TABLE OF CONTENTS

PART ONE – Mind as Modeling

What is thought? I will argue that thought is the maintenance and manipulation of a working real-time model of reality. By argue, I do not mean to present an argument in the technical sense, but rather, to present a persuasive case. I wish to convince the reader that this concept of thought is a reasonable and useful one, a viewpoint worthy of tentative acceptance and critical examination.

Having experiences

The simplest nervous system amounts to a number of sensitive sensory neurons contacting a number of effecting motor neurons, to give a fixed set of reflex responses to environmental change. The sensory neurons notice certain changes in the environment and trigger particular motor neurons to respond to the changes. The only learning that is possible in this structure is habituation, the temporary loss of a reflex when it is triggered at a high frequency.

A more flexible behavior is possible when inter-neurons are interposed between the sensory set and the motor set. With small numbers of inter-neurons, the behavior is still reflexive but the inter-neurons allow more complex and discriminating patterns of movement. A small amount of non-habituation learning is possible. Further, the inter-neuron net can also maintain spontaneous rhythms. The spinal cord and much of the lower brain stem is such a network. It generates the rhythms of heart beat, breathing and alternate limb movements; it mediates the reflexes of avoidance movements; it maintains postural muscle tone. But its ability to learn is limited.

In order to learn, in the sense of gaining from past experiences, it is necessary to have experiences, to store them and evaluate them. When inter-neurons increase in number and become organized into large sheets and nuclei, it is possible to deal in experience. The inter-neurons function to build a coherent model of experience, store an edited version of this model as a sequential memory, simulate future events with the model and compare results with expectations. Now there is the ability to learn from experience, there is thought.

Being conscious

It is important at this stage to point out that in this concept of thought, there is a difference between the model (and its production and use) and our consciousness of it. The phenomenon of consciousness is associated with the formation of an edited form of the model and the storage of this summary. Consciousness is not the modeling activity itself.

If we ask the question, “How do I know that I am conscious at this very instant of time?” it is difficult to answer with referring to the continuum of past, present and future experience. Our awareness of having been without consciousness is simply a discontinuity in the memory.

Consciousness can be pictured as the leading edge of memory, having the same detail and structure as recent memory. It is one of the finished products of thought, not the process of thought. Far from being uniquely human, consciousness will occur in all animals with brains of the same memory fabricating type, at least the higher vertebrates. The elaborateness of the edited model would dictate the elaborateness of the consciousness of it.

We can examine the model in three ways: one, through its edited form using introspection of our memory and consciousness; two, through its results using behavioural studies; three, through its physical apparatus using studies of neural anatomy and physiology. These approaches will yield different but equally valid information. A successful theory of thought must be congruent with information from all these perspectives.

No sense datum available

The idea of a sense datum as an un-interpreted bit of information registered by the brain is fictitious. The eyes and ears at least, and probably all the senses, transmit highly processed signals to the brain. Further, the brain does not assemble these signals first and then examine them. The brain is concerned with obtaining a ‘best-fit’ scenario of incoming signals with its model of what is happening.

The ear does not transmit details of changes in air pressure or even of the movement of the eardrum. The ear transmits a partial Fourier synthesis of these complex changes in the form of the sine-wave components of these oscillations. Because the ear is tuned to only a portion of the sine-wave spectrum, its signal cannot, even in principle, be used to reconstruct the original air pressure changes; it is in this sense that the ear's analysis is only partial.
The eye does not transmit the amount and frequency of light reflected from each point in space or even of the light falling on each point on the retina. The eye's signal has very enhanced contrast, edge definition and sensitivity to movement. There is a somewhat arbitrary relationship between frequency of light and colour. (There is simply no frequency of light which corresponds to some shades of purple.)
The usefulness of transmissions from the eye is negligible unless they are fully corrected for movements of the eye, head and body to create a stable visual field. This type of correction is also fairly important to the other senses. The intimacy of the integration also produces reflexive orientations of the head towards loud or bright sources. It helps to produce impression of movement when the movement is passive rather than active and certain illusions of movement such as the train-in-the-station type.

Events

At a very early stage in the processing of sensory information two very interesting phenomena occur which give that information a context, spatial and temporal. The context cannot then be confused during the time it takes to fully process the information.
Thus the phenomenon of 'events' is born during this very early processing. Any abrupt change in the sensory signals is labeled with a time reference, so that during and after processing, which will take a significant and variable duration, each event can be referred to its place in the sequence of events.
The sensory signals which enter the higher, model building and perceiving parts of the brain are partially processed to give, for instance, colour or pitch and are labeled with a spatial and temporal reference. These attributes are constraints on any model of reality. We already (probably from birth) have the format of the model: our orientation and movement in a three-dimensional spatial framework over a linear continuum of time marked by sequential events. It can be noted in passing, that this is a long way from the idea of the mind as a blank white paper on which on experience is written. Further, many of the processes that fit sensory information into this format are indicated or limited by the structure of the brain.

Best-fit

Best-fit analysis of sensory information is complex. Depth perception, for example, results from: the matching of two images with a parallax error between them, an attempt to preserve size consistency, Gestalt construction of incomplete contours and object-ground relationships, construction of molding effects using colour and textural consistency, elimination of discontinuities between the present image and the immediately previous one, a matching of possible perceptions to our memory of the same or similar experiences. The process produces the 'best-fit' possible given these constraints and probably many others you to be discovered. The perception we see is the interpretation with the least ambiguity and surprise. When the sensory information is scant and ambiguous we are subject to illusions. But although we may 'know' them to be false, we still 'see' them as the best fit available.
It is important to differentiate between our visual depth perception and the three-dimensional framework of our model. It is not the fitting of our perception of depth into our model that produces space. It is the spatial framework that makes depth perception possible. If we are blind, we still perceive in terms of the same spatial format but use acoustics and tactile/movement sensation to experience this space. The same type of multi-parametric best-fit analysis is applied to these sensory systems as is applied to the visual system, with the same results, a highly stable and dependable experience but one liable to similar illusion errors when input is sparse or ambiguous.
The best-fit process does not end with our perceptions such as depth. The different sensory modes, our concepts of reality, our memories and our predictions are integrated. This process can be felt when a small inconsistency, completely and instantaneously changes our perception. An object falling at an angle is perceived and then in a flash, the whole orientation of the world changes so that the next object falls straight down. A very real flower is for an instant unusually

cold to the touch and then becomes a china flower. A drawing that has two equally valid best-fit scenarios is never seen as both or neither but alternately as one then the other. But when a small, almost inconspicuous addition is made, a spot or an extension of a line, the perception immediately stabilizes on the scenario that has benefited. So although uni-modal perception is usually dependable enough and based on a sufficient redundancy of information to produce a best-fit which requires no refitting in light of other constraints, when a global best-fit would give a different perception, then the global analysis is imposed on the offending sensory perception.

Movement

We leave sensory perception at this point and examine another large area of thought, that of movement. This is other side of the sensory-inter-neuron network-motor neuron bridge. In order for a model to be useful in motor control, it must be a predictive model. This also dictates the nature of the model.

A baby, when it is born, is able to walk when its weight is supported. Why does it lose this ability soon after birth and then, later, have to slowly and painfully acquire walking again? It is not obvious what the function of walking is in a new-born human, but its loss is clearly functional. A baby has a rudimentary model of its environment and of itself. Its model is not sufficiently predictive to protect the baby from gross and continual injury if it continued to be able to walk. Movement is only safe and useful from within a dependable predictive model of reality, whether that model is largely in-born or largely learnt.

Our attachment to causality is not the result of our understanding of physics; a proto-causality is at the foundation of our continuous predictive activity. We must have a model to project forward in time in order to predict. We cannot use simple previous experience. No experience is identical in every detail to a previous experience and we have neither the space not time to store and sift through all past experience at the level of detail that would be required. Instead it is our modeling of relationships between objects and events, our understanding of experience that allows prediction. By comparing or predictions of experience against actual experience, we learn to model reality with deeper understanding.

Stereotypical sequences

When we initiate an action such as walking, a complex process begins. This process causes muscles to contract in a particular stereotypic sequence. At the same time, the process projects the nature of signal that will be received as a result of these movements: signals received from the muscles, changes in sensory perception, movement signals from the inner ear, changes in sensory perception of the world. As a result of continual monitoring of prediction against result, the stereotypic sequence is altered to give a smooth, fine-tuned, accurate accommodation to the terrain. If we initiate reaching, a different stereotypic sequence is put into effect with predictive feed-back control. These two discrete programs can be carried on at the same time with, say, a moving object as the target of the reach. Now the programs also accommodate to each other and to a tracking system intent on the target object. Once this more complicated movement has been done many times, it will become highly coordinated and can be called up as a single stereotypic sequence. Such sequences, which are well learnt, habitual and flexible, can be initiated with only the bare essentials of the processes involved being consciously apparent. We remember reaching for our coffee cup but not the details of the action.

When something happens that renders us unable to predict our experiences accurately, our motor programs are un-useable. We feel extreme disorientation, discomfort and tentativeness until we learn to deal with the new situation. Consider someone first encountering life on a boat in rough water. The person clings to some part of the boat for support and is violently ill if he tries to move. But soon the person begins to predict the boat’s movement, to stabilize his visual field, and to use them along with the sensations of acceleration to tailor his movement to the new situation. He has gained his ‘sea legs’. He can again simply walk without committing to short-term memory every minute prediction, mismatch and correction required.

In light of this ability to predict with accuracy and the almost constant use of this ability, it is not surprising that long-term planning and decision making appear to use this modeling facility.

Self

We have discussed the format of the model and its use, but not yet discussed its elements. The actual elements depend to a large extent on past experience and are therefore somewhat unique to individuals. However, there appear to be classes of elements associated with various modeling subsystems and these subsystems appear to be innate. There is no exhaustive list of modeling systems used in thought, but a few stand out in my mind as extremely important and fairly accessible. I will discuss the modeling of self, other beings, places, and categorical symbols.

The model's 'me' is closely associated with motor control. We move, more or less as a unit, not leaving bits behind. So 'me' is in one sense my motor control unit. When we control a tool such as a hammer or needle, our identity extends to include the tool to a surprising extent. When we drive a car, the same thing happens. But there are no passive extensions of identity through movement; we do not extend our identity to include a train or an electric stove. In these cases, the motor control is not intimate enough or not ours at all.

There is also a sense in which 'me' is bounded by my sensations: sensations of pain, muscle tone and tactile feeling give a boundary. Thus we have two boundaries to our identity, one based on motor control and one based on bodily sensations. These usually coincide, and in any animal with no material culture, would always coincide. When the boundaries do not coincide, we may experience something odd, such as flinching and feeling surprised that there is no pain felt when the car we are driving is scraped by another in traffic.

Some would say that the 'me' is associated with memory or motivation. However people can lose large tracks of memory in amnesia without lose in 'me-ness'. They can forgo their motivational structure as in hypnosis and still have a sense of identity. Further, if we imagine that we could install a memory or motivation in the mind of another, this trick does not include an enlargement of our own identity in our imagination. None the less, our memory and motivation do accrue to the modeling element 'me'. When we lose memory or motivation we feel our identity has been weakened.

<u>Others</u>

A category of model element that arises as an extension of identity is other beings, individuals resembling ourselves. We are not forced by the nature of the object to model it as another being. A doctor can model the occupant of a particular bed as a case of typhoid or a clerk can model the object at the wicket as a particular file. By and large, however, we model other humans and most animals as other beings.

This modeling system is largely non-verbal, even non-symbolic. Other people do not need names or labels, even implied ones, to be recognized as individuals. We have a poor vocabulary to express the discrimination we make when recognizing an individual. We analyze faces at a structural level that allows us to recognize individuals who have changed drastically on a superficial level: someone who has gained a mustache, lost his glasses, put on weight and aged ten years. But if we are asked to describe someone, it is these superficial attributes that we are forced to use because we cannot articulate the structure of the face in words.

As well as faces, we also can analyze the voices, mannerisms, gait etc. of other people in great detail. We use a special method to predict the behaviour of the objects modeled in this way. We project onto individuals a motivational structure, habits of thought, emotional tendencies and so on which add up to a personality. In a particular circumstance we might expect a very specific reaction from someone we knew well. But when we see that the person is in a depressed mood, we change our prediction to a different but equally specific expectation. This system is based on the subtle physical changes that accompany emotional states: changes in pupil size, eye movements, body language, gestures, tone of voice, muscle tremor, teariness and so on, including odour. (Some people would deny that odour has this function in humans even though they would agree on its importance to other animals. This seems odd in light of the well developed sense of smell in humans, the production of emotion related scents in humans, the ability of a smell to evoke memories of places, people and moods with instant clarity and our ability to discern faked emotional states even when fairly well acted. We cannot examine any questions of smell by introspection because the olfactory tract enters the brain be the back-door, so to speak, and does not participate in memory formation in the same fashion as the other senses.)

We have this well developed system for recognizing other individuals, ascertaining their mood and motivating, and predicting their behavior. Although the system is not fool-proof, the power of this analysis is truly amazing. We are not restricted to only using it on other human beings. We can and do apply it to other animals when convenient and other animals apply it to us. Card players can treat the 52 cards as individuals with face/personality in order to easily remember which cards have been played and when. Primitive farmers can use the ‘face’ and ‘mood’ of the sky to predict the weather. The system can be applied to anything. It may not always be used appropriately, but it is so powerful a system of prediction that a very deep understanding of some other type is required before the face/personality system is overtaken in flexibility and effectiveness. This is probably the most profound and complex non-verbal modeling system we possess.

Place

There is another fairly elaborate system of modeling found in ourselves and other animals. This is the recognition of ‘place’ and ‘path’ to produce, in effect, a predictive map. Again, this is basically a non-verbal form of thinking.

A disruption of our map produces the peculiar feeling of being lost. It is not the three-dimensional framework of our model that has disappeared; it is contact with a map imposed on that space that is gone. We cannot trace on our map how we got to this place and we cannot predict where we will arrive if we follow any particular path. This lost feeling is uncomfortable and we expend a good deal of mental energy in maintaining our map.

Place is a complex concept. It has enormous associations with a kind of memory summation: habits of action, habits of thought, and habits of mood. Our behavior is different depending on where we are: home, a friend’s home, a stranger’s house, the street, the bush. In a sense places belong to us or others or all of us or no one in a way that objects do not.

The network of paths that connect places is in essence a two-dimensional structure unless three-dimensions are actually required. Places are experienced in three-dimensions, but our map is usually two-dimensional. This facility to draw predictive lines to connect entities and form a map can be extended to analyze other entitles than place and other predictive modes than directional movement. We can map any complex array of concepts with connections we happen to have between them. We can then think with this map without becoming ‘lost’. Although we tend to use directional, causal, logical or memory association paths, the connections can be extremely vague and still increase our useful understanding of an array. We can map music, for instance, using connections of harmony and rhythm.

Language

There are other systems of non-verbal modeling besides the two discussed here, such as the system used to model bodily movements, and so to think kinetically. In these systems, understanding can accrue to an entity which requires no label, or if dealt with verbally, acquires a proper name. Further, this understanding can be productively useful although it is difficult to communicate that understanding in purely verbal terms.

When we come to verbal modeling there seems to be two systems at work: one concerned with category/property relationships and the other concerned with linear sequencing for communication.

We can form a set of categories and still use the two-dimensional mapping system or the multi-dimensional face/personality system rather than the linear language system to examine and model the set. However, by and large our categories are manipulated as symbols such as words, numbers and signs in natural or formal languages.

If someone has found it useful to have a category and the languages they have learnt give them no symbol for that category, they will simply use the image of an archetype. A proper name or metaphor, suitably ‘coined’, can become the symbol for a whole category. By and large, mature languages can supply symbols for any categorical distinctions that have been found useful by the culture using the language. These shared language symbols are somewhat different from personally derived ones. They often have many rather than one central meaning, as an analysis of the word ‘game’ has shown. They also have wider, vaguer and less natural boundaries of classes.

The question arises, “do we use a personal system of categories to model the world and translate them for communication of do we use the categories supplied by our languages to model with?” Without doubt, we learn much about our world and how to model it when we learn our language and when that language is used to instruct us. Certain difficulties in communication point to at least some use of personal categories in our thought. We search for appropriate words to express ideas which we feel are fully formulated; we are forced to express simple ideas in complex phrases; we occasionally learn a new word which exactly meets a need we have felt for some time; we are sometimes forced to use bits of shared jokes or stories to express a complex idea which has become a unitary concept for us. It seems we model using our own personal categorization of experience, which bears a great resemblance to the one we have been taught but is not identical.

A moderate lack of correspondence between the system of categories of thought and of speech is not too surprising because the uses of category are different in modeling and in communication. We think in categories in order to make modeling, memory and prediction easier. We communicate in order to affect someone else's model of reality and therefore his behavior. In communication English speakers find little need for a separate word for a plastic glass as opposed to a glass one. The word 'glass' tends to have a functional meaning. When the need arises, it is simple to specify which type of glass is meant. But the difference between plastic and glass is very important in a model of reality; we must think differently about them. When handling a glass, we must perceive the category immediately because we must predict the approximate weight of anything we are about to lift. If we fail to predict the weight accurately, our hand will shoot down or up in an uncontrolled movement when we take the load.

There are many who believe that the meaning of language is embodied in its grammar and syntax, and that therefore it is necessary to use these forms when thinking. It is not so. There are two injuries to the brain which show the separation of grammatical form from meaning. The one injury produces a person who responds to questions with meaningless nonsense housed in good grammar and syntax. The other injury produces a person who responds with good sense and appropriate words but with non-existent grammatical form and rudimentary syntax. The purpose of grammar appears to be to render a meaning relationship between symbols into a standard linear form suitable for communication. The relationships and thoughts we which to express may very neatly correspond to grammatical sentences or they may be extremely awkward to package.

<u>Discreteness</u>

In the systems of thought we have examined, both verbal and non-verbal, there is a common structure, which I suspect is true for all thought. All the systems have discrete elements; they treat the reality they are modeling as if it was actually composed of discrete, separate and bounded entities. This is not necessarily 'true' – the world may not be composed of objects or things. However, we are forced by the nature of our thought to experience a world of discrete things. This commonality of structure is very important for it allows the integration of thought. Elements and relationships can be slipped easily from one modeling system to another. Metaphor and analogy is possible between modeling systems. Our model of reality can in the end be a singular one.

The problems with discreteness arise when we attempt to model reality at extremes of scale and find few natural boundaries and an arbitrariness in our fragmentation of reality. At these same extremes of scale, we encounter problems with three-dimensional space, a unidirectional continuum of time, causality and even, some would say, logical relationships. Our brains were never prepared by quantum mechanics and so we find it and similar theories difficult to comprehend except in with formal symbolic ways such as mathematics.

<u>Focus</u>

There is another aspect of modeling which we have not examined. We know that we model similar situations differently on different occasions. To a large extent this is a product of our focus and our mood. Changes in attention, motivation and emotional state are continuously affecting the way we think. But they do not supplant or bypass our model. Rather, they affect our behavior through our model.

At a very deep level there are indicators of threat and injury which impinge on thought in very powerful ways; pain, lack of air, thirst and so on are not to be ignored. We will and do, however, suppress driving motivations if, from within our model, that is the reasonable thing to do. We do not drink a poison because of thirst. It is our modeling that makes a particular liquid poisonous. As thirst becomes more life-threatening, more of our thought is directed to finding relief; we become single-minded in our focus on finding water, it becomes our only motivation. But we still use our model as a method to try desperately to predict and plan how to find water.

Suppose someone has several powerful driving motivations at the same time. This person is in pain from an injury, is very cold is thirsty and hungry. He will not simply follow the dictates of his most powerful drive; he will try to understand his situation and act on that understanding. He may keep moving for warmth although he feels pain more keenly than cold. He may expose his skin to bind his wound although he feels the cold more keenly than his pain. He may eat snow although it increases his chill more than it reduces his thirst. He may ignore his hunger, although it may be most powerful, because he sees it as the least life-threatening. His thoughts are not concerned with the obvious or immediate satisfaction of his drives but with planning the best course of action to save his life. His life will depend on his ability to model his situation and act on the basis of that model.

There are a number of reflexes which operate outside our model: blinking when an object approaches the eye, withdrawing from a very hot object and so on. These true reflexes are faster than thought. However, if within our model, it is important to avoid giving a reflex response, we can suppress the reflex. We can only do this when we anticipate the occurrence of the reflex in our model. Although we naturally withdraw from flame, we can 'steel our nerve' and run through a flame to safety.

The situation is somewhat more complex than an examination of the classic drives and reflexes would indicate. We possess a constellation of 'hard wired' drives and reflexes but we also acquire similar motivational substrates through habit. For instance, it is difficult to distinguish on the basis of speed or the need to anticipate the response in order to suppress it, between our withdrawing from heat and our orientation and attention response to hearing our name. Not is it easy to distinguish between a drive like thirst and a powerful motivation to succeed at some arbitrary game. This is no need to draw a firm line between inborn and acquired motivation. Some motivations we are born with and lose, some we are born with and never lose, some we acquire because we were born with the necessity to acquire them, some we acquire without any inborn tendency, and some we acquire in spite of inborn tendencies to the contrary. Except for the extremes of the spectrum it is not possible to categorize motivation into genetically structural components and products of thoughtful learning. For the bulk of our motivational substratum, the 'nature-nurture' question is sterile, vacuous and counterproductive.

The focus of our thought is not only shifted by motivations arising outside our thought but by the substance of our thoughts as well. The model itself may dictate its preferred focus. If something is difficult to model we devote more time and effort to it. If the model indicates that something is important, it becomes one of the foci of our thought. Whether the source of motivation is a drive, pseudo-drive or modeling imperative, the result is the same. Our mind concentrates on a particular aspect of our experience. Our perception, prediction and motor control are never without the influence of attention. At the extreme of concentration on one area, we may not see or hear anything and may be very sloppy in or actions. We can, of course, find ourselves in a relaxed alert mode, where our thought can shift easily and quickly. It seems like a focus-less state. But when we examine this state afterwards, it usually turns out to be achieved by attention on the act of perception itself or an idling period in which nothing is actually being thought, as in mediation. Our thoughts, our model, almost always have a focus.

<u>Emotion</u>

Another major influence on the model's operation is a chemical switching system we call emotion. This can be envisioned as possessing many different minds which operate slightly differently but share the same memory. Our perception changes as we move from one emotional state to another, so does what we plan, and our range of behaviors.

An example of an emotional state is fear. It may vary in intensity from mild apprehension to terror. It can arise from almost any source: a loss of balance, an object approaching quickly, a sharp pain, a phobia of spiders, a threatening gesture, a situation we know from experience to be dangerous, a situation we have never encountered before but model to be dangerous. Fear arises when we feel threatened but we may fool this way because we have been genetically programmed to feel so, or because we have learnt to, or because we understand it is the appropriate response in a situation. The result, independent of cause, is a surge of chemical mediators to prepare the body, including the brain, for escape from danger. Fear can be mimicked by an injection of nor-adrenalin. This is an artificial state for two reasons: nor-adrenaline is only the major chemical of a number released during fear, and its distribution in the body is slightly different compared to natural release. What results is a peculiar state of mind – a distinct feeling of apprehension but with no source to focus on. This is not what we normally mean by 'fear'. Fear includes both the change of thought and bodily function, the feeling, and the change in attention focus, the motivation.

Mood

There is a group of classic emotions which can be mimicked chemically such as fear with nor-adrenalin, rage by adrenalin, ecstasy by morphine, and so on. But mood is a more subtle notion, a mixture of very low level emotional states and many other inputs, which tailors our modeling activity to our situation. Focus and mood keep our modeling relevant. Instead of a group of primary colours, mood is like an available rainbow of pastels. The metaphor of colour for emotion and mood is a very apt one. A particular mood change does not seem to change the structure, line or shape of the model but its hue. Or to take a musical metaphor, it is not the tones or rhythm that changes but the timbre.

Mood is not simply mixed and low intensity emotional states. It is also dependent on hormonal levels and the chemicals of the body's metabolism such as sugar. As well as its chemical determinants, mood probably depends on gross electrical fields within the brain. Such fields exist and are measured with EEGs. The workings of the brain are almost entirely dependent on the voltages across neuron membranes. A neuron cannot distinguish changes in potential caused by chemical and electrical causes (or electrical effects caused by magnetic fields for that matter). Even external fields, if sufficiently strong, have mood effects. There is reason to believe that smells have a strong and direct mood altering effects. How we understand reality through our model of it would also affect mood.

The function of this complex and multi-layered system of motivational attention and emotion/mood can be seen by imagining its absence. Thought would then be without any means of maintaining appropriateness, perspective, a sense of priority, speed or efficiency. It would lose much of its ability to use genetic wisdom. It would lack flexibility. Its ability to learn would be taxed by a lack of discrimination. A man with no fear is indeed a fool, and so is a man without anger or joy.

Memory

We have examined many aspects of the model's formation and use, but not, so far, its storage. We remember a sort of motion picture of our model. However, it is a very simplified record; there is no usefulness in storing every detail. Only the general picture and the significant detail are required. But, what are the criteria of significance or usefulness? To a certain extent it is obvious whether some particular item should be stored. If it is part of a complex on-going process then we may need it again. If it is very surprising we may find it significant. If it is the sort of thing we have found useful in the past, there is a high probability that we may need it.

A good safety mechanism would be to store as much as possible and then clear out the storage of useless items later. There are at least four types of memory: immediate (working) memory, short-term, intermediate-term and long-term. As time passes and a particular parcel of memory is transferred to a longer-term memory, its detail is lost as it becomes more obsolete. Older memories are also reworked by more recent ones. It is not just that memories lose detail and become amalgamated; they also change. Being confronted by an object of memory or a photograph of an event we remember, can show the extent of this change.

Consciousness

But the immediate and recent short-term memory is still a quite accurate version of the model as it was, with only the detail being progressively lost. The immediate memory is almost identical to conscious experience; an edited form of the model. When we introspect, when we try to include in our thought our on-going processes of thought, when we model our modeling, then many aspects of the model that would ordinarily not merit memory are significant. This gives a false impression of the wealth of detail in our consciousness. That wealth and more is always in the model but usually not in our consciousness. To think that consciousness always contains a detailed version of the whole model, is similar to thinking that our vision always has a clear image of the whole visual field because any small area we look at directly is clear and highly resolved.

A way to gain a rough impression of the amount and type of detail normally made conscious is to leave the room without preparing for it ahead of time and attempt to reconstruct a detailed image. We remember who is in the room and where they are sitting, their general posture and facial expressions. But the details of position, posture and expression are not available. Most of the non-animate objects are sketchier or absent. Even those amazing people who can draw from a 'photographic' memory miss more obvious detail then they retain. The arrangement of cups on the table or shape of leaves on a plant may have been rendered by these people – certainly not by others – but still remain subtly inaccurate and sketchy. We remember the gist of conversation, its structure, many key phrases, but cannot give the exact details of the last exchange including exact stress, rhythm and intonation. We have remembered what is useful, new, interesting and the rest must be reconstructed.

It is the confusion of consciousness with thought that make though appear so puzzling. When we view the process in a different way, the puzzle disappears. We think; we have a memory of our thought which is incomplete. Consciousness is that memory during its formation. As we can solve conceptual problems in our sleep, it appears that memory formation is not required during thought.

In summary

We can view ourselves as having a nervous system with many functions, one of the major ones being the function of 'mind'. This mind function (singular not plural) consists of building, maintaining and refining a model of reality; using the model to predict, plan, decide, initiate and control responses to the world; storing an edited form of the model as a memory of experience for comparison and learning. During introspection we only have access to consciousness, the memory being formed, and not the model itself.

PART TWO – Brain as modeler

My viewpoint, to be convincing, must be consistent with what we understand of the mind, but it must also be consistent with what we know of the brain. Neural anatomy and physiology are in their infancy, however it is still possible to use what scientific knowledge does exist to examine the process of thought.

Neo-cortex

In the same way that the conventional account of consciousness was a difficulty in my examination of mind, the conventional account of the pre-eminence of the new-cortex is a difficulty for me now. The logic of giving the neo-cortex a position as the only or the main organ of thought seems to be based on a philosophical rather then scientific foundation. The argument goes: humans think but animals do not, the greatest anatomical difference between ourselves and other animals is the size of the neo-cortex, therefore we think with the neo-cortex. However, there is no reason to believe that our thought differs in kind as opposed to degree from that of other intelligent social animals. Nor is the neo-cortex the only area of the brain where a difference in proportion is found between ourselves and other animals. Finally, although it is a fairly good guess that differences in thought will correlate with differences in anatomy, the correlation is as likely to tend towards the subtle as the simplistic.

Neurons

Let us start by examining the types of information processing that neurons can perform. A neuron can either transmit a signal or not, there is no in between. This has led some to assume that the brain must function in the same way as a binary digital computer. However, a neuron can transmit more than one type of signal, so it cannot be described in terms of on-off, ones and zeros or any other binary system. A neuron can be thought of as transmitting more than one bit of information at a time. Firstly, a neuron transmits either a positive or a negative signal because it is either excitory or inhibitory in its chemical type. Secondly, a neuron carries a graded intensity of signal encoded in frequency of firing. The frequency code for intensity is lost at the actual synaptic junctions between neurons and the signal becomes a continuous graded function of intensity (completely analog rather than at all digital). Third, a neuron's signal carries spatial information about the origin of the signal, modal information about the originating cell type and temporal information about its time of onset. This is because, unlike the sequential regime of computers, neurons make-up a parallel system of signal transmission. So when a neuron transmission reaches another neuron it carries the information that a negative or a positive message of such and such intensity has arrived from a particular type of cell in a particular place and it started at a particular time. There is a big difference between a brain and a silicon type digital computer.

As well as being a transmitter of information, a neuron is a processor of information. A neuron fires if the voltage across its membrane falls below a threshold level. The neuron can integrate all the influences on its threshold level and its membrane potential into its firing rate. This allows a neuron to function as what in computer terms is called a logic gate.

A neuron ‘and’ gate is simple to imagine. Suppose a neuron A and a neuron B both have excitory synapses on our and-gate neuron. But neither A or B alone have enough synapses on the and-gate neuron to overcome its threshold. Neither alone can fire the and-gate but together they overcome the gate cell’s threshold. Thus the and-gate fires if both A and B fire. If either A or B alone can overcome the threshold, we have an or-gate. The or-gate fires if either A or B fires. If we suppose that A and B are inhibitory rather than excitory, we produce ‘nor’ and ‘nand’ gates. Simple 3 neuron gates are only the start. Hundreds of neurons could contact a ‘gate’, some inhibitory and some excitory with varying connection strengths to realize extremely complex logical expressions.

As well as the logic gate functions, neurons have other processing capabilities. Suppose that we have two neurons, C and D. Each makes many inhibitory synapses on the other. Then if C is firing, it is extremely difficult for D to fire, and vice versa. Now we have a discriminator; the most intense signal (or the first signal, depending on the details of the arrangement) will be transmitted by the discriminator and the weaker or later signal will disappear. A discriminator can be used to produce definition and for comparison. Neurons can also act as delays. If a signal is transmitted through an extra synapse, it will acquire a delay. Thus a signal can be distributed sequentially with fixed delays to a number of other neurons.

It is doubtful that these are the only functions that a neuron can serve. Further, it is doubtful that any particular neuron would have only one function. A typical large neuron may have a very complex discriminator embodied in its dendrites, c complex logic-gate function over its cell body, a sequential delay line distribution system along its axon and some powerful over-riding logic gates immediately up-stream of its terminal synapses. Such a cell would also be sensitive to global and local brain activity through its sensitivity to the electrical and chemical environment. Finally, the neuron would be able to grow and change in response to its activity. The brain contains an astronomical number of these little processors.

<u>Structures</u>

It would be very difficult and inefficient to try to determine the functional roles of individual neurons in the brain. If we were faced with the million million neurons in a random arrangement in the brain, we would be at a loss to ascribe function to each one. But we are not faced with such a situation. The million million neurons have discernable form. By and large they are laid out neatly in sheets, in modules, in clear groups. The axons bringing information into an area stream through the area with a particular pattern, the dendrites with glean from these axons are arranged in particular shapes and sizes of tufts and arbors. The contact between neurons within the area follows standard local patterns. The axons carrying information out of the area form defined tracts. The structure of sheets of regular parallel processing units almost states their function, at least in a general way.

<u>Neo-cortex</u>

We can see the neo-cortex as a large but fine-scale discriminator. The cortex is a thin sheet of neurons with a standard structure. Over its whole area it has six layers, each with its own distinctive cell types and connections. It can be divided into several million modules, each a small column reaching through the six layers from one surface of the sheet to the other. Adjacent modules powerfully inhibit each other and sample only slightly different portions of the local input signals. For instance, a representation of the retina is laid out on part of the cortex with its spatial relationships conserved. A particular module has a particular pattern, colour, orientation or movement in one particular area of the retina to identify. It will make a particular sampling of the input from that area of the retina to identify its target. The module will use internal networks to arrive at a final state of excitation corresponding to the fit of its input to its target. As the neighbouring modules are targeted on similar patterns and overlapping areas of retina, the mutual inhibition between modules allows only the one with the highest state of excitation (the closest fit) to output signals from the cortex neighbourhood. The cortex can thus be seen as a facility for making a fine-grained, high definition analysis of input signals.

An odd feature of the neo-cortex is its weird specialization. Specific areas of the sheet do very specific discriminations and when they are damaged, the resulting lack of skill is particular. For instance, the prosopagnosia the injured person can recognize people from the total person, from the voice but not from the face. They can however recognize facial expressions. There is an area whose damage produces visual-object agnosia. This area deals only with objects that can be manipulated with the hands and not other objects (even if they are small and in the foreground). Long lists of such gnosia and aphasia conditions are associated with damage to particular areas of the neo-cortex.

Until recently, there has been a particular pattern to studies of neural anatomy. An input from some sensory organ is traced to its first destination in the brain, from there to its second and so on, until it reaches the neo-cortex. At the neo-cortex this process stops because the assumption is made that the neo-cortex is not just another processing way-station along the pathway but the organ of thought, the end point of the pathway. It is assumed that information marches across the cortex sheet from the more primary sensory areas to the associative and then the motor areas. The problems with this picture have been ignored on the justification that though is a very complex thing and we cannot expect to understand it quite yet.

Although the areas of the neo-cortex are interconnected, they are not massively interconnected. The large bundle of fibers that connects the two hemispheres can be cut with very little disturbance of function. Only under elaborate experimental conditions is this 'split brain' state obvious. Some of the other fiber bundles that connect parts of the same hemisphere do appear to be necessary for normal function, but they do not seem to have the importance of connections between the neo-cortex and the rest of the brain. Leucotomy, for instance, is not the severing of the frontal lobes from the rest of the cortex as most people assume but the severing of the frontal lobes from as area of the mid-brain called the thalamus. The large changes in personality and behaviour that follow leucotomy are due to a lack of two way traffic with the rest of the brain rather than the rest of the neo-cortex.

<u>Thalamus</u>

The frontal lobe is not alone in having an important two-way traffic with the thalamus. The thalamus is the source of all the sensory input to the neo-cortex except those of smell and internal body chemistry. And further, the bulk of the output from the cortex goes right back to the areas of the thalamus which supplied the input in the first place. Even the associative areas of the neo-cortex receive the bulk of their input not from the primary areas of the cortex but from associative areas of the thalamus. The thalamus and the neo-cortex seem locked in a reciprocal relationship. The axons of the thalamic neurons stream through the neo-cortex. The axons of the neo-cortex neurons stream through the thalamus. In the thalamus, the sense organs such as the retina, the organ of Corti in the ear, and the skin are all represented on the surface topologically as they are on the neo-cortex. Topologically identical neurons in the two maps are linked by feed-back loops. Areas of the thalamus are interconnected as are areas of the neo-cortex. The two structures mirror one another.

The differences between the thalamus and the neo-cortex help to understand this reciprocal relationship. The anatomy of the sensory areas of the thalamus is different from the neo-cortical modules. The incoming axons pass through an area as overlapping conical arbors of branches. The thalamic cells have thick directional tufts of dendrite fields inter locking the axon branches. The shape of these structures indicates pattern recognition. Their size indicates a local diffusion of incoming information and their density points to a selective security of transmission.

The second difference is that the neo-cortex receives the bulk of its incoming axons from the thalamus or other areas of the neo-cortex while the thalamus receives axons from other sources besides itself and the neo-cortex. Practically all of the optic nerve enters the thalamus and so does the ascending tracts of the other senses. Thus the thalamus is in a position to compare the result of neo-cortical processing with the raw data on which it was based. It is in the thalamus that in-coming sensory modes are first mixed. For example, a region of the thalamus called the pulvinar receives mixed axons from two near-by areas of the thalamus, one of which receives the optic nerve and the other of which receives both the auditory and somatic sensory pathways. The reciprocal projections of the pulvinar to the neo-cortex are to the areas of visual association, auditory association and complex association. It is a fairly easy leap of imagination to see this immense set of feed-back loops, set in parallel but with a great deal of overlap and intermixing as a system which will stabilize only on the 'best fit' scenario of sensory input. As long as we have a sequential mindset, complex feed back loops are difficult to understand. There is no place to begin to trace effects and no place to end; everything affects everything else. And yet, such systems are predictable, they are understood, but only in a holistic sort of way using the language developed in electronics, systems behavior and other fields that build or encounter complex feed-back networks.

When we look at the neo-cortex/thalamus feed-back in detail we find that it is interwoven with a number of other feed-back loops of similar size and complexity which involve the midbrain and forebrain associates of the neo-cortex and thalamus. Nestled inside and under the cerebral hemispheres are a number of small but important areas. These ganglia are heavily interconnected and damage to them or their connections cause disturbance to memory, cognition, emotion and motor control. The overlapping and interleaving circuits of this area are the epicenter of all the lines that can be drawn entering or leaving the brain. The thalamus is one of two doors into this epicentral region. Through it enters most sensory information and the excitation of the reticular formation. The other in-door is the hypothalamus which we will examine later.

Reticular formation

The effect of the reticular formation on thought is to control its intensity and general focus. The formation itself has a very unusual anatomy. The neurons are small with very delicate and long dendrites and axon. It extends form the spinal cord through the middle of the brain stem all the way up to the thalamus. Along its whole length it receives input from the areas of the brain what surround it and the nerves that pass to and from the spinal cord. The impression is that nothing happens in the nervous system that is not an input into the reticular formation. But unlike the rest of the brain there is no order, no topological mapping. There is one feature of the anatomy that can give a clue to the functioning of this structure: the dendrites of each neuron tend to form a large horizontal lacy disc while the axon projects up and down in a vertical orientation. This may be the largest add-gate in the world. Each neuron will only fire when the thin disc of its dendrites receives sufficient signals to overcome its threshold. Its excitation is communicated to other discs but forms only a small part of their inputs. When a small number of the neurons are excited there is very little spread of the excitation. But if a large number of the neurons are excited simultaneously, the excitation spreads like a grass fire. Waves of excitation travel into the thalamus from the reticular formation associated with important and intense occurrences no matter what their nature or source. The message is clear,"Wake up, something is happening, find out what it is!" If the reticular formation is quiet then we are asleep; if it is active, we are awake; if it is only moderately active, we are only moderately attentive. No other brain activity correlates so well with the intensity of our experience.

<u>Hypothalamus</u>

The other in-door to the brain's epicenter is the amygdale and the hypothalamus. It is here that the sense of smell enters. The information carried by this sense is already highly processed by a small area of cortex which is anatomically separate from the neo-cortex. The sense does eventually reach the neo-cortex, but not as a sense like others to be filled into a sensory construct. Rather, smell is integrated with the more or less complete construct in the prefrontal lobe of the neo-cortex and its associated area of the thalamus. These areas: the amygdale, hypothalamus, prefrontal cortex and the medio-dorsal area of the thalamus are part of a circuit which has long been called the limbic system and has long been associated not so much with smell as with the elaboration of emotional states.

Besides smell and some vague input from the reticular formation, the hypothalamus measures directly the state of the body. For instance, there are neurons in the hypothalamus that react to dehydration, low blood sugar, body temperature, levels of various hormones and so on. The hypothalamus directly controls the autonomic nervous system and the pituitary gland. Though these controls, it is the hypothalamus that triggers the basic set of strong emotions such as fear. It is also the source of the classic drives such as thirst. Interestingly, the information available to the hypothalamus in addition to smell, the state of activity of the reticular formation, and the chemistry of the body, is of a very high level. It is by and large fed from above rather than below with the most complete and integrated products of the rest of the brain. Its reaction to the needs of the body is in no way automatic or vegetative, but informed. The difference can be seen when the hypothalamus is slowly destroyed and lower centers assume its functions without reference to the mind. I say slowly because otherwise accommodation is impossible and death results.

<u>Out-doors</u>

The thalamus and hypothalamus are the in-door to the epicenter. Where are the out-doors? The hypothalamus is also one of the out doors. Its exports commands to the smooth muscle and glands of the body. The brain controls the skeletal muscle through two distinct but cooperative systems of motor control.

Neo-cortical modules that lie in the motor cortex contain large neurons called pyramidal cells, and some of these neurons send axons directly to the motor-neurons of the spinal cord. This constitutes a fast and direct bypass of more subtle motor control. But the movements produced by this direct pathway are jerky and uncoordinated compared to normal movement. This is no doubt due to the fact that this pathway acts on the individual motor-neurons and not on the network of neurons in the spinal cord which sequence groups of related motor-neurons. The motor cortex also acts via other brain centers on this spinal network to product coordinated movement.

The motor cortex is no different from the rest of the neo-cortex in its relationship to the thalamus. It is one-half of a large complex of parallel feed-back loops. Its surface is not a map of the retina or the skin but a map of the muscles of the body, a homunculus. The process is similar except that the analysis is not of in-coming information but of out-going signals to the muscles. The process of feed-back between the modules of the neo-cortex and the neurons of the thalamus will stabilize on the pattern of motor control that best fits all constraints on movement. The neo-cortex and thalamus are not alone in this enterprise. A number of other areas are involved. There is a cortex to striatum to thalamus loop, a cortex to cerebellum to thalamus loop, and a cortex to reticular formation to thalamus loop. All of these important feed-back loops add information, pattern and constraint to the basic interaction of the motor cortex and the ventral thalamus. The cerebellum is essential to habits of action: it is the cerebellum that 'never forgets how to ride a bicycle'. It can be thought of as a library of movement sequencing patterns. Through the striatum loop enter signals from the substantia nigra. This small area does not function properly in Parkinson's disease with the result that voluntary movement is difficult to initiate although it can be planned.

Memory

So far we have a partial black box understanding of a region of the brain which takes in sensory information and gives out motor commands. Within this region we can the outline of an immense number of parallel feed-back loops, the neo-cortex-thalamus group and others. This sort of structure over-lapping parallel feed-back loops is the type of structure that can stabilize quickly on a 'best fit' solution. This system is in communication with another set of feed-back loops involving the hypothalamus. This structure integrates the best-fit perception and action with information about the bodies needs and drives. The result is emotion and the effects of emotion are felt by the rest of the brain was well as the body. Let us now look at the centers of memory and consciousness.

Unfortunately, we do not know the nature of the memory trace or even whether it has a single nature. It may be electrical, chemical, anatomical growth, some other mechanism or probably a combination. It is known that certain areas and circuits are involved. When an area called the hippocampus is removed the result is a person who can recall memories he had before the operation was performed but cannot form any new memories. The condition of Korsakoff syndrome, which occurs in chronic alcoholism and thiamine deficiency, destroys neurons in some areas of the brain especially the mammillary bodies. The result is difficulty in memory storage, recall, accuracy and indexing. Between the hippocampus and the mammillary bodies runs a nerve tract called the fornix which contains more fibers than either the optic nerve or the cortical-spinal tract. It is in this area that we will look for memory and examination shows that the hippocampus is the point of overlap between two large feed-back loops, and that these loops carry the most highly processed information that can be found in the brain.

Informational Circuit

One of these loops goes hippocampus-mammillary bodies-anterior thalamus-cingulate cortex-entorhinal cortex-hippocampus. Into this circuit is fed information for all of the neo-cortex both directly and via the frontal lobes. The neo-cortex input enters the cingulated and entorhinal cortexes where complex cognitive maps are formed from the many overlapping but more specific maps of the neo-cortex. Inside this loop is a smaller one between the hippocampus and the entorhinal cortex. Each of the neurons of the entorhinal cortex is in direct two-way contact with the calls in a thin slice of hippocampal tissue. We have our old friend here, a complex parallel set of feed-back loops which will stabilize on the best-fit scenario. And because this set of loops preserves the topology of this cognitive map, we can assume that the larger loop through the through the mammillary bodies and thalamus do likewise. There is an interesting feature of the larger loop. A particular type of neuron in the hippocampus reacts to a signal after 2 or 3 presentations. Its target cell in the mammillary bodies reacts after 3 or 4 and the next cell down the line in the thalamus reacts after 4 to 7 times. Finally the corresponding cell of the cingulated cortex reacts after 15 presentations. The entorhinal cortex is thus input with signals and their echo after about 15 iterations. The small loop deals with all the information in the entorhinal cortex while the larger one only completes its circle for more long lived information. We can call this complex of feed-back loops the 'informational circuit'.

Regulatory Circuit

The other loop which passes through the hippocampus is of a very different kind. It goes hippocampus-lateral septal cortex-hypothalamus-reticular formation-medial septal cortex-hippocampus. The passage of signals along the reticular formation for a short distance makes the nature of these signals somewhat vague and amorphous, but it does give access to the attention producing part of the brain. This pathway has two other consequences: it imposes on the loops a 3 to 6 Hertz rhythm which synchronizes them and it gives a reference point to the timing of events. Through the hypothalamus part of this circuit is added the emotion and drive produced by the hypothalamus and amygdale. The circuit appears to have a regulatory function and to impose criteria of importance and relevance on the information in the other circuit. We can call this complex of feed-back loops the 'regulation circuit'.

Another circuit goes lateral septal cortex-medial dorsal thalamus-prefrontal cortex-entorhinal cortex. This third circuit connects the informational and the regulator circuits as does the passage of both circuits through the hippocampus. These circuits and others form the central feature of what I have called the epicenter of the brain.

Hippocampus

In the interplay between the entorhinal cortex, the septal region and the hippocampus lies the model of reality which is central to 'mind'. It is here that past, present and future become a continuum. It is here that processed information becomes consciousness and then memory. All the minute discriminations that have been made by the neo-cortex to sensory, spatial, temporal, motor, semantic and even emotional data is now not only given its final assembly but also filtered according to the significance of the data to both the model itself and to the biological needs at that time. In the hippocampus we find cells that anticipate events.

The anatomy of the hippocampus is complex and I will not describe it here. However, it is important to note that the groups of neurons within each slice have bifurcated axons which allow each to output signals and also to feed those same signals into the dendrite fields of other neurons within the slice. Each slice then becomes a rind circuit which can reverberate for a long time after it is stimulated. The turning-on and turning-off of reverberations as opposed to the transmission in and out of the slices, seems to depend on whether signals arrive in synchrony from the informational and regulatory circuits. This is very reminiscent of random access memory units and is probably the site of the immediate memory. It is not however the site of short or long-term memory.

This account of the brain has omitted a great deal more than it has included. Whole aspects of neurology are missing; the chemical nature of neural transmitters in different areas of the brain is just one example of important omissions. Much detail is hazy and may be overtaken by new scientific discoveries. The purpose was not to examine the brain in detail but simply to indicate that the brain is capable of functioning as described in Part One.

In summary:

Our brains are capable performing the function of 'mind'. The brain can maintain a model of reality and use it to interact with the world. It can also be conscious through the immediate memory of the model. The key to understanding how this is done lays in the ability of massive numbers of parallel, overlapping feed-back loops to quickly stabilize on a stable state, a 'best-fit' solution.

PART THREE – Modeling in Philosophy

As well as being consistent with a convincing image of both the mind and the brain, the concept of thought as a modeling process is philosophically useful. The concept has a bearing on several philosophical questions. In this part I will discuss three of these. What is the relationship between mind and matter? What is the nature of our knowledge of reality? To what extent is our will constrained? I hope to show that a different approach can shed light on these questions. From my own perspective, these problems have been solved by the use of the concept that thought is modeling. To illustrate what I mean by using a concept to solve a philosophical problem, Zeno's paradox of motion is a good example.

Solutions

Zeno's paradox goes as follows. An arrow flies from A to B; we see it fly and see it land. But we can analyze the flight and say that in order for the arrow to reach B, it must reach a point half way between A and B, call this point C. In order to get from C to B, the arrow must reach a point half way between C and B, call this point D. We can go on with this process of halving the distance left through the rest of the alphabet and on without end. As the process cannot be ended the arrow cannot reach B. Here is the paradox: the arrow does reach B and cannot reach B. The concept that has solved this paradox to most people's satisfaction was the concept of the sum of an infinite series. The series that consists of one half plus one quarter plus one eight etc. has the sum of 1. Thus we can see the flight of the arrow as two infinite series, one of distance and the other of time. In half the time the arrow travels half the distance and so on. But when the time reaches unity so does the distance. When the time of flight has elapsed, the arrow reaches B. It does not matter how fine we slice up the distance, if we also slice up the time in the same way. Zeno's paradox was once an important philosophical problem is now a minor and esoteric intellectual diversion. It is not that the problem has really disappeared, but that it has been transformed into a mathematical concern with the exact relationship between the limit of an infinite series and the sum of an infinite series. To the extent that the concept of the sum of an infinite series is convincing to a particular person, to that extent the paradox is solved for that particular person. Most people are so comfortable with the sum that they now find it difficult to even grasp that a paradox exists. What a new conceptual framework does is to transform the problem and make it trivial.

Mind-Matter Problem

Let us now turn to the mind-matter problem. We all have an existence with is of the mind and also of the physical world. We find it practically impossible to deny that we have physical bodies. An attempt to maintain that we do not runs into brick walls, figuratively and literally. We are of and in the physical world. It is equally impossible to deny that we are of mind. If we maintain that we are not mind, then the question immediately arises, 'what is doing the maintaining?' although when we have nothing important to do, we can sit and pretend that we are only body or only mind, these ideas are both extremely unconvincing. We are usually willing to accept that we are both of mind and of matter; this is, in fact, our normal, functioning, unselfconscious view. The problem is not with either half of our nature but with the relationship between them.

There are three types of answer to the question of the relationship between mind and matter. We can say that they have no relationship, or that they have a relationship of such and such a sort, or that the question has no meaning.

If we assume no relationship, then how do we explain their correspondence? Almost every time a particular frequency of light is reflected into my eyes, my mind contains a red surface. There is a certain amount of predictability about the state of mind that accompanies states of brain and vice versa. How is this possible if mind and matter do not interact? Why do they keep a more or less parallel course? A classic answer is that God wills it so. This simply transfers the question from one of interaction between my mind and my body to one of God's mind and the physical universe. The problem is still there, just on a cosmic scale. If we are not going to resort to well intentioned sleight of hand, or denial of the existence of mind or of matter, we are reduced to coincidence to explain the correspondence. The odds against such a coincidence must be truly astronomical and render the whole proposition unconvincing.

If we assume that matter and mind do interact, then how? No reasonable mechanism has ever been proposed for this interaction. Figuratively, there is no lever and no fulcrum with which mind can move matter or matter can move mind. We could assume that there is a special type of hybrid mind/matter that can interact with both ordinary mind and ordinary matter. This special type of matter is the essence of vitalism. For several centuries vitalism has been in retreat. First it was assumed that life was made of different elements that non-life. When that became untenable, it was assumed that the elements were combined differently in life (organic chemistry was different in kind from inorganic). When that had to be abandoned, it was assumed that the processes in life were different (biochemistry was different in kind from organic chemistry). When that became untenable, vitalism took its last stand in the neo-cortex. Considering the long and sustained effort that has gone into proving the vitalist case, it does not inspire confidence that not one atom of special mind-tainted matter has ever been brought to light.

There is a very real sense in which the question can be made meaningless. This is by an exact and complete identity between mind and matter. For instance, if we ask what the relationship is between Venus and the evening star, the answer is that it is a silly question because they are the same thing. We can reasonably ask what the relationship is between the two words but not between the object and itself. So the question becomes, can we draw a convincing identity between mind and matter? On the surface this appears to be as difficult as any other approach. The identity must be complete; either all of the mind or all of the matter must be included in the identity. It seems natural to approach this condition by assuming that it is all of the mind that has identity with the brain or part of it. But we must include all of the mind. The other condition is that the identity must be exact. There must be no sense other than a semantic one in which the two sides of the identity differ. Here is the rub. If we accept the conventional view of mind and the conventional view of brain, we appear doomed to failure. They will not match.

<u>Identity of Mind-Brain</u>

This identity has been proposed in ways that are not convincing to me and a great many others. The attempt to draw analogy between states of mind and brain states runs into the problem that neither the mind nor brain is sufficiently static or repetitive to have ‘states of’ in any meaningful sense. Each instant is somewhat unique for both, never to be repeated in that exact conformation. This is not to say that there are not predictable correspondences between the general activities in the brain and mind. Far from it, there is a very good correlation by and large. ‘States’ is just the wrong approach to dynamic processes.

Further, the attempts are usually awkward because identity is sought between multiple minds (conscious and unconscious) and only the neo-cortex rather than the whole brain. The idea is that somewhere on the neo-cortex there must be a region which maps to the mind. We simply find this region and the problem is solved. However, the more we learn about the mind and about the neo-cortex, the more difficult it is to believe that this approach will be successful.

Another approach is to say that our ignorance of the brain and mind is such that we are not in a position at the present time to outline an identity between the two. But as we are materialist, we can simply assume that one exists and will someday by found. The problem with this is that, for the bulk of humanity, the one thing that stands between them and a materialist philosophy is their inability to see any way out of the dualism of mind and body. Although it is true that our ignorance for outweighs our understanding of the brain, it does not seem impossible to at least formulate an outline of a possible identity that is comfortable and convincing.

I have used the approach of comparing the functional activity of the brain with the functional activity of the mind in the widest sense. As soon as this functional activity is envisaged as a model building and using system, all that we have been calling mind can be seen as an integrated whole. It is a complex structure; but what we know of mind can be fitted into the context of building, manipulating, editing and storing a model of reality. This same context can be successfully used to understand our information about the brain. We have bypassed the structural dissimilarity between consciousness and the neo-cortex while retaining their detailing functional correspondence.

To state explicitly the identity I am proposing is between the whole of the mind and the activity of part of the brain. Both halves are a functional activity (like a verb not like a noun), and that activity is modeling reality. Another word for this functional activity of modeling reality is thought. There are no differences between the activity of mind and that part of the brain; they are one and the same thing. I am not drawing an identity between mind stuff and matter. The one side of the identity is not matter but the processes of a material system. The other side of the identity is not mind stuff but again the processes of a material system. Matter we have, and its functional activity as a system. Mind stuff, as the thing of which mind is constructed is not a necessary idea. There is no ghost in the machine but there is the mind activity or thought.

What is knowledge?

Let us now turn to the basic epistemological question. What is knowledge; of what can we be certain? There have been three basic answers to this question with fairly wide acceptance. We can be certain of the content of our consciousness; we can be certain of logical structures; we cannot be certain of anything.

If we are in pain, we can be definitely certain that indeed we are in pain, unless, of course, we are not actually in pain. Here we have the problem with the first answer. This is not the kind of knowledge we are usually interested in. When we think we see a book, we are not concerned with the certainly that we assign to whether we do really think we see a book. The concern is whether when we think we see a book, there is really a book there to see. To believe that we have a privileged knowledge of our consciousness is not an idea with which many people would disagree. But it is a singularly useless one. It can be enlarged with some increase in usefulness and becomes the idea that we have privileged knowledge of our minds. However, this idea is not se convincing. Why would the notion of an unconscious mind have been invented and gain wide acceptance if we could indeed by certain of the content of our minds? Why would we have the phrase 'self deception'?

We can be certain of logically derived conclusions from necessary premises. If, of course, that is how we have defined certainly. Even here we have two choices: to accept the conclusion or reject a premise. Again there is a problem of usefulness for arguments which can make at least a good show at being logically impeccable tend to be esoteric and formal. By the time they are translated into our ordinary lives, they have lost the purity of their logic. Further, the ideas of logic and of necessity are products of our thought processes. Their only claim to special status is their a priori nature. And as we have difficulty justifying a priori notions on any basis other than our inability to escape their spell, we have a problem here that we bypass only by defining certainty in terms of logical necessity. Even this situation is made somewhat weak by Godel's theorem.

It is very convincing to believe that there is no such thing as certain knowledge in any useful, non-tautological sense. So what. We still trust our senses and our intellect to an enormous extent. Once we have rid ourselves of any hope of certain knowledge, we can take seriously the question of why we have this trust and how valid it is.

The concept that thought is a modeling process opens a door here. If an economist makes a model of the economy, the last thing he hopes to gain from playing with his model is knowledge. The thing he hopes to gain is understanding. Conversely, if I say that I know a person, this is very different from saying I understand that person. To understand is to appreciate structural relationships, to be able to predict. The only way I can understand something is to model it. The only way I can tell if I understand something is to be able to model it. The object of our thought is not knowledge but understanding. Understanding is useful to us.

Why do we trust our understanding? Actually, we don't trust it on many occasions. Anytime that we find our modeling is not giving us consistent, dependable, predictive understanding, our trust in our understanding of the situation vanishes. We are in interactive contact with reality. As long as our modeling works well, we relax and trust our understanding. When it does not, we don't; we get concerned, suspicious and put a great deal more effort into trying to improve our understanding. Our understanding is continually being tested and that is why we usually can and do trust it.

We are aware of optical illusions and find them entertaining and intriguing. They are not, however, particularly disturbing. Our modeling facility can be fooled by such constructs. But we realize that they are carefully arranged and arbitrary constructs. Normally we have a redundancy of information with which to model. But in the case of optical illusions, one type of modeling information is taken out of context and in isolation. It is precisely those modeling tools that make the accuracy of our vision automatic that make us prone to optical illusion. Instead of a logical system which will not commit itself to uncertain conclusions, we have a modeling system which commits itself to the best conclusion it can find and has a long evolutionary honing towards biologically relevant accuracy.

The great communal scholarships such as history or science do not give us knowledge. They produce very extensive, complicated, predictive and very convincing models which we can

incorporate into our understanding. We accept them and use them as long as they are acceptable and useful. We need not concern ourselves with the absolute truth of these models. We have the answer to that question in practical terms. We do concern ourselves with their consistency, relevance, and predictive accuracy.
To be explicit about this epistemological viewpoint, thought is a modeling process and therefore incapable of formulating or testing for the existence of certain knowledge. Thought is capable of understanding. Because thought modeling is interactive with reality, it can be tested. But the criteria of the testing are not 'truth' but relevance, consistency, extent, and predictive accuracy.
<u>Will</u>
I will now turn to a problem which underlies our moral concerns. To what extent are our wills constrained? There are two historically entrenched positions with differing approaches to morality. We have complete freedom of will and therefore when our actions are judged this is also a judgment of our worth because we must take moral responsibility actions that are freely taken. Conversely, our actions are completely determined by causality and therefore with our actions are judged this is not a judgment of our worth because we are only causally responsible for our action. Both these approaches are highly problematic.
The idea of absolute freedom of physical action is absurd. I cannot will myself to levitate, to become invisible or to be immortal. I can only will what circumstances allow me to achieve. Even the most ardent believer in free-will has not trouble with the credibility of inevitable plot development in novels. They can accept ideas of fate, luck, reflex action, justification by appeal to circumstances and so on. They hold the view that we have free-will while rarely acting on that premise, not for its own merits, but, because they feel they need to believe it to underpin their morality.
The determinist has very little difficulty making a convincing case for the causal inevitability of our actions. This case is in essence as strong or as weak as the case for causality in general, an overwhelmingly powerful appeal to an a priori notion. But he is left with no justification or explanation of his moral nature. He still agonizes or decisions, makes choices, initiates action, feels guilt and feels self-righteousness. This is not a very satisfactory position to live with.
The obvious way around this impasse, is to separate ideas of morality from those of free-will. We can then accept without any hedging that we are probably part of a determined universe and still inquire into the morality of our actions. The concept of thought is a modeling process is helpful here.
How does choice have real meaning without freedom? When we make a choice, the process amounts to, first, having identified a point in our model of reality where there is figuratively or literally a fork in the road. We learn from our experience to identify these points. Secondly, we model the outcomes of various optional courses of action. This involves using our model as a simulator in order to predict the consequences of possible acts. Thirdly, we evaluate in light of our model of reality the advantages and disadvantages of each course of action. We then have made a choice and initiate the most advantageous action. There is nothing in this process which contradicts a material determination and yet it is still a very real and specific process for which that is not better name than 'choice'. There is nothing illusionary about the choice; our modeling system has just modeled and evaluated a situation and selected the preferable action. But the choice, real as it was, was not free.
If a person has chosen an action but not with any freedom of will, then what is the reasonable position to take on the question of responsibility? Is that person to be judged when his actions are judged? There is no way to escape this assessment. In the same way that we must make value judgments of pieces of meat, chairs, or the weather, we must human evaluate actions. As an action is the product of a particular model of reality, we will be judging that model when we judge the action. If we wish to evaluate people in order to decide who to eat with or sleep with or cross the street to talk to or borrow from etc., we have only the patterns of their actions to guide us. We do connect people with their actions. In a sense we think of a person's nature or personality or disposition as an attribute of their model. We predict their actions by trying to understand their thinking. This is slightly different from what is normally thought of as a moral judgment.
It is a very convincing argument that we act in our own self-interest, especially in our own biological self-interest. We see ourselves as the products of several billion years of evolution, in which all of our ancestors make relatively successful choices. It is, however, not always clear

what our biological self-interest is. It certainly must include our own life and comfort. It would include the life and comfort of our relatives to the extent that they share our genetic inheritance. Further, it appears to include the cohesion and strength of any social groups with which we or our relatives make common cause. It may include the protection of the mechanisms that regulation the competition between social groups. In many cases we are simply not able to model our long-term biological self-interest with the same precision with which we can model our immediate self-interest. It is for this reason that I feel we are endowed with a moral sense. The function is not to judge others or ourselves in terms of good and evil. Its function is to weigh the scales of evaluation with the accumulated wisdom of both genetic and social evolution as to what is in our long-term self-interest. The way this is done is to use guilt/shame and self-righteousness/pride. When we do not do the selfish thing because we do not want to feel guilty this is equivalent to not doing the selfish thing because we have reason to believe is a not actually in our long-term self-interest. We do, of course, take actions occasionally when we know we will feel guilty; the advantages have outweighed the disadvantages even including the disadvantage of feeling guilty. To summarize, the modeling, our thought, is capable of choice but this is a process of modeling and not free from causal relationship. The criterion for choice is self-interest; but, we are not capable of modeling our long-term biological self-interest without a sense of morality.

POSTSCRIPT

This was written in the mid ‘80s. A lot has changed since then. So what would I change if I was writing it now. I would change many very small things which make little difference to the general argument. I would also add the following more substantial items.

1) It has been shown many different ways that our conscious ‘now’ is about a third of a second behind the reality ‘now’. That appears to be how long it takes to produce an update of the model. We can be shown to make decisions some time before we are conscious of making a decision.
2) There has been work on altruism and ‘tit-for-tat’ that would add to the discussion of morality. Our sense of morality may also have the function of protecting our social groups from individuals who would take advantage of altruism.
3) There is work on ‘mirror neurons’ that would add to the discussion on modeling others. Our understanding of others may depend on such neurons.
4) It might help to include a discussion of differences between analogue and digital computers, sequential and parallel processing, and programmable verses purpose built machines, programmed against learning computers. If computers are going to be the almost only metaphor used in describing the brain then it show be more flexible.
5) I think that the way we use language might illustrate many of our a priori concepts.