an entire hemisphere, half of the cortex. It is a drastic measure, but when a tumor has deeply invaded one side of the brain, cutting out one whole hemisphere can save the patient's life and affect a cure of sorts. Hemispherectomies obviously affect the contents of consciousness on the side of the cut, though some function is taken up by the other side. *But as massive as they are, they do not lead to coma* .

In sum, losing tiny areas in the core of the brain leads to a loss of waking consciousness, and massive damage elsewhere does not. This is exactly the kind of "contrastive case" that allows us to think about consciousness as a variable. Consciousness is not some fuzzy or indeterminable business at the level of the brain. It requires the small ILN centers and the reticular formation in core of the brainstem.

There are many unanswered questions. What do the ILN clusters do? Many neuroscientists believe that the fibers emerging from the ILN to cortex seem too sparse to support a full, conscious experience of a symphony orchestra or a cheering football stadium. One hypothesis is that the ILN generates a regular waveform about 40 times per second, which may serve to coordinate and "bind" many specific areas of cortex into a single, conscious experience. Neuroscientist Rodolfo Llinás has become a major advocate of this hypothesis. A related idea is that ILN signals may facilitate re-entrant loops, cycling back and forth between the sensory cortex and the sensory relay stations of the thalamus, looping and building on itself until its activity rises above the threshold of conscious activity and drowns out the uncoordinated firing of other neurons. Such a coherent re-entrant loop may correspond to a conscious experience, according to Nobelist Gerald Edelman and others. Both the 40 hz hypothesis and the reentrant loop notion may be true. They are viable hypotheses today, which do not yet command a consensus. They may be definitively accepted or rejected over the next few years.

We are always half-ignorant, of course, never yet at the point where we understand the most interesting issues. But we now know enough, I believe, to narrow down the possibilities; even to make some firm claims about some of the long-standing debates. Among those is the question of animal consciousness.

Are animals conscious?

Just to illustrate how clearly the evidence has begun to accumulate on some hotly debated issues, consider the question of animal consciousness. It has raged for centuries, from Aristotle's proclamation that animals have a kind of soul (hence the word animal, from *anima*, soul), to Descartes' more exclusive view that our animal cousins are merely automata. We know that the brain evolved cumulatively, each new layer on top of ancestral layers that are mostly unchanged. All vertebrates have brainstems, required for wakefulness; all mammals have sizable midbrains on top of the brainstem, containing the thalamus and its ILN clusters, again a requirement for waking consciousness; and all mammals have perceptual cortex, mushrooming over these older structures, which we need to experience sensory consciousness. (Figure 1-3)

Insert Figure 1-3 about here.

Now let us forget for a moment our self-serving wish to have the human species stand out as utterly different from other animals. Pretend to observe life on earth with the cold gaze of a visitor from another star system. What do you notice about consciousness?

Only one species on earth is *articulate* about its experience, having devoted the last two millennia to writing about it. Yet the brain anatomy that supports the *state* of consciousness in this most talkative species can also be found in other mammals. The brainstem centers involved with sleep, wakefulness, and emergency wake-up calls are much the same in our brains as they were in ancestral reptilia, and the resulting electrical activity of the brain looks identical. Species closer to us, including all mammals and birds, have midbrain thalami in addition to the reptilian brainstem. As for sensory cortex, the visual cortex in macaque monkeys resembles the human visual brain so closely that macaques are routinely studied as a source of information about human vision. Thus *perceptual consciousness* appears to be quite similar in monkeys, apes, and humans. Dogs and cats lack color vision, but they probably have a more acute sense of smell.

The behavior we show when we are awake is not all that different from other animals. All animals engage in purposeful action when their brains show the electrical activity of waking, seeking food, mates and the company of others. All mammals snarl when threatened and writhe when wounded, and while we cannot tell directly whether they relish eating, any pet owner can testify to the frenzied excitement of hungry dogs and cats before a meal, their apparent eagerness to eat, and all the signs of lazy satisfaction afterwars. Animals seem to work as hard as humans do to obtain food when hungry, and sexual contact when in heat. All mammalian mothers protect and suckle their young, and both sexes engage in the eternal dance of mate selection. Other animals investigate novel and biologically significant stimuli as we do, ignore old and uninteresting events just like us, and share our limited capacity for incoming information.

Do animals show all the observable aspects of consciousness? The biological evidence is a clear yes. Are they then likely to share the subjective side as well? Given the long and growing list of similarities, the weight of evidence, it seems to me, is inexorably moving toward yes.

Is there still controversy about animal consciousness? My sense is that the scientific community has now swung decisively in its favor. The basic facts have come home at last. We are not the only conscious beings on earth.

Human uniqueness.

Human beings *are* different of course. We have mastered spoken language, a single biological adaptation that is crucial for human culture, technology, and mass society. Indeed, humans seem to spend most of our waking hours silently talking to ourselves. We also come equipped with a uniquely large frontal cortex, useful for long-term planning, abstract thought, and voluntary executive functions. Purposeful use of mental imagery is likely to be limited to humans. All these uniquely human capacities make use of consciousness, even as they extend its reach. Human consciousness seems to differ from that of other animals both in its *contents* and in its capacities, such as our ability to pay long-term, purposeful attention over months or years to a remotely apprehended purpose, that is, to control what we are conscious of in pursuit of long-term goals.

But we can no longer pose absolute barriers between ourselves and other animals. As always in this Age of Darwin, we must come to terms with the fact that we are half animal, half human. The Cartesian notion that humans are the only conscious species flies in the face of the evidence.

We can already see some clear results from contrastive phenomenology. Some of these results may not sit well with some critics. But that is not a *scientific* problem. The science speaks for itself.

But is it the real thing?

Some philosophers maintain that the experiential descriptions we have collected in a century of sensory science may *parallel* conscious experience without actually resulting from it. Prominent philosophers of science like Ned Block of MIT have recently advanced arguments of this kind. But this seems utterly implausible to an empirical scientist. If the overwhelming majority of people say a pencil is red, if they can match it with other red things and distinguish it from blue and green pencils; if their eyes have red receptors, and color cells in the visual cortex fire a red code, what else would they be having but a conscious experience of red?

There may always be those who maintain that what we learn from contrastive phenomenology has nothing to do with real consciousness. The best reply I can think of is to ask skeptics to try one of the demonstrations in this chapter, and ask, *is this your own experience?* If yes, an honest skeptic should say that we are indeed dealing with genuine consciousness as a valid topic for exploration.

Could the evidence regarding consciousness just be a clever imitation of the real thing? Such arguments remind me of Darwin's most adamant critics, who after years of debate proposed that God created the geological fossil record merely to test our faith. It was a last, desperate move against the evolutionary hypothesis. As an answer, we need only notice that *consciousness as an object of scientific scrutiny fits our personal experience remarkably well*. That is not likely to be a coincidence.

In sum, do we have hard evidence bearing on conscious experience?

Yes.

Further reading:

Readers interested in philosophical issues may want to read Owen Flanagan's *Consciousness Reconsidered,* one of the clearest and most up-to-date overviews of a complex field. A more advanced view is given by John Searle in *The Rediscovery of the Mind,* and Daniel Dennett's *Consciousness Explained.* There are many excellent philosophy books, but most advance a specific, controversial point of view, so that a newcomer will only obtain one of several perspectives.

A good introduction to the brain processes involved in conscious experience is provided by Francis H.C. Crick, *The Astonishing Hypothesis: The Scientific Search for the Soul*, and by Robert Ornstein and Richard F. Thompson, *The Amazing Brain.* William H. Calvin and George A. Ojemann have presented an excellent account from the perspective of a brain surgeon in *Conversations with Neil's Brain.*

A more advanced effort is Gerald M. Edelman's *The Remembered Present: A Biological Theory of Consciousness.* Recent breakthroughs in brain imagery are showcased in a remarkable work called *Imaging the Brain*, written by two wonderful scientific pioneers, Michael Posner and Marcus Raichle. It has extraordinary snapshots of brain activity during simple mental tasks, based on PET scans, many obtained for the first time in history.

Remarkably, the best source on the psychology of consciousness (and just about everything else) is still William James's *Principles of Psychology,* first published in 1890, but never out of print. His *Psychology: Briefer Version* is a short introduction based on the 1,300 pages of the great work.

James's thought must be understood in historical context, but the phenomena have not changed one bit.

Part II.

A unified image.

We have so much evidence today about conscious experience that a single integrative image is helpful to keep it all in place. The theater metaphor of mind is both ancient and modern. Plato and Aristotle used it, as did the Vedanta philosophers and William James. Modern researchers who have developed scientific theater models include Herbert A. Simon, Alan Newell, and John R. Anderson in cognitive science, and Francis Crick in neurobiology. I.P. Pavlov referred to the “bright spot” in cortex which integrated all sensory input into one united activity. All these views seem to reflect a similar set of insights. In this section we explore the image of the theater of consciousness.

Chapter 2

The theater stage has *limited capacity* but creates *vast access.*

,,, can this cockpit hold
The vasty fields of France? or may we cram
within this wooden O the very casques
that did affright the air at Agincourt?
Piece out our imperfections with your thoughts;
Think, when we talk of horses, that you see them
Printing their proud hoofs i' the earth;
For 'tis your thoughts that must now deck our kings,
Turning the accomplishment of many years
Into an hourglass...

--- Shakespeare, *Henry V,* Prologue

Let your eyes sweep along a great canvas and the richness of detail may seem overwhelming. But to really "see" a fine painting we need to see it thoughtfully, many times. The reason is that the capacity of conscious contents at any single moment is suprisingly limited: we cannot read this sentence and listen to a conversation at the same time. Nor can we pay attention to the s p e l l i n g of a word, without taking a risk that we will miss some of its meaning.

In fact, our sensory systems pick up only a little bit of conscious information at any moment. If you close one eye and focus steadily on the "+" sign (below) from about eight inches away, it becomes difficult to see the surrounding numbers. Try it, without letting your focus wander:

5	4	3	2	1	○	+	1	2	3	4

High-resolution vision uses the tiny patch of retina called the fovea, packed with light receptors, which we point at the world when we want to see something. The fovea is limited to about 4 degrees of visual arc --- a tiny keyhole of clarity in a very fuzzy visual field. If you stretch out your arm in front and hold up two fingers, they will subtend about 4 degrees of visual arc. We all have the illusion of seeing far more, because the brain cleverly takes foveal snapshots of high information regions in the visual scene, and fills in the rest with plausible guesswork.

Take another look at the last word above, "guesswork": you can become conscious of it with a single foveal glance. It is one of about 100,000 words in your recognition vocabulary. Though "guesswork" is a *relatively rare word, most readers will grasp its meaning in a fraction of a second. Merely being conscious of a target word seems to trigger a search through a mental lexicon to access its meaning; but searching the lexicon is of course unconscious. Once aware of its meaning we immediately gain access to many other abilities; we can define "guesswork," generate paraphrases, and distinguish it from subtly different words. The details of all these abilities are of course unconscious. We can also use the word in an endless array of grammatical sentences, control the high-speed movements of hundreds of muscles in the vocal tract that are needed to pronounce it, and detect errors at any level, including spelling, pronunciation and usage. In sum, conscious access to a tiny visual keyhole allows "guesswork" to become conscious, one lonely little word out of hundreds on this page; but that brief event creates immediate access to a vast world of knowledge. The payoff for limited capacity seems to be vast access.*

Consciousness is the private arena in which we live our lives. Many of us maintain a coherent narrative, some framework in which growing up, adulthood, and old age make up the parts of a meaningful tale. Traditional cultures provide plot patterns for such stories, opening and closing doors that every male and female is expected to encounter, with ceremonial rites of passage to tell the story to the clan. Our inner life narrative seems to happen within some such framework, even though modern culture creates far fewer occasions to give a symbolic accounting of oneself to a chorus of public witnesses. Sometimes we seem to address an inner jury instead.

Here is a version of the theater model that allows us to think about the evidence in a unified way. (Footnote 1) Imagine entering a theater just before the beginning of the show, and seeing the stage, the chatting audience, and a few side doors leading backstage. As the house lights begin to dim and the audience falls silent, a single spotlight pierces the descending darkness, until only one bright spot, shining on stage, remains visible. You know that the audience, actors, stage hands and spotlight operators are there, working together under invisible direction and guided by an unknown script, to present the flow of visible events. As the house lights dim, only the focal contents of consciousness remain. Everything else is in darkness.

Figure 2-1 about here.

----

1. *Working memory is like a theater stage.* All unified models of the mind have a small "working memory" that is closely associated with conscious processes. Working memory is that inner domain in which we can rehearse telephone numbers to ourselves, or, more interestingly, in which we carry on the narrative of our lives. It is usually thought to include *inner speech and visual imagery.*

Inner speech is what you hear yourself saying while silently reading a difficult passage; or after a heated argument with a friend, when you remind yourself of all the good points you should have made but couldn't think of at the time. It has an *speaking* and a *hearing and comprehension* component. Verbal working memory appears to make use of the same parts of cortex that are involved in outer speaking and listening.

Visual imagery helps us to display and solve spatial problems. Most people use it spontaneously without realizing that they are doing so. If you try for a moment to imagine the streets you see coming home from work, and remind yourself to stop at the local grocery store to buy some milk for the cat ... can you see it? That is the mind's eye. It, too, seems to involve the same parts of cortex we use to see the outer world.

Over the last few decades we have gathered a great body of solid evidence about the verbal and visual components of working memory, but one thing has not changed: Working memory is remarkably *limited*. We can keep seven unrelated items in the verbal part, and in the mind's eye perhaps four.

All working memories operate serially, one thing at a time. They show a *stream* of events, just as the spotlight shining on the stage of a theater may show each actor speaking to the audience, one at a time.

2. *The players on stage: the contents of conscious experience.* Consciousness shows *competition and cooperation* between different "players," the different sources of conscious experience that try to reach the stage. Listening to a conversation in a crowded party is one example; it is helpful at a loud party to combine visual and sound information by reading the lips of speakers you want to hear. This is a case of cooperation between two channels of information. But watch a film with a soundtrack lagging half second behind an actor's mouth movements, and the eye and ear begin to compete. One of the two sources inevitably wins at any given moment. Our conscious experience is always a coherent mix of cooperating elements; incompatible input is simply excluded from consciousness.

Our capacity for conscious contents is radically small: Essentially we can take in only one dense train of conscious contents at a time. Fortunately, we can pack a vast amount of information into that single flow of experience, and sometimes, as in talking to a friend while driving, we can jump back and forth between different streams.

The limited capacity of consciousness is one reason for the theater metaphor. Like other scientific analogies it allows us to package a large amount of evidence into a simple, organized image. As you read this book you may notice that when we speak of the *theater of consciousness,* your mental image will become more and more densely packed and elaborated, but you will still be able to understand it as a unit.

3. *The spotlight of attention. Conscious contents emerge when the bright spotlight of attention falls on a player on the stage of working memory. But the spotlight has a fringe ...*

To illustrate, look at the following numbers carefully. Ready?

11, 23, 4, 61, 3, 17

Now close your eyes for about ten or twenty seconds, and then write the numbers down. What was conscious when you closed your eyes? Most people report that at any given moment, whatever number is *being said* in their inner speech is conscious; numbers that are momentarily not being said are not. There seems to be a clear phenomenological contrast between currently conscious and briefly stored numbers. Try it a few times to see if you agree. If the results are reliable and if we can verify them independently, we generally take them to be valid.

The spotlight of attention has a crucial role in our version of the theater metaphor, for whenever it falls on some particular actor he or she comes to consciousness. Actors in the spotlight are privileged in the theater metaphor. They are the only ones who can disseminate information to an audience of specialized experts --- who represent the unconscious resources of memory, knowledge, and automatic routines. The audience in turn may hiss or applaud, asking to hear more or less from any given actor. Audience members can also exchange information among themselves and form coalitions to bring other messages to the stage. But there is only one way to reach the audience as a whole, and that is by way of an actor in the spotlight on stage.

What is *fringe consciousness?* Imagine that the bright spot on stage is surrounded by a darker penumbra, to represent a very interesting phenomenon that William James called *fringe consciousness.* If we take focal consciousness to include immediate, detailed experience, the fringe would cover those cases in which we have reliable access to information without being able to experience it explicitly in detail. Cognitive scientist Bruce Mangan has helped revive a long philosophical tradition about the fringe, including such experiences as feelings of knowing, of familiarity, beauty and goodness, of something not quite fitting, or a sudden profound feeling of rightness. A surprising amount of our mental life is occupied with fringe events, which may be experienced as fuzzy or vague, but which have properties suggesting that something very precise is going on.

Take the "feeling of knowing" that comes when we ask a question like, *What is the name of the flying reptiles of the dinosaur age?* Most of us have trouble finding the answer right away, but we know that we know it, and rightly so. Feelings of knowing have been studied quite a bit, and the evidence indicates that (1) they are quite accurate most of the time; (2) they receive high confidence ratings; but (3) they do not involve detailed, structured experiences, unlike the sight of a coffee cup, where we can talk about shape, color, shading, texture, and many other details.

We have feelings of knowing about items in working memory that are not currently conscious. Moreover, we seem to have feelings of knowing about things that are readily *available* to consciousness though they are not conscious at the moment --- Our ability to find known words, our mood, our ability to act and control some mental functions, our basic knowledge about friends, relatives and ourselves, and much more.

4. *Contexts operate behind the scenes to shape events on stage.* THE THEATER HAS ~~This point involves~~ unconscious *contextual operators* that set the background against which the brightly lit actors play their roles. They are invisible behind the scenes. but have profound effects on consciousness. Your experience at this moment is shaped by unconscious expectations about the syntax of this sentence, as well as by conscious thoughts you may have had years ago about human consciousness. Context effects are ubiquitous and powerful.

Among these behind-the-scenes operators are a *director,* which performs executive functions --- what William James called “the self as agent and observer.” *We* ---- whoever *we* are --- have voluntary control over parts of working memory. At times we can decide what will come into consciousness next, and whether to interrupt the current stream if something more urgent happens. Voluntary functions involve frontal cortex, and injury to frontal cortex often degrades executive control. (See Chapter 7).

Finally, we have

5. *The audience.* This is what it’s all about; it is the *raison d’être* of the whole design. We do so many things unconsciously, and the neuronal

networks that perform unconscious functions are so widely distributed throughout the brain, that the notion of a vast *society* of specialized systems has become very natural. If we think of the brain as a distributed system with millions of specialized abilities, the question becomes, how do you recruit and mobilize all of the specialized unconscious networks in pursuit of survival and reproduction? This is presumably why the unconscious society of the brain requires a stage, a spotlight, and a director. Consciousness, in this view, serves to disseminate a small amount of information to a vast unconscious audience in the brain. It is the publicity organ in the society of mind.

It is terribly important to keep in mind that audience members are merely metaphorical. Automatic routines tend to be relatively separate, specialized and autonomous; but in fact they work together with others to carry out the details of even the simplest action. Audience members, if you will, have a vast hardwired telephone network connecting each to each, enabling them to carry out routine tasks without consciousness. Consciousness is not needed for many of these routine collaborative activities; it seems to be needed for new combinations of ingredients. It is likely that today's routine collaborations between separate automatic units were created in the past with the help of consciousness. Today's automatic processes emerged from yesterday's effortful and elaborately conscious projects.

All unified models of cognition today suggest some sort of unconscious audience, including unconscious memory archives and automatic routines that are triggered when their "calling conditions" appear in working memory. In the brain the audience seems to consist of functional networks and routines --- collections of neurons that work together to perform some job. We can think of them as people sitting in the dark audience, unconscious but with great local expertise.

Some audience members specialize in memory, including the 100,000-word lexicon of English that you are using at this very moment; your autobiographical memory, a noisy record of a lifetime of conscious experiences; implicit memory, the patterns, skills, and regularities you have learned since birth; semantic memory, the knowledge that is needed to understand this paragraph; and declarative memory, which includes explicit beliefs about the world and oneself.

Here is an example of an unconscious, automatic process. Try to read the words below *without* saying them to yourself:

inchoate
Pappa Doc
infundibulum

I cannot avoid sounding them in my inner speech, no matter how hard I try. The mere visual experience of the printed word seems to trigger auditory experiences. This is one example of the automatisms that control our delicate and purposeful eye movements, linguistic analyses, visual scene interpretation, bodily posture, and all the moment-to-moment inferences we make when we read or see or hear.

In many ways the audience acts as a legislature. Audience members may hiss or applaud certain messages from the stage, and they may build coalitions to help their favorite actors compete against others for access to the stage. Different impulses may try to become conscious in a dynamic game of king of the hill, as we can guess by observing the impulsive actions of young children. Impulsive goals may eventually coalesce into working hierarchies that support a coherent series of actions. The adult self may involve a more settled version of the child's impulsive, short-term goals.

A great body of evidence can be understood in terms of these six basic ideas --- the stage, the bright light of attention, actors and their speeches, the audience, contexts, and director. The actors in the spotlight may fret and strut their hour upon the stage, cued by the director against a background created by context setters. The spotlight selects the most important events on stage, which are then distributed to an audience of all the unconscious routines and knowledge sources in the hall.

These are the basic ingredients of the theater; a few ideas to limn a vast biological brain in broad outline. If our metaphor seems puny in the face of that immense reality, that is surely true. The question is, can this simple set of ideas capture some basic set of facts about human experience? Let us see.

Theater diagrams.

It is helpful to visualize new ideas, and the theater metaphor is no exception. We can show its essentials with a few lines and circles, allowing us to sketch various ways the theater might work in the following chapters. (Figure 2-2) Theater diagrams are simplicity itself. The stage is a small rectangle, with a small oval to symbolize the spotlight. Actors are shown as arrows going into the spotlight. The contexts that operate backstage are simple horizontal brackets, and the audience is just a collection of little circles. If we want to see a stream of conscious events over time we need only shrink these basic symbols and draw a series of boxes with a time arrow, as shown.

Insert Figure 2-2 about here.

Two different time scales are known to be important. The time needed for conscious perceptual integration is on the order of one-tenth of a second. If two identical pictures are presented to the two eyes more than .1 second apart they will not fuse into a single image; but time differences of less than .1 second experienced as a whole. That scale is shown in Figure 2-2 (a).

However, working memory plays out on a scale of seconds, as you can tell simply by counting to yourself in inner speech. Various measures show the duration of working memory to be about ten seconds without rehearsal. This is the time scale on which we talk to ourselves and make thoughtful decisions. This more common sense grain size of experience is shown in Figure 2-2 (b).

Notice, by the way, that today's neuroimaging techniques do not have the temporal resolution needed to show what is happening at these time scales, though improvements are coming so fast that we should be able to see brain activities over seconds, if not milliseconds, within a decade. A complete understanding of our experience must be able resolve brain events happening over seconds and tenths of seconds.

Spontaneous problem solving: That little pause before the answer comes to mind.

Consider the following questions:

(1) What is the name of a herbivorous dinosaur?

(2) What technology develops artificial limbs?

(3) What are three synonyms for "talkative"?

Did you experience a brief pause before the answer came to mind? (If you are not sure, try it one more time.) That brief "dead time," of simply waiting for the unconscious to do its job of finding the answer, is typical of spontaneous problem solving. It starts with a conscious moment when the question is asked or the problem assigned; followed by some period of unconscious incubation; and ends with a return of the answer to consciousness. We certainly don't need to be deliberately trying to solve the problem. All we need is some incomplete .... , to start a problem-solving process.

Spontaneous problem-solving can cover lifetime issues, fantasies about the future, finding a memory of elementary school, control of one's own body, searching for the right word at the right time, understanding a sentence, trying to influence other people, and an endless variety of other things people think about.

Figure 2-3 shows a theater diagram for the principal elements of spontaneous problem-solving: conscious problem priming, unconscious problem solving, and conscious display of the solution. Or, if we want to be a little more complicated, we can show a goal that is achieved by way of a number of sub-goals. This is much more realistic in examples like mental arithmetic; multiplying 324 x 11 has the subgoal of multiplying 324 x 10. Each subgoal seems to have the same three-stage format as a top level goal.

Insert Figure 2-3 about here.

Creative processes in art, science and mathematics seem to show the pattern of conscious assignment, unconscious processing, and conscious emergence of the answer. But so do short-term events like word-search, question-answering, the interpretation of ambiguous words, action control and the like. The stream of consciousness may be a flow of experiences created by the interplay of many goals, each tending to make conscious whatever will promote its progress. When all these spontaneous problem-solving patterns are intertwined, the stream of consciousness may *seem* random and without purpose. Researchers like Jerome Singer report that on closer examination, spontaneous thought seems to contain many unresolved problems, to which we return time and time again until we hit on a satisfying answer.

Notice how much we need to trust the competence and creativity of the unconscious. Chances are that unconscious incubation makes use of all the highly practiced automatisms that we have thought about consciously over our lifetimes. Obviously word search is an unconscious type of problem-solving in the mental lexicon, but it can include all the words we have paid attention to in our lives. It seems that the unconscious mechanisms that are quietly buzzing away before the answer is returned are themselves the working residues of earlier conscious thoughts.

The role of unconscious problem-solving was described a century ago by the mathematician Henri Poincaré, who devoted much thought to the psychology of mathematical creation. He wrote,

> Most striking at first is this appearance of sudden illumination, a manifest sign of long, unconscious prior work. The role of this unconscious work in mathematical invention appears to me incontestable, and traces of it would be found in other cases where it is less evident. Often when one works at a hard question, nothing good is accomplished at the first attack. Then one takes a rest, longer or shorter, and sits down anew to the work. During the first half-hour, as before, nothing is found, and then all of a sudden the decisive idea presents itself to the mind. ... There is another remark to be made about the conditions of this unconscious work: it is possible, and of a certainly it is only fruitful, if it is on the one hand preceded and on the other hand followed by a period of conscious work.

This is also emphasized by the poet Amy Lowell:

> How carefully and precisely the subconscious mind functions, I have often been a witness to in my own work. An idea will come into my head for no apparent reason; 'The Bronze Horses,' for instance. I registered horses as a good subject for a poem; and, having so registered them, I consciously thought no more about the matter. But

> what I had really done was to drop my subject into the subconscious, much as one drops a letter into the mailbox. Six months later, the words of the poem began to come into my head, the poem --- to use my private vocabulary --- was 'there'.

Of course creative people are often conscious of some intermediate events in this process. And not all creative work is experienced as spontaneous --- some of it is deliberate hard work. This mixture of ingredients goes to make up a completed piece. Listen to Mozart,

> When I am, as it were, completely myself, entirely alone, and of good cheer ... my ideas flow best and most abundantly. *Whence* and *how* they come, I know not; nor can I force them. Those ideas that please me I retain in memory ... If I continue in this way, it soon occurs to me how I may turn this or that morsel to account, so as to make a good dish of it, that is to say, agreeably to the rules of counterpoint, to the peculiarities of the various instruments, etc.

A major creative work is not accomplished in a single conscious-unconscious-conscious leap. Mathematical creation requires a long string of insights and new problem assignments, most of them minor, and only a few dramatic enough to stay in memory. Poincaré may simply have forgotten some intermediate events between the first effortful period of conscious problem-assignment and his memorable Aha! experience. Most problem-solving requires multiple dives in and out of the stream of consciousness.

Convergence and divergence.

Here is perhaps the single most important feature of the theater image: As Figure 2-1 shows, it combines *convergent input* with *divergent output.* Onto the stage converge the competing actors and their speeches, the makeup artists and scene designers, the playwrights, directors and acting coaches. Whatever comes to mind reflects a compromise between competition and cooperation, fusing whatever is compatible and excluding for that moment anything that is not. Every dramatic moment, each syllable spoken on stage reflects this convergence of input. Yet as soon as a syllable is pronounced (and this is the aim of the whole enterprise, of course) it floats out to the audience with effects that are largely unknown, but which depend on each individual listener, who makes of it whatever they will.

Hyppolite Taine, a French historian of the nineteenth century, emphasized these features:

> One can compare the mind of a man to a theater of indefinite depth whose apron is very narrow but whose stage becomes larger away from the apron. On this lighted apron [i.e. front of the stage] there is room for one actor only. He enters, gestures for a moment, and leaves; another arrives, then another, and so on ... Among the scenery and on the far-off stage ... . unknown evolutions take place incessantly among this crowd of actors of every kind, to furnish the stars who pass before our eyes one by one, as in a magic lantern.

Taine suggested that only one actor at a time can come to the footlights, a fact that the nineteenth century called the “narrowness of consciousness,” and which we call “limited capacity.” It is a very robust finding indeed. “Unknown evolutions take place incessantly” in the far off, invisible part of the stage, perhaps behind the scenes, “to furnish the stars who pass before our eyes one by one, as in a magic lantern.”

A message is broadcast globally, but it is interpreted locally in the mind of each audience member; the director and playwright, listening backstage, are also taking in global messages to guide the next performance. In sum, there is massive convergence of information onto the stage, but once it has come together there, it flows divergently to the audience.

In the next few chapters we will explore the theater metaphor. It is a simple arrangement that we can see working in daily life. This is how classrooms are arranged, and legislative bodies, and scientific conferences. But the metaphor is only a scaffolding, to be discarded without regret once it has outlived its useful days.

Scientific metaphors.

Aren't metaphors pretty crude for thinking about difficult scientific problems? Actually, they have a long and honorable history in science as tools for making the perilous leap from the known to the unknown. The clockwork metaphor of the solar system was of great help to astronomers in the Sixteenth century as a way to think about the swirling interplay among sun, earth and moon. About 1900 physicists found the Rutherford model of the atom as a tiny solar system useful for generating testable hypotheses. Darwinian evolution was a powerful qualitative metaphor for its first century of existence, and it is often still used so today. Many scientific theories begin in this humble fashion.

Perhaps the best example in the history of science is William

Harvey's idea in the 1630's that the heart acts as a pump, pushing blood around the body. Harvey's insight challenged a medical tradition going back two thousand years to Galen and Hippocrates. In his time medicine was just beginning to gain an understanding of the body as a great assemblage of cells that all required nutrients, and the notion of the "heart as a pump" cast of a great shaft of light into a poorly understood corner of reality. It is a functional metaphor, of course, describing how the heart might act serve a certain purpose by propelling the blood stream through the veins and arteries. Like Harvey's pump metaphor, the theater metaphor describes how conscious and unconscious processes might serve a certain set of functions. There is no literal theater in the brain any more than the heart has a little windmill carrying tiny scoops of liquid over a dike, as wind-driven pumps moved water in Harvey's seventeenth century. The theater image is just a way to describe the flow of information in the brain.

Scientific metaphors are scaffoldings that help us to grasp and simplify complicated problems. But scaffoldings are useless without real bricks and concrete, and these models are mere doodles on a napkin in the absence of a large constellation of specific research findings: studies of the cellular wiring of the visual cortex, for example, and careful analyses of the remarkable effects of selective brain damage, the role of perception in action control, a century of psychophysical studies, research on memory, imagery, language, and much more. Those results are indispensible for any unifying theory. Fortunately, we have a wealth of information today that allows us to fill in many of those gaps. Like any useful framework, the theater metaphor will point to new openings, new questions that can now be asked in a clearer way.

Like any metaphor, the theater architecture is only useful up to a point; we will keep track of its flaws as well as its uses as we go along. Yet it does provide a useful starting point. In the following chapters we use it to think about a large body of evidence about consciousness. We will not treat the theater metaphor as theory, though all current integrative theories can be thought of as theaters. We will use it just to simplify the evidence.

Limited capacity and vast access: two views of the same mountain.

There is an amazing difference in the way psychologists and brain scientists have looked at the brain. Psychologists traditionally see a nervous system that can do only one thing at a time, and which seems to do fairly simple things like mental arithmetic with high rates of errors and a great deal of annoying interference. To the psychologist we know only *a single coherent event* in each moment --- a visual scene, a mental image, or a fleeting thought. We cannot do two conscious things at a time, such as carrying on an intense conversation and driving in busy traffic. If we don't need to think much about talking, we can drive at the same time, and vice versa; but the more conscious involvement is needed for each task, the more they will compete against each other. Viewed psychologically, the brain appears to solve simple problems in seconds or even minutes; it makes many errors, tends to sequence even things that might be done at the same time, and its efficiency in solving novel problems is not impressive.

The vast, unconscious brain .

From William James to the present psychologists have thought of consciousness in terms of *selectivity,* a reduction in complexity, while the neuroscientist, looking at the nervous system more directly, finds plentiful evidence for global brain *arousal* but much less for selection. Brain scientists see vast orderly forests of neurons, each receiving input from thousands of others via bushy dendritic twigs and branches, and each with a single output called an axon. Neurons send electrical pulses an average of forty times per second through their axons, and they are all active at the same time. The brain is massively parallel, largely unconscious in its details, and widely decentralized in any given task.

At the level of the cells, a structure like the cerebral cortex is immense --- a vast, looming starship unto itself, containing by recent estimates 30 to 55 billion neurons. Seen from the outside it is an elaborately folded structure with many hills and valleys, neatly tucked into the upper half of the cranial cavity. If we could carefully unfold the great cortical mantle we would see a sheet about three feet square, with six layers, each composed of myriads of bushy neurons surrounded by supportive cells. This layered sandwich can be parsed into millions of vertical columns and clusters of columns, so that we can imagine a vast six-layered array in three dimensions. Each layer seems to specialize in input, output, or internal connections in the cortex.

Cortical neurons are connected by vast tracts of axonal cables, wrapped in a white sheath called myelin. If we simply slice through cortex we see mostly white matter, an indication of how many connective cables there are. With the nake eye cortex lookS like a cheese cake covered with a thin layer of cell bodies, the gray matter that we see from the outside. But the white matter containes miles of tiny fibers that descend from the gray cell bodies and end up coming back to cortex on the opposite side of the brain, while an equal number OF FIBERS loop beneath THE cortex and come back up on the same side. The uppermost layer of the six-layered sandwich, Layer I, is so densely woven horizontally that it has been called a *feltwork,* a large slice of tight webbing on top of the sandwich. While most long-distance communication in cortex seems to run through long vertical output fibers (axons), the top layer is so tightly interconnected horizontally that many brain scientists believe there is also a great deal of spread within Layer I. Recent evidence suggests that in visual cortex at least, the *innermost* layer of cortex may be the home of conscious sensations.

Cortical neurons project in vast elegant fiber bundles to the neural organs nestled tightly under the cortex, like golf balls beneath a baseball glove. Among the golf balls the *thalamus* serves as the great traffic hub, the way station for all messages going in and out from cortex, and therefore a strategic control point.

Other large pathways project from one hemisphere to the opposite one in mirror-image symmetry, or hang suspended in great fiber bands beneath each hemisphere, mapping more frontal and more posterior neurons. Most communication between neurons takes place by means of massive fiber bundles. In fact, most of the cortex consist of "white matter," long fibers wrapped in a protective coating of why myelin, which gives the inside of the cortical mantle the appearance of cream cheese. The traffic through these fiber bundles is dense. Current estimates for the left-to-right fibers that cross in the corpus callosum --- the fiber bridge that is cut to create split brain syndrome --- is on the order of 600 million. Each of these fibers send an electrochemical message from 40 to almost 1,000 times per second, making for message traffic ranging from two to 600 billion events per second. The whole elegant arrangement obviously limns some regular and mysterious symmetry, if we only knew what it might be.

When we become conscious in the morning the brain is globally activated, every part showing a sharp increase in the speed and complexity of neural activity. As novel or surprising events catch our attention a vast electrical tidal wave rushes all over the brain, three-tenths of a second after the event. Like so many aspects of consciousness in the brain, these facts of consciousness look not like a reduction but an *increase* in complexity.

Why is conscious capacity so limited?

Why does consciousness seem so limited in a brain with 100 billion neurons? This is a key question that is not asked as often as it should be. It isn't just that we have a limited capacity to *do* things --- only one mouth to speak with and two hands to fiddle with. The conscious limits are particularly strong in perception, the *input* system, so the limitations of hands and mouth are not the reason. And it is not that the brain lacks sheer processing capacity --- its ability to store and transform information is far beyond our current ability to describe. The limits are not in the size or complexity of the brain, but in the fact that we can be conscious of only one unified experience at a time.

If two heads are better than one, why did Mother Nature not find a way to package more than one into a single cranium? This is not just a playful question. In the natural world survival depends on one's ability to drink from a waterhole even while keeping an eye out for predators, making sure all the time that one's offspring don't get lost. It is easy to imagine a host of selective pressures pushing the brain toward an ability to do more than one conscious thing at any given time. But evolution has given us a one-track mind --- to be sure, with lots of unconscious things going on at the same time --- but only one stream of consciousness. Why?

Here is a possibility. Conscious limits seem to reflect trade-offs in the functioning of the brain. Every evolutionary adaptation involves trade-offs of some kind: adding a huge cortex to our primate brains cost a great deal in allocation of oxygen, glucose, and the like. It is thought that the wide hips of mature women reflects the need for the large head of the infant to pass through the birth canal. But the size of the newborn's head also makes birth more difficult and painful than it seems to be in other primates. Our sizable cortex carries heavy costs, and presumably there are balancing biological benefits in our ability to plan, to control ourselves, and to think. The following point may be evolution's consolation prize for the narrow limits of consciousness.

Consciousness is the gateway to the unconscious mind.

A few examples give some sense of the global reach of conscious events. We have mentioned the 100,000 word recognition vocabulary of an educated speaker of English, but that is an underestimate since most word have multiple meanings, to which we gain access automatically when we hear the word in context.

The size of memory is unknown, but we do know that with recognition tasks we can look at a sequence of up to 10,000 distinct pictures one week, and without attempting to memorize them, simply by *paying attention,* we can recognize this week's pictures and distinguish them from a different set a week later, with more than 95% accuracy. That implies that the brain must have stored distinctive information about 10,000 pictures after conscious exposures of perhaps 20 seconds each. It is an awesome accomplishment.

In everyday life we sometimes encounter this large access to memory when we recognize a film seen only once, perhaps many years ago, and can even anticipate the next scene. What about all the musical phrases you can recognize as familiar? All the faces, the art objects, the artifacts, the people you have known from birth to the present? The fact that we are always being conscious of *something,* and that mere consciousness of distinctive events leads to excellent recognition memory, suggests that the amount of information stored is large indeed. Not all of it can be retrieved under ordinary conditions; but when we make retrieval easy by presenting the same event we experienced before, the effects of short conscious exposures hint at an indefinitely large storage capacity.

A second example of the wide access of consciousness involves biofeedback training. A few decades ago brain scientists discovered that humans can control a number of neural functions when they are given immediate, conscious feedback. For example, the EEG spontaneously displays a regular alpha-rhythm of eight to twelve cycles per second some of the time, especially over the visual cortex when the eyes are closed. It is possible to set up a computer to detect alpha rhythms and have it emit an easily audible tone whenever alpha waves occur. The surprise was how much control this immediate conscious feedback provided over the alpha rhythm. People can learn to "go into alpha" at will.

The evidence has now mounted that any single nerve cell, or any population of nerve cells in the brain, can come under voluntary control. That includes single cells in cortex, in the thalamus and brain stem, and in the peripheral nerves that pervade all of the body. But entire neural populations can also come under control, such as those in a gigantic nucleus of the thalamus, in a structure called the hippocampus, in parts of cortex. The alpha wave activity mentioned above involves large numbers of neurons. The power of biofeedback training is an extraordinary finding, and one that scientists perhaps disregard too often. The implications are significant, because it suggests that all neurons in the brain can become involved with the conscious, voluntary, limited-capacity system: The part of us that controls most of our actions.

It is not emphasized often enough that biofeedback training always requires *conscious* feedback. To gain control over alpha-waves in the EEG we pay attention to a tone or light corresponding to the increased alpha activity; to gain control over single cortical neurons we play back a conscious click for each peak of electrical activity, and so on. Under such conditions, people can learn to control an extremely wide range of physiological activities with surprising ease. A small needle electrode in the base of the thumb can record the activity of a single motor "unit" --- one muscle fiber controlled by a motor neuron coming from the spinal cord and a sensory one going back to it. When a single neuron fires, and the electrical event is amplified and played back through a loudspeaker, it sounds like a sharp click. A subject can learn to control the click in about 10 minutes, so that it occurs at will, and some have been able to play drumrolls on their single motor unit after 30 minutes of practice. Distraction, subliminal stimulation, habituation, or drowsiness all impede biofeedback learning.

If we keep in mind the fact that you and I have no idea how we control our muscles, a complex and subtle process that is not available to consciousness, the question arises, Who or what is doing the learning in biofeedback? Common sense says that "we" are learning, as if there is some centralized self that is in control of all the details. But that cannot be true, because the everyday "we" has no access to the necessary information. If we think of the brain as a largely decentralized society of biocomputers it may be more realistic to say that the motor system is learning something, *based on conscious information that is made available to many unconscious local control systems.* *We* are not doing the learning; *they* are.

We can draw an analogy between biofeedback training and finding a child lost in a large city. At first we search *locally,* around home or school. But if the child cannot be found there it makes sense to broadcast a global message to all the people in the city: "Have you seen Jerry or Martha?" Only those who have relevant information will answer back. The message is global, but only local experts feed back their information. This is what we would expect if conscious feedback were made available throughout the brain, and local processors decided whether to respond to it.

And this is of course the message of the theater metaphor. That is what we mean when we say that consciousness seems to creates vast *access* to perhaps all parts of the nervous system.

Although potentially it seems that any neural system can come under voluntary control, it would be absurd to try to control heart-rate, or the movements of the intestines, because doing so would take up working memory capacity with all its limitations; and we do not have the conscious knowledge and wisdom needed to run the heart, for example, as well as the body does unconsciously. Most of the business of body and brain are conducted with great skill by unconscious processes without moment-to-moment conscious control. Consider automatic components of action:

talking, listening, viewing a scene, emotional reactions, interpreting a social situation, relating to others, playing chess, driving a car, walking, loving, dancing, teasing, arguing, making peace after an argument. Automatisms can be evoked by a distinctive stimulus: such as the automatic patterns that fire quickly when you see any *known word*, the gestures of a friend, a familiar face from a ten years ago, a signal of instinctive danger. Even years after we have learned to swim, if we are accidentally thrown into water we can recapture in seconds the automatic components of swimming.

Learning as a magical process.

The idea that consciousness is a gateway, something that creates access to a vast unconscious mind, has interesting implications for our understanding of learning. It suggests that the real work of consciousness in learning is simply to point to some information to be learned, with the detailed process of learning taking place unconsciously. It is as if learning occur magically, without effort or deliberate guidance, carried out by some skilled squad of unconscious helpers.

For the last twenty years cognitive scientists have tried to encode the knowledge human experts have about physics, computer programming, and medicine in large computerized semantic networks. The big lesson of those twenty years is that expert knowledge is highly *domain specific,* that is, that visual knowledge, for example, is so different from linguistic knowledge that almost nothing in one area applies to the other. And yet as human beings we do pretty much the same thing with anything we need to learn: we merely bring it to consciousness, and learning somehow occurs. It looks like sheer magic.

The radical simplicity of learning is quite extraordinary. We direct our attention to the formula $x = y + 3$, play with its elements and rules, and somehow, with no *detailed* conscious coding of the information, we acquire the ability to grasp it as a whole with a genuine sense of understanding. We learn to see new visual patterns simply by paying attention to a set of X-ray photos or a series of Flemish still-life paintings. We learn to hear in new ways by merely listening to bird songs or symphonies. The regularities of language are acqured just by paying attention to the sentences we hear.

Yet we know that language activates an utterly different part of the brain than visual events, which are yet different from planning and feeling, fine motor control, learning what foods taste good, and hundreds of other specialized interpretations of conscious information. Paying attention --- becoming conscious of some material --- seems to be the sovereign remedy for learning all these very different kinds of information. It is the universal solvent of the mind.

Children learning language surely don't label the words they hear as nouns or verbs. Rather, they pay attention to speech sounds and the underlying grammar is learned *implicitly.* We rarely become conscious of abstract patterns --- the regularities of grammar, the harmonic progressions of a symphony, or the delicate brushwork of Vermeer. Most knowledge is tacit knowledge; most learning is implicit.

Did evolution discover theaters first?

Theaters are useful. Their fundamental features are found in classrooms, scientific conferences, television sets, public broadcasting, armies, bureaucracies and business organizations. It may be that evolutionary biology discovered the same style of functioning as well. Quite different animals may solve similar problems in similar ways, and human technology occasionally rediscovers one of those biological solutions as well. Computers were invented a billion years after nervous systems, but they have important similarities. The Romans invented the arch eons after evolution discovered the rib, but the principle of strength through arched construction is the same. Homologies like this are the rule, not the exception, and the case of the theater architecture may be just another example.

We have now sketched one way of thinking about our experience. Over the next few chapters we will see how far it will take us. What can we make of the stage, the spotlight, the audience and those mysterious goings-on behind the scenes?

Further Reading.

Allan Newell and Herbert A. Simon's *Human Problem Solving* is still the best source on the first twenty years of cognitive modeling by means of "theater models." More recent works along these lines include Newell's last book, *SOAR: Toward a Unified Theory of Cognition.* John R. Anderson has written several books tracing the development of his unified theory, most recently *The Adaptive Mind.* Perhaps the nicest informal presentation of a theater model is in Donald A. Norman and Peter Lindsay's *Human Information Processing,* an outstanding introductory text that may be difficult to find. The history of the unconscious is thoroughly covered in Henri Ellenberger's *The Discovery of the Unconscious.* These books are rewarding but not easy.