CONVERGING LINES OF EVIDENCE ON MIND/BRAIN RELATIONS

EDWARD F. KELLY

Like most other working scientists, I find it altogether too easy to become completely absorbed in the endless details of ongoing research. For me, this conference therefore represents an exceptional opportunity to escape this mole-like everyday perspective and hopefully to acquire some renewed sense of location and direction within a larger intellectual context.

The central problem before us is whether the properties of minds are reducible to—or identifiable with—those of brains. I will assume without further discussion that this is largely, if not wholly, an empirical problem and not "just" a metaphysical or linguistic one. I further believe that the empirical status of the problem has shifted considerably within the last several decades. Certainly one of the perennial obstacles in approaching the relation of mind and brain has been the scarcity of detailed information about either term. Without exaggerating the volume of knowledge we now possess, I believe it is still fair to say this ignorance has been substantially reduced and is being further reduced at an accelerating rate.

The main contention of this paper is, therefore, that we may be approaching the point where converging lines of evidence from several sources can intersect to produce a decisive resolution of the mind-body riddle. In developing this theme, I will draw principally upon the three disciplines which have engaged me professionally, namely, cognitive psychology and computer modeling of human cognitive capacities, psychophysiology and contemporary neuroscience, and parapsychology. I certainly do not mean to imply that these supply the only relevant considerations—they simply represent relevant scientific areas with which I feel I have sufficient direct acquaintance to make realistic judgments. I particularly regret my scant acquaintance with the pertinent technical literature of philosophy. In what follows, I hope I do not commit too many philosophical *faux pas*, but perhaps my fellow participants will be able to correct or at least identify such defects as shall undoubtedly creep into the presentation.

An essential aspect of my thesis is that the complementary perspectives provided by these various disciplines can and must be brought to bear simultaneously. Historically, there have been relatively few discussions of these problems which successfully crossed traditional disciplinary lines, and none to my knowledge which has brought all the parties together with a realistic appraisal of the possible contribution of each. Many contemporary discussions continue to be infected either by an appalling ignorance and even disregard of neighboring disciplines, or by naive over-valuation of their results.

In this paper, therefore, I hope to take a small step in the proposed direction, by describing the framework of ideas within which I perceive this convergence of perspectives to be developing. Because of the large amount of ground to be covered, the presentation will necessarily be a bit telegraphic—really more of an outline than a finished presentation. However, with the exception of the fragmentary and speculative lattermost parts of the discussion (section 5), I feel confident that the framework is basically sound and that in a more detailed treatment the remaining gaps could, in principle, be filled.

1. Background-the biological perspective

Any contemporary discussion of the mind-body problem must take into account the enormous advances made during this century in our understanding of the brain. New manifestations of mind appear everywhere to be closely associated with modifications of structure or process in brains. In phylogeny, for example, we see the gross correlation across species between behavioral complexity and the level of organization of the nervous system. The rapid post-natal mental evolution of the human infant likewise is associated with massive structural and functional changes in the developing brain. And as human adults we are all presumably familiar with numerous facts—such as the normal diurnal fluctuations of consciousness and the effects of mild cerebral trauma induced by fatigue, alcohol and other substances, thumps on the head and so on—which reflect in a general way the dependence of mind on brain. But what about the details? In recent decades, freed at last from the rigid behaviorist prohibitions of an earlier era, brain researchers have begun to "open up the black box," using a vast variety of clinical, pharmacological, surgical, electrophysiological and behavioral approaches in increasingly successful efforts to understand what brains can do and how they do it.

Scientists and philosophers confronting the mind-body problem even as recently as a century ago knew only that the brain was in some gross undifferentiated way the organ of mind. Today we know a great deal more, although our knowledge undoubtedly remains in many respects extremely primitive relative to the virtually unimaginable complexity which the human brain presents. We know a lot about the structure and operation of neurons and even lower-level constituents. We also know something about the basic structural organization of the brain and, most important of all, we have begun to learn a fair amount about its functional organization, the contributions that different components of brain structure make to the overall content and organization of behavior.

Correlates of mind have spatial distribution inside the brain. In particular, the kinds of performances that we regard as most characteristic of human mental functioning are known to require specific, coordinated patterns of activity of a multiplicity of identifiable brain regions (particularly regions of the cerebral cortex), patterns which differ systematically from function to function (Luria, 1973). Similarly, one of the most conspicuous trends characterizing recent work in the psychology of perception is its increasingly physiological cast. In vision research, particularly, theory has become steadily more deeply anchored in increasingly detailed knowledge of structural and functional characteristics of the physiological pathways from eye to brain. Indeed, a variety of subtle phenomenological aspects of human vision such as color, brightness and numerous contrast effects have even been directly linked to identifiable physiological mechanisms, some as far out in the periphery as the retina (e.g., Lindsay and Norman, 1972; Uttal, 1973 and, for some especially elegant examples, Julesz, 1971 and Ratliff, 1965).

In short, the connection between mind and brain so far appears to grow ever tighter and more detailed as scientific understanding of the brain advances. Certainly, it is not unreasonable, particularly in light of these successes, to assume as a working hypothesis that this process can continue indefinitely without encountering any insuperable obstacles, and that properties of minds can ultimately be explained entirely by those of brains. For most scientists, however, this useful working principle appears to have become more like an established fact, or even an unquestionable axiom. Whereas the plain man, confronted by the phenomenological duality of private conscious experience and the public physical world, opts for interactive dualism in the tradition of Descartes, current opinion among brain scientists on the mind/body question—if they stop to think about it at all—strongly gravitates toward one or the other of the theories that assume a strict one-to-one relationship between mental events and physical events in the brain.

Epiphenomenalism, the old favorite, acknowledges mental events as a separate category of existents, but denies that they have causal efficacy; conscious experience is merely a passive by-product of causal processes occurring wholly within the brain. Identity theory and its variants, the current favorite, asserts that the duality exists in terminology only, and that mental events (should they ever need to be referred to) are in some difficult-to-analyze sense "identical" to certain subclasses of physical events occurring in the brain (Beloff, 1965; Shaffer, 1968). The identity theory in particular thus attempts to provide an intellectually respectable theoretical foundation for what everyone would presumably agree is currently sound scientific practice. Let us call this majority theoretical viewpoint and its relatives collectively "the official brain doctrine."

2. *Heresy in the temple of neuroscience—some opponents of the official brain doctrine*

In light of the above, it is of particular interest to study the opinions of those few eminent brain scientists who have publicly repudiated the official doctrine in favor of more or less explicitly interactionist positions. The principal examples I have in mind are Charles Sherrington (1955), Wilder Penfield (1975), and J. C. Eccles (Popper and Eccles, 1977), although there may well be others. I will not attempt to review their arguments here in any detail, but will focus chiefly on Eccles as the most contemporary and radical representative of the group. In no case do I think the arguments succeed.

For Eccles as well as the others, the bottom line of the argument is clearly the difficulty one intuitively feels in reconciling the unity of conscious experience—and especially self-consciousness—with spatially and temporally diverse physical processes in the brain. This feeling by itself, however, is certainly insufficient to establish interactionism. Therefore, it is of much greater importance to examine Eccles' more substantive arguments, in which he attempts to supply positive evidence of aspects of human activity that, in his view, resist explanation in terms of brain function alone.

For concreteness, let me give a few examples. First, he cites phenomena which suggest to him that the relation between neural events and conscious experience may not be 1:1 in the manner required by the official doctrine. In particular, he cites the work of Libet on "antedating." The idea may be expressed as follows:[1] The elaboration of conscious experience following stimulation takes time;

in particular, suppose that it takes longer to hear an isolated *soft* drum-beat than an isolated *loud* drum-beat. Yet, if loud and soft beats are interspersed in a sequence of equal short intervals, we hear them in their proper temporal relations—hence the experience of the soft beats is somehow referred backward in time. Eccles doubts that brain mechanisms alone can account for such effects (Chapter E-2, Section 9.2; Chapter E-7).

Second, he reviews some of the modern work on sensory coding in mammalian nervous systems. For example, while giving due credit to the impressive results on hierarchies of feature-extraction mechanisms and the like in the visual system, he correctly points out that the known processes provide only a fragmentary analysis of the optically available information, and do not yet come anywhere close to explaining the overall visual synthesis which ultimately emerges. A fortiori, we cannot at present explain the larger synthesis of perceptual experience over multiple sensory modalities. Although Eccles does not explicitly argue or even claim that this synthetic activity could not in principle be carried out by the brain itself, it is evident that this is what he in fact thinks (Chapter E-2, Section 10.3; Chapter E-7, especially p. 358).

Similarly, he describes work by Kornhuber on elementary voluntary movements (Chapter E-3, Section 19). The critical finding is the gradual development of a surface-negative potential in the cerebral cortex preceding simple movements such as a self-initiated finger-tap. This "readiness potential" takes surprisingly long to develop (almost a full second), and is distributed surprisingly widely over the cortex (p. 285). Eccles argues that this is best explained as the effect of a weak influence cumulatively exerted upon the ongoing activity of the cortex at critically poised "liaison" regions by an independent entity which he terms "the self-conscious mind," and which he feels must be responsible for all the "higher" mental phenomena, including the integration of both perceptual experience and voluntary motor activity.

In this last example, we are shading over into the sort of higher-level argument that is really central to Eccles' presentation. I think he sees the power of his theory as residing not so much in an incontrovertibly superior ability to explain *particular* phenomena as in the ease with which it can assimilate a wide *variety* of phenomena. Other things being more or less equal, this would certainly be a defensible attitude. In this case, however, I fear other things are far from equal. None of the individual examples given is remotely persuasive as an argument against the possibility that brain processes might in principle account for everything. Even in the antedating experiments, the information about the presence of the weak stimulus is already available at the cortex in the required time frame; the fact that we do not know at present how to account for the antedating phenomenon in terms of brain processes, therefore, by no means constitutes a demonstration that in principle we never will.

The theory is central, an imaginative construction that goes very far beyond presently available data. In fact, I think it is fair to say that Eccles uses his neurophysiology, in essence, to rationalize or interpret a pre-existing theory whose real intellectual origins lie elsewhere. He has been a declared dualist for many years; his dualism is apparently founded chiefly upon the Sherringtonian intuition that the phenomena of brain and mind are irreducibly incommensurate (*cf.* Sherrington, 1955, pp. 205, 218, 244, 252), but is also appears to have sources in his Catholicism, in a powerful experience (on OBE?) he had as a teenager (Popper and Eccles, 1977, p. 357), and in a naive belief that only dualism can "restore to the human person the senses of wonder, of mystery and of value" (p. 374). In the present exposition, he has updated his familiar dualistic theory considerably in light of recent neurophysiological discoveries. In particular, his hypothesis that the self-conscious mind interacts with critically-poised regions of the brain is now greatly refined. The critical regions of the brain where interactions with the self-conscious mind can occur now seem to him most likely to involve the upper layers of vertically oriented cortical modules recently revealed by Szentagothai's microstructural investigations. Moreover, on the basis of Sperry's split-brain studies and related clinical investigations he further believes the relevant modules to be confined exclusively, or at least primarily, to the dominant hemisphere, in particular to areas concerned with language comprehension (Wernicke's area, Brodman areas 39, 40) and to the prefrontal cortex (p. 363).

While the increasing specificity of the theory is welcome and good, the theory itself remains, in my view, highly premature. The phenomena adduced in its support, while perhaps consistent with a dualistic interpretation, cannot possibly serve to establish it since alternative possibilities of explanation based on brain processes alone are nowhere decisively excluded. Moreover, and more seriously, the Popper/Eccles theory seems to me to exemplify conspicuously what I take to be the central scientific problems confronting any dualism:[2]

1) First there are the diachronic questions: Where does immaterial mind first appear in the phylogenetic spectrum, and at what point in the course of ontogeny? Closely related to these (but more interesting to me) is the synchronic or functional question: How shall we

determine which human functions are mediated entirely by mechanical processes in the brain, and which functions require intervention of the immaterial mind? There are many complex regulative and goal-directed processes going on which never require conscious attention (indeed may well require the *absence* of such attention), or which once did but no longer do. Some of these, such as driving a car, clearly entail complex perceptual and judgmental processes of a rather sophisticated sort quite impossible for lesser brains. In perception, thinking and the rest, where do brain-processes end and mind-processes begin?

2) Supposing that we feel compelled to admit such a division of labor among two or more irreducible parts, what account can we provide of the relations among them? In this respect, Eccles' account is hardly more satisfactory than the crude instrumental analogies of earlier times. Instead of an immaterial pianist playing the bodily piano, we are offered a self-conscious mind which scans, interprets and reciprocally influences the activity of "open" cortical modules. This is surely unsatisfactory. Except by way of very indirect technical arrangements, we have absolutely no acquaintance with events taking place in our brains.[3] Moreover, this and all similar metaphoric accounts seem to me implicitly to confer upon the mind an entirely unrealistic degree of independence of the body. Although as a parapsychologist I am well aware that many transient experiences of such separability have been reported, these experiences so far remain open to alternative explanations, and in no way alter the fact that the vast bulk of every individual's mental life appears to go forward in strict conjunction with, and dependence upon, a *particular*, *intact* and *functioning* brain.

3) Finally, what really is gained by ascribing complex mental functions of whatever sort to an entity of the type Eccles proposes? We are still faced with exactly the same problems we might have hoped to solve in terms of observable or inferable brain mechanisms; the problems simply recur in a relatively inaccessible domain. William James (1909) identified this problem with characteristic precision:

"You see no deeper into the fact that a hundred sensations get compounded or known together by thinking that a 'soul' does the compounding than you see into a man's living eighty years by thinking of him as an octogenarian, or into our having five fingers by calling us pendactyls. Souls have worn out both themselves and their welcome, that is the plain truth. Philosophy ought to get the manifolds of experience unified on principles less empty. Like the word 'cause,' the word 'soul' is but a theoretic stop-gap—it marks a place and claims it for a future explanation to occupy." (p. 209)[4]

Problems of this magnitude should certainly make us tremble before the prospect of embracing interactionism. At least, we should certainly demand much more compelling evidence than Eccles has so far been able to offer. I think he is entirely correct in stressing how little we really know, as against the hubris of many contemporary brain scientists, and indeed, his theory or something like it might ultimately prove to be correct. On the other hand, he displays a large amount of hubris himself, by dismissing, in effect, any possibility that mechanistic models will ever come significantly closer to explaining the central capacities of the human mind.

The dismissal takes two forms, really. On the one hand, Eccles seems strangely reluctant to entertain the possibility that future discoveries about the brain itself might run counter to his hypothesis and strengthen the official doctrine. Much more damaging—and indeed this seems to me the most fundamental flaw in the book—neither Popper nor Eccles seems to have any appreciation of the fundamental developments that have transformed cognitive science in recent decades. For example, particularly in their "Dialogues" it becomes clear that one of the central facts driving them toward dualism is the *activity* of the mind as against what they seem to believe is the necessary passivity of a mechanical apparatus based on brain processes alone. Related to this, they point to a number of phenomena—such as incorrigible illusions, reversing figures, recognition of accuracy of memory retrieval, awareness of awareness, etc.—which suggest a kind of hierarchical organization of levels of integration of conscious experience; although no compelling arguments are provided to show that any portion of any such hierarchy necessarily lies outside the brain itself, they seem to believe that the phenomena cited can in principle only be explained by a dualistic theory of the sort they propose. In a particularly revealing example which exemplifies both of these attitudes, Popper rails at length against the old linear associationist theory of thinking, correctly describing it as a transposition of earlier and outmoded philosophical doctrines into an artifically simplified conception of conditioning processes in the brain. He and Eccles both seem to believe that simplistic theories of this sort are the inevitable outcome of mechanistic theorizing and that, in demolishing them, they are *ipso facto* establishing interactionism.

In this expectation, they are without doubt gravely mistaken. The image of the brain that appears implicitly to guide and limit their assessment of the capacities of mechanism—as it had Sherrington's decades previously—is essentially that of a gigantically complicated automatic telephone switchboard, a network of passive relays. But, since Sherrington's time, a vastly more powerful mechanistic technology has come into being, one which is certainly capable of capturing many more of the properties—such as activity and hierarchical structure—that Popper and Eccles correctly view as fundamental to human mental processes, and that in fact have long since become cornerstones of modern cognitive theory. Let me briefly review these developments.

3. The analysis of mind as an information-processing system–cybernetics, artificial intelligence, and cognitive psychology

The old concept of "machine"—and no doubt for most of us still the everyday concept—is that of a physical contraption which transforms energy by means of pushes and pulls involving gears, pulleys, shafts and so on. The fundamental insight underlying the modern developments is the recognition that these physical arrangements are quite secondary. The essential attribute of the machine is rather that its normal behavior is *bound by rule*.

This insight opened the way to an enormous enrichment of our concept of mechanism, beginning with the contributions of logicians and mathematicians in the 30's and 40's and continuing into the present day. These developments, furthermore, immediately began to have profound impact on the questions we are addressing in this conference.

For example, it was quickly recognized that machines could transform information as well as energy, and that a machine could in principle utilize information about its performance to regulate its own behavior. These ideas had immediate and urgent practical application in the construction of servocontrolled antiaircraft systems during the second World War, but their general theoretical implications for the understanding of behavior were no less apparent. Rosenblueth, Wiener and Bigelow (1943) showed that, from the point of view of an external observer, a device constructed on this principle of "negative feedback" behaved purposively, i.e., as a teleological mechanism. Thus, mechanism penetrated one of the last strongholds of vitalist thinking.[5]

These analogies were developed much more systematically by Wiener in his influential 1948 book *Cybernetics* (significantly subtitled *Control and Communication in the Animal and the Machine*). In addition to providing a general analytic theory of feedback control processes, Wiener gave a number of concrete examples of physiological phenomena that seemed to fall within the province of the theory.

Nevertheless, the direct applications of cybernetic theory at this level have remained relatively limited. The real power of the ideas emerged in conjunction with the extremely flexible applications technology provided by the digital computer.

To appreciate the full significance of these developments, it is necessary to follow the generalization of the concept of "machine" to its ultimate development in the hands of the English logician A. M. Turing and several others.[6] Turing devised an abstract representation which formalized his intuition of the core meaning of the concept of mechanism, as applied to the theory of computable functions. Any computation can be regarded as the transformation of a string of input symbols by a sequence of rule-governed steps into a string of output symbols. A procedure which is guaranteed to lead to the desired output in a finite sequence of steps is called an "algorithm" or "effective procedure." Turing envisioned a machine consisting simply of a read/write head operating on an infinite tape ruled off into squares. The behavior of the machine is governed by a set of rules organized as quintuples; given the machine's current state and the input symbol written in the current square, these rules instruct the machine to change state, write a new symbol on the tape and move one square left or right. By altering the number of states, the size of the vocabulary and the behavioral rules, an immense variety of behaviors can be realized by such devices. In fact, Turing argued persuasively that *anything* that would naturally qualify as an algorithm can be represented by a suitably constructed machine of this sort. He further showed that he could construct a "universal" Turing machine which would simulate the behavior of any other Turing machine. The intuitive notion of "effective procedure" was thus explicated in terms of the formal notion of "realizable by a Turing machine." That this is not an arbitrary result, but, in a fundamental sense, exhausts the possible meaning of the term is strongly suggested by the fact that several other workers arrived at provably identical results from widely different starting points.

Because of their utter simplicity, Turing machines do even very simple things, such as adding two numbers, in extremely cumbersome ways. Their direct significance is theoretical, not practical. But the link to a vast field of applications was made when it was demonstrated that any digital computer is also in effect a universal Turing machine. McCulloch and Pitts (1943) further showed that networks suitably constructed from artifically idealized neurons could realize the elementary logical functions of computers, and other workers promptly demonstrated that the same capacities could be realized using richer elements that more nearly approximated the characteristics of real neurons. Thus, it became apparent that brains in principle have access to the capacities of Turing machines. They may conceivably have additional capacities as well, but if so—and this is the essential point—these capacities probably lie beyond the reach of understanding based on mechanistic principles alone. To the extent that mind and brain are governed by formalizable rule, their activities can in principle be modeled on a suitably programmed general-purpose digital computer. Indeed, to those who are sufficiently committed *a priori* to mechanistic principles, the very existence of a given class of behavior virtually *entails* the possibility of such formalization.

There were other more specific theoretical results that further strengthened this point of view. Consider, for example, some of the early results in theoretical linguistics. The skills underlying use of language, certainly one of the central manifestations of human mentality, defied explanation in terms of simple conditioning and S/R models (although many were offered). Chomsky (1963) and others showed that the possible classes of formal models of these skills (generative grammars) formed a hierarchy, in which the weakest or most highly constrained class (finite-state grammars) was obtained from the strongest or least constrained class (unrestricted rewriting systems) by the application of progressively severe constraints on the form of the permissible rules of the grammar. Furthermore, the hierarchy of grammars corresponds to a hierarchy of classes of automata derived from Turing machines in parallel fashion. Formal results from automata studies thus transferred to the analysis of candidate grammatical theories. Chomsky (1957) was able to show that the existing psychological and linguistic proposals for theories of language, when formalized, corresponded to the weakest members of the hierarachy of grammars, and that these grammars were in principle too weak to account for systematic structural properties of many kinds of sentences in natural languages such as English. He was thus led to his famous theory of transformational grammar as the weakest class of theory which is still strong enough to account for the known grammatical facts of language. The result that the corresponding automata are weaker than Turing machines greatly strengthened the presumption that linguistic behavior might be formalizable for computer modeling.[7]

The central idea that computers and brains could fruitfully be regarded as two varieties of a more general class of information-processing mechanisms quickly took root. The ground was very well prepared. Indeed, these developments seem to me an inevitable outcome of our western scientific tradition. This is not meant disparagingly, however; I have stressed these results about Turing machines and so on precisely to underscore the impressive depth of the theoretical foundation on which the ensuing applications rest, a foundation which I feel has not been adequately appreciated by many critics of this kind of work.

In practice, the applications came a bit slowly at first. In part this was due to purely technical factors. The early computers were small, slow and highly prone to malfunction. More important, in the early days programming a computer was an exasperating business requiring detailed familiarity with the lowest-level details of its hardware organization. The primitives of the available languages referred to data structures and operations virtually at the hardware level. As the technology advanced, however, machines grew larger, faster and more reliable, and so-called "higher-order" languages (such as FORTRAN) were created whose primitives refer to data structures and operations at a level which is relatively natural for human problem-solvers. Programs written in the higher order language congenial to the user are then translated by special programs into the internal language of the computer for subsequent execution by the hardware.

I mention these details because they relate to the other main reason for the delay, which is more theoretical in nature and involves a basic question of strategy. The fantastic complexities of the brain can obviously be studied at many different levels from the molecular or even sub-molecular on up. At what level shall we seek scientific explanations of human mental activity? Many scientists, particularly those working at lower levels of the hierarchy of approaches, assume that events at the higher levels are in principle reducible to events at lower levels, and that reductive explanations employing the concepts of the lower levels are necessarily superior or more fundamental. Like many other psychologists, I strongly disagree with this view. Furthermore, I believe it is easily shown to be false (or at best inefficient), even without appeal to the (controversial) notion that higher levels may display "emergent" properties not predictable from those of lower levels.

Consider, for example, the problem of understanding the behavior of a computer playing chess under the control of a stored program. It seems obvious that we might observe the behavior of its flip-flops forever without gaining the slightest real understanding of the principles that govern its behavior. Similarly, one of the essential characteristics of both human and animal behavior is that functional invariance at a higher level may be coupled with extreme diversity at lower levels. Thus, the rats whose cortex Lashley surgically disturbed in efforts to locate the engram could wobble, roll or swim to the food box, and I can write a given message with either hand or even with my feet if necessary. Attempted explanations based on activities of the participating muscle-groups, neurons, and so on would probably never get to the essential feature which is the invariance.

Thus, it seems appropriate in general to seek a distinctively "psychological" level of explanation of mental processes. For the computer simulation approach in particular this entails identifying an appropriate set of elementary information structures and processes which seem powerful enough to account for the relevant behavior. [Strictly speaking, for a consistent mechanist and identity theorist there is another requirement. The hypothesized information structures and processes should conceivably be physically realizable in the brain given the known properties of neurons or, at least, should not be demonstrably inconsistent with those properties. However, it is important to recognize that successful use of the computer as a tool of psychological understanding does not require the (obviously false) presumption of literal identity of computers and brains.][8]

By the late 1950's and early 1960's, a number of higher-order languages had been created which emphasized the capacities of computers as general-purpose information-transforming devices, rather than their algebraic and "number-crunching" capacities. These languages (such as IPL-V, SLIP, LISP, etc.) provided facilities for creating and manipulating complex tree-like data structures consisting of hierarchically ordered and cross-referenced lists of symbols. Structures of this sort played a central role in theoretical linguistics, and in this and other ways the new languages seemed to many workers to fall at about the right level of abstraction to support realistic efforts to model many aspects of human cognition.

Previous generations of workers had been obliged either to try to force mental processes to conform to artificially simple but relatively rigorous behavioristic models, or to lapse into the uncontrolled introspection and mentalistic speculations of an earlier era. Now, suddenly, we were provided with a conceptual and technical apparatus sufficiently rich to express much of the necessary complexity without loss of rigor. The black box could be stuffed at will with whatever mechanism seemed necessary to account for a given behavior. A complicated theory could be empirically tested by implementing it in the form of a computer program and verifying its ability to generate the behavior, or to simulate a record of the behavior. The seminal ideas of Craik (1943) could at last be put into practice.

Enthusiasm for the information-processing approach to human cognition fairly crackles through the pages of influential early books such as Miller, Galanter and Pribram (1960). Their enthusiasm has been amply justified in the ensuing flood of theoretical and experimental work based on these ideas. In addition to the specific efforts at computer modeling of cognitive functions that is our main concern here, the rise of information-processing psychology has also brought in its train a healthy reawakening of broader interests in many of the old central concerns of psychology, such as mental imagery, thinking and consciousness.

The depth to which the metaphor of the human mind as an information-processing machine has penetrated cognitive psychology can, perhaps, best be appreciated by referring to recent introductory books such as Arbib (1972), Lindsay and Norman (1972), Oatley (1972) and many others. There can be no doubt that this metaphor has been and will continue to be enormously productive scientifically. However, it currently appears well on the way toward hardening into literal dogma, a companion to the official brain doctrine. I wish now to examine some indications that as dogma it is false.

4. Some difficulties in the mechanical concept of mind

A convincing demonstration that the essential capacities of the human mind can be embodied in technical artifacts would, in practice, presumably be fatal to an interactionist theory such as proposed by Eccles. Although it would not be strictly disproven—since the mind or "parts" thereof might conceivably be immaterial, but still governed by mechanist principles—it would certainly seem rather superfluous. On the other hand, a convincing demonstration that *any* capacity of the human mind can in principle *not* be embodied in such artifacts refutes mechanism and opens the way to various alternative possibilities.

In the following discussion, I wish to maintain a behaviorist perspective on the concept of "mental capacities" in order to avoid being immediately drawn into first-person questions about whether robots could be conscious, have feelings and so on (cf. Anderson, 1964). This is not because I think that consciousness is not important, or irrelevant, but because I think I can establish the point I wish to make without becoming entangled in these very difficult and quite possibly inconclusive arguments. Similarly, I follow Turing (1950) in excluding as inessential various matters pertaining to the physical embodiment of the artifactual intelligence—its sensory and motor appendages and so on. By "essential capacities," therefore, I shall mean rather the kinds of general behavioral skills that ordinary persons display in perceiving, thinking, producing and understanding language and so on. I am also willing to relax the requirements still further by allowing that a demonstration of such skills could be convincing even if not terribly general, provided that the principles embodied in the machine were convincingly generalizable. This loose characterization will suffice for present purposes.[9]

In his justly famous 1950 article, Turing himself expressed the expectation that by the end of this century machines will have advanced to the point where ". . . the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted" (Anderson, 1964, p. 14). Although this remark is perhaps a bit guarded, it is clear from other parts of his discussion that he fully expects machines to display what any reasonable person would be obliged to regard as general intelligence. Indeed, the major part of the article is devoted to clearing away various kinds of possible *a priori* objections to this thesis. Of course, as Turing acknowledges, this is still all theory. How far has this work progressed?

With the appearance of suitable higher-order languages in the late 50's and early 60's, numerous research groups set to work to endow computers with capacities for varied kinds of skilled performance, including game-playing (especially complex games such as checkers and chess), problem-solving (for example, proving theorems in the propositional calculus), pattern-recognition (such as recognizing sloppy hand-written characters), question-answering (in restricted domains such as baseball) and natural language translation.

An important strategic difference quickly appeared, separating these efforts roughly into two streams, often called computer simulation (CS) and artificial intelligence (AI) respectively. Workers in CS remain faithful to the aim of increasing psychological understanding, in that they seek to reproduce not only particular kinds of human performance, but also possible models of the mechanisms by which humans achieve those performances. AI workers, by contrast, disavow any direct interest in psychology and seek rather to achieve high-level performance by whatever means possible. Although a powerful simulation might seem rather more interesting to us than an AI result of equal power (in the unlikely sense that each reproduces the same range of skilled behavior), I think it is reasonable to admit that results of either type would in practice have substantially the same impact on our central question; and hence I will discuss the two streams of work together.

I will not attempt to review the substantive accomplishments here.

Persons interested in following the history of such work can refer to landmark publications such as Feigenbaum and Feldman (1963), Minsky (1968), Schank and Colby (1973), Winograd (1972) and Winston (1975). Suffice it to say that many of the individual pieces of work represent considerable intellectual and technical achievements. Computers have already done a number of quite remarkable things and there can be no doubt that in the future they will do many more.

However, without intending to disparage these attainments of digital computers, I must add that so far they still fall very far short of what anyone could plausibly describe as general intelligence. I would not even bother to mention this were it not for the extremely inflated image many people have of the progress of this work, an image which in large part has been promoted by the research workers themselves. Many extraordinarily grandiose prognostications have been made on the basis of relatively modest concrete accomplishments.

Of course, these predictions might conceivably some day prove correct. The theoretical foundation is deep and after all the work is still in its infancy. Might we not simply be in the early stages of an evolutionary process in which machines will inevitably attain at least the equivalent of human cognitive abilities?

Recently, an important book has appeared which profoundly challenges this vision of the future of mechanical intelligence. Hubert Dreyfus (1972) systematically attacks both the progress and the prospects of CS and AI. He begins by reviewing early work in game playing, problem-solving, language translation and pattern recognition. Work in each domain was characterized by a common pattern of early success, followed by steadily diminishing returns. Subsequent work in artificial intelligence, according to Dreyfus, has fared little better, achieving its limited successes only by operating in artificially constrained environments that do not exemplify basic difficulties handled with conspicuous ease by everyday human intelligence. Dreyfus argues that the extreme efficiency of human intelligence, which becomes progressively more apparent as we move toward problem domains more typical of its ordinary application, rests on a complex of related abilities which he terms fringe consciousness, ambiguity tolerance based on efficient use of context, essential-inessential distinctions and perspicuous grouping. Human performance is characteristically guided by an overall grasp of the problem situation, with essential aspects at the foreground of attention set against an organizing but implicit background. Phenomenologically, the *situation* is primary. Specific facts or features of the situation may only become evident by a deliberate attentive effort of a quite secondary sort. By contrast, Dreyfus argues, for the computer all facts must be specified in advance as explicit bits of atomic data; whatever crude representation of the situation the computer can achieve is necessarily constructed by explicit calculation upon these situation-independent facts. Building upon these and related observations, Dreyfus attempts to establish his thesis that central human mental skills are *in principle* not reproducible on digital computers.[10]

I will not attempt to develop Dreyfus' thesis here in further detail. Suffice it to say I have no doubt whatsoever that he has identified a cluster of problems which at the very least constitute extremely difficult problems of practice. My confidence in this view arises in large part from extensive and sobering experience of my own in attempting to fathom what might be involved in enabling computers to understand and use common words in English (Kelly and Stone, 1975, Chapter 4). Our group was principally concerned with the applied technical problem of reducing lexical ambiguity in keypunched English text, the practical aim being increased precision and power of automated content analysis procedures. In approaching this problem, we constructed a concordance of some half-million words of "typical" behavioral science text. This concordance identified the frequently-occurring words and supplied information about their typical ranges of use. Although our main concern was to build computer routines to recognize pre-established word senses with a useful degree of accuracy, we also went to some pains to determine whether the brute facts of everyday language as we are seeing them could successfully be captured by existing theories of semantic representation.[11] In a word, they could not. I argued that all of the existing schemes utterly failed to capture what to me had become the most characteristic property of word-meaning, namely, a felt Gestalt quality or wholeness, at a level of generality which naturally supports extensions of usage to novel but appropriate contexts. The available proposals could only represent the content of a general term by some sample of its possible particularizations; thus, no computer system in existence then or now could distinguish systematically between metaphorical truth and literal falsehood. For details, see Kelly and Stone (1975). Clearly we were approaching from a different angle a domain of problems strongly overlapping those identified by Dreyfus. Although we acknowledged that the properties we stressed appeared likely to be very difficult to capture in a digital representation, it is only candid to admit that, at the time, it was quite unthinkable to me that they might be not merely difficult but *impossible* to capture.

Since that time such a notion has become thinkable, but I remain unconvinced that it is correct. I cannot go into details here, but I do not find Dreyfus' theoretical arguments convincing, and I fear that his polemical fervor occasionally blinds him to the merits of the views he attacks. I will content myself with two main points of rebuttal.

First, it strikes me as significant that the difficulties primarily revolve around "Gestalty" phenomena with a distinctly analog flavor. It still seems at least vaguely conceivable that these difficult but fundamental qualities of embodied cognition might be deeply rooted in the special properties of the nervous system as a particular kind of computing machine. Whereas early discussions focused on the all-or-nothing neuronal spike discharge as the equivalent of a digital relay, in more recent times we have become increasingly sensitized to the complex analog processes of spatial and temporal summation occurring prior to the axon hillock. Workers in CS and AI have rather cavalierly assumed that they could safely disregard these low-level structural and functional properties of neurons and pitch their efforts at a level of abstraction which happens to be congenial both to them and to present-day machines. However, what we may be seeing is that these low-level properties enter into the overall computational possibilities of the brain in a much more fundamental way than we have heretofore suspected. Perhaps the "missing" cognitive attributes are reducible to such low-level characteristics. It is at least suggestive that the relevant phenomena appear continuous with phenomena that are both ontogenetically and phylogenetically of lower order (see Kelly and Stone, loc. cit.). Also relevant here are the facts that, in humans particularly, the cerebral hemispheres tend to show considerable differential specialization, and that in right-handed adults the right hemisphere displays more prominently, though certainly not exclusively, various kinds of Gestalty skills, particularly in relation to visual synthesis. Thus, one is led to imagine the possibility of hybrid devices with a mixture of digital and analog components. Alternatively, schemes might be invented for capturing analog properties with a digital representation, and even if that representation were so clumsy as to render implementation as a broad scale transcomputable in the sense of Bremermann (1977), enough might be accomplished to convince us of the generality of the underlying principles.[12]

Second, it seems likely that computers can in fact progress a good deal further toward grasping situations than they so far have. It is particularly interesting that in recent years workers in AI have showed increased appreciation of the fact that the only *general* intelligence we know is that associated with the human body and brain. Thus, intelligence artificers now show great interest in studying human cognitive development, and there are many interesting signs of convergence between researchers in cognitive psychology and in AI. In particular, both groups have now clearly recognized (apparently quite independent of Dreyfus) the critical role played by our overall sense of the situation in guiding various forms of intelligent behavior. Thus, several proposals have recently emerged for representational schemes by which computers might be enabled to construct and manipulate more global understandings of this sort (see for example Neisser, 1976; Schank and Colby, 1973 and especially Minsky, 1975). The basic notion is to have the computer store "frames" representing typical situations at varying levels of generality. In a given situation, the appropriate frames are then to be extended or specialized in a manner appropriate to that particular case. It should be stated, however, that efforts in this direction so far remain largely programmatic, with relatively primitive concrete accomplishments. I would also argue that the notion of general *frames* harbors severe difficulties quite analogous to those which I earlier argued arise in the case of general *terms*, and hence that these too may be intimately bound up with specialized analog properties of neural computation. The two classes of problems thus seem to me strongly interdependent.

To sum up, although there appear to be major difficulties ahead, I nevertheless think that the kinds of problems noted so far might not prove fatal for the mechanistic approach, and that on these fronts we have to remain agnostic for the present. I concur with Weizenbaum's (1976) judgment that we cannot at this point reliably predict the possible progress of computer mentality. We must wait and see what can be accomplished. Meanwhile, we should strongly support further efforts along these lines, for as CS and AI evolve they are constantly sharpening our understanding not only of what computers can do, but of what we ourselves do in a broad range of cognitive activities. Even apart from its ultimate success or failure, each new attempt to extend computer mentality into a novel area of human performance forces us in the first instance to try to understand in detail what capacities are presupposed by that performance. As our understanding of the capacities of minds deepens and becomes more detailed, it thus should become increasingly possible to judge to what degree those capacities may be explainable in terms of the properties of brains.

There remains, however, one area of human capacities regarding which I think we can conclude immediately that present day mechanistic principles must necessarily fail. I am referring, of course, to parapsychological abilities. Taking advantage of the theoretical equivalence of brains and computers under the mechanistic theory, we can simply ask what it would take to endow a computer with psi ability. Although there are some slight complications that I will not go into here, it seems quite clear that generalized psi ability, and particularly psychokinetic ability, could not conceivably arise from any kind or degree of complication of the basic capacities presently available to machines. It is certainly not a matter of developing more complex or subtle algorithms, data structures, or the like, but of providing one or more *extra capacities* of a fundamentally novel sort.

I was astonished to discover, upon re-reading his 1950 article, that Turing himself had anticipated this argument. Indeed, he evidently took it quite seriously, since it appears last in his list of *a priori* objections to the possibility of mechanizing human abilities, a list which appears to be presented, at least roughly, in order of increasing difficulty. His attempted solution to the problem, moreover is patently defective—as he probably knew, since it requires a telepathy-proof room(!)—and seems to me to reflect little more than his ardent and freely expressed wish that the phenomena should simply go away. In this he was of course adopting a strategy that has been widely practiced by contemporary philosophers and scientists.

I believe that this argument shows clearly that mechanism is false, or at least incomplete. It succeeds without appeal to consciousness, by displaying a behavioral capacity of the human mind that lies entirely beyond the reach of mechanistic principles as currently understood. Although psi phenomena thus play a pivotal role in the argument, I want to stress that they seem to me only the leading edge of a pattern of connected developments, other elements of which might well eventually point to similar conclusions, and all of which merit continuing study by those of us interested in the mind-body problem. Let me explore this pattern a bit further.

5. *Beyond mechanism-what?*

First to review: I began by characterizing identity theory and its variants as the official brain doctrine, to which most contemporary scientists subscribe, at least in their more philosophical moments. Subsequently, I introduced the mechanistic theory of mind as a companion doctrine which is (or should be) accepted by a consistent materialist as the framework within which the explanation of cognitive capacities in terms of brain processes must develop.[13]

We saw that Eccles and a few others have repudiated the official doctrines, on grounds that there seem to remain large gaps between the properties of brains and those of minds, gaps which they believe can only be filled by postulating interaction between brain processes and the activities of an immaterial mind. However, I rejected these skeptical conclusions as premature. The experimental evidence presented was insufficient to refute the official doctrine, and the theoretical position seemed to raise at least as many fundamental difficulties as it resolved (or perhaps dissolved). Furthermore, Eccles and the others have not adequately appreciated the full possibilities of the mechanistic approach, nor the depth of the theoretical foundation upon which it rests.

However, precisely because of the depth of that foundation, we were able to pierce through the official doctrines at one small but significant point, based on the inability of mechanistic principles to account for the human capacity for psi performance. I further suggested that psi abilities may represent just the sharp corner of the wedge, and that broader kinds of mental capacity may subsequently turn out, as both neuroscience and the cognitive sciences advance, to be likewise irreducible to mechanical brain activity.

These latter possibilities so far remain to my mind uncertain, but even the one small definite result already in hand opens the way to a wide spectrum of theoretical positions lying beyond the currently orthodox doctrines. I wish now to outline very briefly where things seem to me to be headed.

Some extension or modification of prevailing views is clearly going to be required, but just how drastic need these revisions be? I believe that fundamental considerations of scientific policy enjoin us to confine the damage as much as possible, and to accept only such minimal revisions as we can judge are absolutely necessary to repair the limitations of our present outlook. In this way, we shall also minimize the risk of running afoul of the powerful empirical and theoretical considerations raised in the first two sections of this paper.

Although my ideas are not definite and clear enough to spell out in any detail, the minimum scenario, which represents my current position, would seem to go something as follows: Practically all of the behavioral aspects of human mentality may be accounted for by brain processes alone. It may subsequently prove necessary to refer to consciousness to fill a few small but irreducible gaps in our understanding of the synthesis of perception and voluntary motor action, as well as memory, thought and so on; but perhaps consciousness somehow "emerges" in physical systems that reach a certain threshold of complexity. (Indeed, some have suggested that machines themselves might become conscious, and it certainly would be very difficult to *prove* that they are *not* conscious even now, however unlikely that may seem). The difficulty with psi phenomena, however, seems to require some more radical kind of emergentism, which might or might not be associated with the appearance of consciousness itself. Nevertheless, the kind of theory I am pointing toward could conceivably remain very close to present orthodoxy. Taking advantage of the greatly expanded horizons of modern physics, as opposed to the old-style materialism, it might remain a physicalist theory, and I think in the same way it could still be an identity theory, although perhaps of a sort somewhat different from what most current proponents have in mind. Let me denote theories of this type collectively by the term "augmented official brain doctrine."

I would be very happy to leave the story here, in hope that these vague outlines and obscure concepts will take definite and intelligible form with the steady progress of future scientific research. Such a story is, to me, intellectually and temperamentally congenial in a general way, and it is specifically congenial to the sort of research program on possible physiological correlates of psi processes in which our group at Duke is making such a heavy investment.[14]

It is with considerable reluctance, therefore, that I acknowledge increasing discomfort with the kind of picture I just sketched, and several variants of it. I fear that the gaps may ultimately prove substantially larger than I have so far indicated. Although I and (I presume) most other scientists would be willing to tolerate a good deal of straining of the minimum scenario to accommodate new difficulties, at some point one reaches one's elastic limit. That will occur in different places for different persons, of course, depending on complex temperamental factors and the range of evidence they are willing to consider seriously. Although I emphatically do not feel compelled to abandon the augmented brain doctrine, I do feel considerable, though still ill-defined, strains in maintaining it.

My thinking is really just beginning to take shape in these areas, so again I will not try to go into details at this time, but I would like to mention just briefly, some of the points at which I sense, with varying degrees of clarity, *possible* conflicts emerging between properties of human cognition and the physical capacities of the brain. Please note carefully the emphasis on the word "possible"; I am not claiming that any of the phenomena to be noted currently provide a compelling argument for abandoning the augmented brain doctrine, but only that they reveal points of strain that may merit further investigation in this light. I simply wish to convey in rough-and-ready fashion my overall "sense of the situation."

Some of these phenomena involve "normal" behavioral functions, and can thus be added to the kinds of problems raised in the previous section. For example, consider human long-term memory. It is really an astonishing system, and one about which stunningly little is known even at this late date. There does exist a consensus that possible physiological mechanisms exist in the brain sufficient to account for the required volume of storage capacity. (These might involve, for example, macromolecular configurations and/or structural changes of neural junctions that affect overall patterns of functional activity of the brain.) But there are scarcely any remotely realistic proposals describing details of mechanisms both for storage and especially for *retrieval*. Meanwhile, recent experience with computer systems has cast the situation in an interesting new light. The largest computer memories are still orders of magnitude smaller than a typical human memory, by anybody's way of counting. They are, of course, also physically relatively enormous and functionally very different. The structural elements of machines are individually exceedingly much faster and more reliable than their neural counterparts. Nevertheless, computer "knowledge-bases" of even relatively modest size already present formidable problems of organization and access. Experience with such systems is rapidly sharpening our appreciation of the extreme flexibility and efficiency of human memory processes (e.g., Norman et al., 1975; Anderson, 1976). At the very least it is clear that we need drastically more powerful concepts of data representation and retrieval; hence the increasing efforts to develop so-called "associative memories," which, although so far rather unsuccessful, at least seem to be going in the right direction (Kohonen, 1977).

The memory problem is a good one because it is now widely recognized by both psychologists and intelligence artificers to lie at the very core of generalized intelligence. Whereas earlier workers tried to approach functions such as pattern recognition, thinking and sentence understanding in a relatively piecemeal and isolated way, it has become increasingly apparent that all human skills, and particularly our central everyday skills, are embedded in a continuous interplay between present performance and stored information. While recognizing that fundamental new discoveries in this area may arise both from neuroscience and from research with computers, I think it is appropriate to entertain the question whether the informational characteristics of human memory might not prove inconsistent with the physical characteristics of the brain. This question has, of course, been raised previously, both by persons interested in parapsychology such as Bergson (1911) and Driesch (1935) and, in a curiously fence-straddling way, by the physicist Elsasser (1958), but as our knowledge advances it may eventually become possible to pose it in a sharply quantitative way.

Other phenomena that make me uncomfortable are phenomena that suggest degrees of precision and reliability in at least some mental processes that seem intuitively hard to reconcile with what we might hope to extract from a fundamentally analog device working in a statistical way with components of low precision and reliability (cf. von Neumann, 1956, 1958). For example, the kind of "complete" memory-record phenomenon suggested by certain hypnotic states and by Penfield's cortical stimulation studies with epileptic patients undergoing brain surgery might, if better established, fall in this area. A related phenomenon is that of so-called "eidetic" imagery, as best exemplified to my knowledge in the extraordinary experiments of Stromeyer (1970). It must be pointed out, however, that many psychologists—perhaps because they feel similar discomforts—simply refuse to believe any of these memory results. A further and related phenomenon which certainly occurs, however (although it remains to my knowledge very poorly studied), is that of calculation prodigies. I find it quite astonishing that there have existed people capable of multiplying together two arbitrary thirty-digit numbers, or taking the 23rd root of a 201-digit number, all in their heads, often in very short times and sometimes in the absence of any conscious effort. For an entertaining account see Barlow (1961); F.W.H. Myers (1961), with characteristic prescience, also sensed the possible significance of these phenomena.

Other phenomena more directly involve consciousness and alterations of consciousness, and shade over toward the realm of the paranormal. Another oddity pointed out by Myers concerns so-called "glove anesthesias" and kindred phenomena, in which hysterically induced paralyses or anesthesias may correspond to a "psychological" unit of the body, in apparent disregard of the underlying physiological organization. Another possibly important behavioral manifestation, if genuine, is the concurrent multiple use of a single basic set of skills. Although psychologists are beginning to think that, with suitable training, people can probably do more things at once than we customarily suppose, this generalization seems to apply only to relatively divergent things, and conspicuously fails as the simultaneous tasks become more alike (Neisser, 1976). Although I do not have a reference for this at the moment, I remember reading with considerable sense of alarm that Mrs. Piper would occasionally communicate while in trance with as many as three separate sitters at once, writing with both hands and speaking to the third sitter, all apparently simultaneously. Observations of this sort if firmly established are at the very least of great psychological interest and would pose severe difficulties for prevailing notions of mind-brain relations. A further set of phenomena occurs in conjunction with clinical cases of the "multiple personality" type. The particular occurrence that interests me most in this connection is the phenomenon of co-consciousness, in which the sphere of awareness of one personality may entirely include that available to one or more others. Some interesting physiological observations supportive of clinical descriptions of this phenomenon have been published by Ludwig et al. (1972). Finally, I have read a large number of accounts of transcendental or mystical experiences, and talked with several individuals who have themselves had such experiences. Although better documentation is certainly required, I think we must reckon with a substantial probability that such experiences may often result at least temporarily in marked changes in overall pattern and level of cognitive functioning, at least occasionally including an influx of paranormal capacities. Suggestions of this sort are certainly potentially open to empirical investigation.

In addition, we have not yet exhausted the catalog of paranormal phenomena. In particular, I have so far not mentioned any of the kinds of work falling under the general heading of survival research. A potentially very important transitional class of cases involves so-called near-death experiences. Although to my knowledge no case has yet been reported with sufficiently detailed physiological records corresponding to extended episodes of veridical experience, it already seems likely that at least some such cases confront us with the prospect of elaborate perceptual and cognitive activity apparently taking place under physiological conditions we would have expected to be insufficient to support it. It remains to be seen just how sharply this apparent conflict can be drawn, but it seems virtually certain, given the increasing interest of medical personnel and the increasing availability of facilities for detailed physiological monitoring in hospital environments, that more information will be forthcoming in this area in the near future (Sabom and Kreutziger, 1977).

The more general survival literature of course confronts us still more starkly with evidence suggestive of mental activity occurring without its normal physiological accompaniments. Obviously I cannot discuss that evidence here, nor is my familiarity with it sufficient to qualify me to do so. On the other hand, I have made moderate efforts to acquaint myself with it, and I feel obligated to acknowledge that, on the basis of my studies so far, the weight of the evidence seems to me to tilt slightly in favor of survival of something undeniably mental. Along with most other parapsychologists, I do not find the evidence anywhere near sufficient to exclude decisively alternative explanations based on "ordinary" psi abilities of the living, but, despite its *a priori* repugnance, I feel that the survival hypothesis must be taken quite seriously.

Survival in any form, and reincarnation in particular, would appear to entail some kind of extracerebral representation of memory. Establishment of such phenomena would thus certainly overthrow even the augmented official doctrines, completely and decisively, and very likely precipitate us into some kind of pluralism with all the attendant difficulties. Despite its notorious problems survival research therefore clearly bears on the issues at hand in the most direct way imaginable, and for this reason it seems to me to merit intensive further effort.

I think I shall end my catalog here, for now. I repeat that my purpose has been only to draw attention to a number of areas that all of us interested in the mind-body problem—parapsychologists and non-parapsychologists alike—might do well to keep under surveillance. I will be very interested to hear from other members of this group, and from subsequent readers, both comments on these topics and suggestions of possible further topics for the catalog. For now let me quickly bring this long paper to a close.

6. Conclusions

I am impressed on re-reading this how little of it is genuinely new. Whatever virtue it has lies, I think, primarily in the organization of the material. Let me summarize the main themes.

The lines of evidence reviewed here are converging, but not yet convergent. Part of the difficult lies simply in getting the different intellectual disciplines to acknowledge each other's existence. I have tried to promote this kind of synoptic view by displaying the conflict between parapsychological findings and current doctrine in an unusually well-defined context, one which in addition brings into sharp focus some basic relationships between what parapsychologists are doing and various other contemporary approaches to the study of mind/body relations.

We parapsychologists have a tendency to feel victimized by lack of attention from neighboring disciplines. While agreeing that our findings deserve more consideration from others than they presently receive, I also feel that we have been almost equally culpable in our historical tendency toward professional isolation. I hope this essay will help contribute to the rapid demise of this kind of suffocating academic parochialism, on all the converging fronts.

Meanwhile, we can already see that mechanistic materialism of the currently fashionable sort is surely false or at least incomplete, as demonstrated by its inability to account for psi phenomena and possibly various others. Exactly what will take its place will depend, in a way I believe is currently not predictable, on future results turned up by the approaches reviewed here and perhaps others as well. For the present, something quite close to our present views seems to me to remain provisionally tenable, with significant, but still relatively minor, adjustments needed to patch up the presently identifiable shortcomings of current doctrine. At the same time, there are ominous signs of more fundamental difficulties ahead that may ultimately shatter even this extended picture and drive us on toward much more radical revisions of our basic ideas on the mind-body problem.

FOOTNOTES

1. This analogy was used by Eccles in his invited address to the 1976 P.A. convention, where he also very inappropriately characterized the antedating phenomeon as "precognition" (Eccles, 1977).

2. Here I omit mention of further difficulties of a more narrowly philosophical sort, such as the problems of *individuation* and *identification* of immaterial minds; see for example, Shaffer (1968). I also find it difficult to imagine conditions under which Eccles would *abandon* his theory, and quite surprising that Popper does not press him on this!

3. Popper is aware of this difficulty, and proposes a slightly better analogy: Just as in efficient reading we sometimes seem to grasp meanings directly, without conscious experience of individual letters or even words, so "In perception we read the meaning of the neuronal firing pattern of the brain and the meaning of the neuronal firing pattern is, as it were, the situation in the outside world which we try to perceive." (p. 418). Although this analogy helps, it is still radically insufficient. In particular, we are entirely unable to perform any kind of redirection of attention to neural processes, analogous to our ability to become aware of letters, words, and so on.

4. In light of the foregoing remarks it is particularly astonishing to find Eccles claiming (e.g. p. 366, 512) that his theory, unlike the official brain-process theories, does not require recourse to a homunculus which embodies most of the original problems!

5. Even as recently as McDougall (1911), it had still been intellectually possible to think of embryological development as lying outside the scope of materialistic theories. But, by the time Sherrington delivered his Gifford Lectures in 1937-38, it had already become clear that these tremendously complex and goal-directed events are at least very largely matters of physics and chemistry. Defenders of vitalist theories had thus been forced to retreat again, and the apparent purposiveness of human and animal behavior had become more than ever a cornerstone of their defense.

6. What follows is a brief heuristic introduction to a very complicated subject. For further detail, see, in ascending order of difficulty, Weizenbaum (1976), Trakhtenbrot (1963), or Davis (1958).

7. I am deliberately ignoring, for present purposes, the important (and controversial) distinction Chomsky makes between *competence*, as a formal representation of the skills underlying language use, and *performance*, or the exercise of those skills in speaking and understanding. Also irrelevant to the point I am making is the fact that the currently

most successful efforts at computer modeling of language do not explicitly incorporate transformational grammars.

8. Some of the early workers chose to attempt understanding the brain at a very basic level, by using the computer to model neurons and networks of neurons. Others carried out theoretical studies of so-called "self-organizing systems" (aggregates of simplified neuron-analogs modifiable through experience by elementary principles of conditioning and the like), which, it was hoped, could gradually come to display complex activities. These approaches have never yet come anywhere close to supporting interesting behavior, however, and although work still continues along such lines we will not be concerned with it here.

9. To the degree that machine intelligence begins to approximate our own, the problem of precise specification of adequate criteria will no doubt increasingly call for careful philosophical analysis as foreshadowed—quite prematurely, I think—in Anderson (1964) and numerous other discussions. My general impression is that many philosophers have been too easily influenced by pronouncements issuing from the mechanist camp. In what follows, I am displaying the progress of this work in the most favorable possible context, by restricting discussion to everyday cognitive phenomena which take place on what William James liked to call the "sunlit terrace" of the mind. Even there, I will argue, ultimate triumph of the mechanist approach is by no means assured.

10. Several of Dreyfus' arguments appear in less developed form in various chapters of a much earlier book by Sayre and Crosson (1963). Of particular interest are some remarks by Wittgenstein on mechanical mathematics. More recently, another critical attack on artificial intelligence has appeared, this time from an insider, a computer specialist from MIT (Weizenbaum, 1976). Weizenbaum is entirely conscious of the theoretical underpinnings of AI, and consequently much more reluctant than Dreyfus to argue the nonreproducibility thesis (although he verges on it in numerous places). His central argument is rather a moral one; that because it cannot emerge from a fully human situation, one which takes into account our biological uniqueness among other things, whatever understanding computers may ultimately develop will necessarily be in fundamental respects alien to human understanding; and, therefore, that there are many kinds of tasks—such as psychotherapy—which computers should never be permitted to perform.

11. The main such theory, that of Katz and Fodor (1964), built upon the central notions of Chomsky's transformational linguistics. It is particularly congenial to potential formalization, in that it depicts the representation of the meaning of a sentence as resulting from a rule-bound calculation operating upon semantic representations of the individual words in the sentence; furthermore, the meanings of the words themselves are claimed to be analyzable into an underlying universal set of discrete, atomic features or logical structures built out of such features.

12. The model described by Pribram at this conference and elsewhere is perhaps one concrete example of this sort.

13. Of course, strictly mechanistic theories no longer hold sway even in physics proper, and it is appropriate to ask what implications quantum-theoretic ideas may have for our understanding of brain activity. I am not qualified to discuss the subject in any detail, but I should like to indicate the main possibilities that seem to be open. On the one hand, the micro-structure of the brain certainly offers ample scope for quantum-level effects to take place, such as electron tunneling through synaptic clefts and the like. More broadly (and as noted previously at least as early as von Neumann, 1958) even at higher levels the actual operation of the brain is very unlike that of a deterministic machine, but rather saturated with probabilistic or statistical properties. None of this automatically has any bearing on the conclusions so far drawn, since effects of the sort mentioned might in principle be handled by extending the class of machines to include stochastic machines, i.e., machines which incorporate random elements. On the other hand, it also appears that quantum theory in one or another interpretation might be enlisted in support of a wide variety of much more radical departures from current notions. Certainly it forces us at the very least to acknowledge characteristics of physical systems that go far beyond

those contemplated by Newtonian physics, and in this negative sense has considerably softened the artificially sharp traditional dichotomy between matter and mind. Recent discussions by physicists such as Wigner and Walker further suggest the possibility of much more direct and positive contributions of physical theory to the mind-body problem, but I am unable at present to judge how far these discussions have actually progressed, and how they might supplement the framework outlined in this paper.

14. I should make clear at this point that *failure* to discover any physiological correlates of psi would not for me constitute independent evidence supporting interactionism. Although such failure might well be consistent with that interpretation, it could also readily be understood in terms of currently available neuroscience. It is entirely conceivable that psi processes may indeed have a physiological representation, but one which for any of several reasons would elude detection by present or future technology. For arguments suggesting a more hopeful view of the research prospects in this area, see Kelly, 1977.

BIBLIOGRAPHY

Anderson, A. R., (ed.) *Minds and Machines*, Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Anderson, John R., *Language, Memory and Thought*, Hillside, N.J.: L. Erlbaum Associates, 1976.

Arbib, M. A., *The Metaphorical Brain-An Introduction to Cybernetics as Artificial Intelligence and Brain Theory*, New York: Wiley-Interscience, 1972.

Barlow, F., *Mental Prodigies*, New York: Greenwood, 1969.

Beloff, J., "The identity hypothesis: A critique," in Smythies, J. R. (ed.) *Brain and Mind*, London: Routledge and Kegan Paul, 1965, 35–61.

Bergson, H., *Matter and Memory*, New York: MacMillan, 1911.

Bremermann, H. J., "Complexity and transcomputability," in Duncan, R. and Weston-Smith, M., (eds.) *The Encylopedia of Ignorance*, Oxford: Pergamon Press, 1977, 167–174.

Chomsky, N., *Syntactic Structures*, The Hague: Mouton, 1957.

Chomsky, N., "Formal properties of grammars," in Luce, R. D., Bush, R. R., and Galanter, E., *Handbook of Mathematical Psychology*, vol. 2, chapter 12, 323–418, New York: John Wiley, 1963.

Craik, K. J. W., *The Nature of Explanation*, Cambridge U. Press, 1943.

Davis, M., *Computability and Unsolvability*, New York: McGraw-Hill, 1958.

Dreyfus, H. L., *What Computers Can't Do-A Critique of Artificial Reason*, New York: Harper and Row, 1972.

Driesch, H., "Memory in its relation to psychical research," *Proceedings of the Society for Psychical Research, 43*, 1935, 1–14.

Eccles, J. C., "The human person in its two-way relationship to the brain," In Morris, J. O., Roll, W. G., and Morris, R. L. (eds.) *Research in Parapsychology 1976*, Metuchen, N. J.: Scarecrow Press, 1977, 251–262.

Elasasser, W. M., *The Physical Foundation of Biology-An Analytical Study*, London: Pergamon Press, 1958.

Feigenbaum, E. A., and Feldman, J., *Computers and Thought*, New York: McGraw-Hill, 1963.

James, W., *A Pluralistic Universe*, New York: Longmans, Green & Co., 1909.

Julesz, B., *Foundations of Cyclopean Perception*, Chicago: University of Chicago Press, 1971.

Katz, J. J., and Fodor, J., "The structure of a sematic theory," in Fodor, J. and Katz, J. J., *The Structure of Language*, New York: Prentice-Hall, 1964, 479–518.

Kelly, E. F., and Stone, P. J., *Computer Recognition of English Word Senses*, Amsterdam: North-Holland, 1975.

Kelly, E. F., Physiological correlates of psi processes. *Parapsychology Review* **8**, 1977, 1–9.

Kohonen, T., *Associative Memory: A System-Theoretical Approach*, Berlin: Springer-Verlag, 1977.

Lindsay, P. H., and Norman, D. A., *Human Information Processing. An Introduction to Psychology*, New York: Academic Press, 1972.
Ludwig, A. M., Brandsma, C. B., Wilbur, C. B., Bendfeldt, F. and Jameson, D. H., "The objective study of multiple personality," *Archives of General Psychiatry*, 26, 1972, 298–310.
Luria, A. R., *The Working Brain-An Introduction to Neuropsychology*, New York: Basic Books, 1973.
McCulloch, W. S., and Pitts, W., "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, 5, 1943.
McDougall, W., *Body and Mind: A History and a Defense of Animism*, London: Methuen, 1911.
Miller, G. A., Galanter, E. and Pribram, K. H., *Plans and the Structure of Behavior*, New York: Henry Holt, 1960.
Minsky, M., (ed.) *Semantic Information Processing*, Cambridge, Mass.: MIT Press, 1968.
Minsky, M., "A framework for representing knowledge," in Winston, P. H., *The Psychology of Computer Vision*, New York: McGraw-Hill, 1975.
Myers, F. W. H., *Human Personality and Its Survival of Bodily Death*, Abridged edition, edited by Susy Smith, New Hyde Park, N.Y.: University Books, 1961.
Neisser, U., *Cognition and Reality-Principles and Implications of Cognitive Psychology*, San Francisco: W. H. Freeman, 1976.
von Neumann, J., "Probabilistic logics and the synthesis of reliable organisms from unreliable components," in Shannon, C. E. and McCarthy, J., (eds.) *Automata Studies*, Princeton, N.J.: Princeton University Press, 1956.
von Neumann, J., *The Computer and the Brain*, New Haven: Yale University Press, 1958.
Newell, A., Shaw, J. and Simon, H., "Elements of a theory of human problem-solving" *Psych. Review*, 65, 1958, 151–166.
Norman, D. A., Rumelhert, D. E. and LNR Group, *Explorations in Cognition*, San Francisco: W. H. Freeman, 1975.
Oatley, K., *Brain Mechanisms and Mind*, New York: E. P. Dutton & Co., 1972.
Penfield, W., *The Mystery of the Mind*, Princeton, N. J.: Princeton University Press, 1975.
Popper, K. R. and Eccles, J. C., *The Self and Its Brain: An Argument for Interactionism*, Berlin: Springer-Verlag, 1977.
Ratliff, F. *Mach Bands: Quantitative Studies on Neural Networks in the Retina*, San Francisco: Holden-Day, 1965.
Rose, S. R., *The Conscious Brain*, New York: Vintage Books, 1976.
Rosenblueth, A., Wiener, N. and Bigelow, J., "Behavior, purpose and teleology," *Philosophy and Science*, 10, 1943, 18–24.
Sabom, M. B. and Kreutziger, S., "Near-death experiences," *Journal of the Florida Medical Association*, 64, 1977, 648–650.
Sayre, K. M. and Crosson, F. J., *The Modeling of Mind-Computers and Intelligence*, Notre Dame: University of Notre Dame Press, 1963.
Schank, R. C. and Colby, K. M., (eds.) *Computer Models of Thought and Language*, San Francisco: W. H. Freeman, 1973.
Shaffer, J. A., *Philosophy of Mind*, Englewood Cliffs, N. J.: Prentice-Hall, 1968.
Sherrington, C., *Man on His Nature*, Garden City, N. Y.: Doubleday Anchor, 2nd edition, 1955.
Smythies, J. R., (ed.) *Brain and Mind-Modern Concepts of the Nature of Mind*, London: Routledge and Kegan Paul, 1965. (esp. essays by Beloff, Brain.)
Stromeyer, C. F. and Psotka, J., "The detailed texture of eidetic images," *Nature*, 225, 1970, 346–349.
Trakhtenbrot, B. A., *Algorithms and Automatic Computing Machines*, Health, 1963.
Turing, A. M., "Computing machinery and intelligence," *Mind*, 59, 1950. (Reprinted in Anderson (1964)).
Uttal, W. R., *The Psychobiology of Sensory Coding*, New York: Harper & Row, 1973.
Weizenbaum, J., *Computer Power and Human Reason-From Calculation to Judgment*, San Francisco: W. H. Freeman, 1976.

Wiener, N., *Cybernetics-Control and Communication in the Animal and the Machine*, 2nd ed., New York: John Wiley, 1961.
Winograd, T., "Understanding natural language," *Cognitive Psychology*, 3, 1972, 1–191.
Winston, P. H., (ed.) *The Psychology of Computer Vision*, New York: McGraw-Hill, 1975.

DISCUSSION

PRIBRAM: I have some comments to make—two minor and three major ones. I think you can believe Strohmeyer. There's just no way that one can fake the Julez patterns. He presented one pattern to one eye and then five days later the other patterns to the other eye. There is no way that one can cheat on that particular test. The second thing—about hysterical or hypnotic anesthesia—it is neurologically reasonable to think of the phenomenon as being thalamocortical since it is at the cortex that the whole body representation comes together. All of the various senses come together at the thalamocortical level.

Now, the major things that I want to discuss are the meanings of three words that you used: *mechanism*, *material* and *memory*. I think they may have different definitions for different people in the audience, and I think we should be very clear that they're used differently by different disciplines. Take "mechanism," for instance. If I use the word with my friends in physics, they say, "Well, nobody believes in mechanism any more." What they mean is Newtonian mechanism. When *you* talk about mechanism are you talking about Newtonian mechanism? It must be very clear that mechanistic analysis is a prevailing way for psychologists and physiologists to be thinking; but this approach is different from that of modern physics.

Thus, my second point—regarding "material." Modern physicists don't believe in a material world any more than they believe in a "mechanistic" world. The term "material" is at the same level of discourse as is Euclidean geometry and Newtonian physics. In that realm, things look hard and time and space are stable coordinates. But problems arise in this realm: For instance, light is transmitted from, let's say, a star to us through *nothing*. That's a funny kind of macro-universe. Further, in the quantum and nuclear micro-universe, the materiality of material disappears entirely.

Now thirdly, the problem of "memory." There is no reason why we can't imagine a brain process that operates in the same way that Turing's process works—a brain process that stores an unlimited amount of information which can be retrieved. We simulated a content addressable memory back in the 1960s in our laboratory. Furthermore, we've shown that the frontal lobes have an "execute" function and the posterior convexal cortex functions as a "fetch" mechanism to a content addressable memory. The brain's memory operation is not like that of current computers, that's all.

KELLY: I accept particularly the first of your major points, and if I'd had more time for oral presentation I would have, I think, made similar remarks myself. In a narrow way I don't think that the presence of quantum effects on the brain necessarily changes anything. In a deeper sense, it well might, but unfortunately I'm not qualified to comment on that. I do think, however, that quantum theory, as we've seen recently in parapsychology, might well be compatible with any of the wide spectrum of alternatives lying beyond the majority viewpoints as I characterize them. I also agree about the dematerialization of matter and its potential implication that the traditional hard and sharp outline between the material world of billiard balls and the mental world has softened considerably in the present century. That's an important point. As to the memory part, I look forward with great interest to what you have to say.

STORM: With respect to the analogy with computers, a Turing machine, and this is the case in approximately any other formulation, requires the notion of a fetch-execute cycle to drive the activity of the machine and make it anything other than a static device that might do something. Do you think that it is important or unimportant to look for the real equivalent of the fetch-execute cycle in the human mind or brain and, if you do, do you have any ideas about how we're going to identify it?

KELLY: I certainly have no ideas about how we might identify it. The success of work on computer simulation of psychological processes does not require any assumption of literal identity between the brain and machines, which is obviously false, anyhow. What it does require is that there is some level where things are *similar enough* that structures and processes in the brain can be usefully represented by structures and processes in the computer. The difficulty has been in providing such a level. Now, I do think, that people working in these areas have underestimated the degree to which properties of human cognition are dependent on these low level structural properties—brains—and it may be their inability to provide, say, analogies of the sort you would like to hear about, which may be part of the difficulty.

EHRENWALD: I was much impressed with Dr. Kelly's brief summary and criticism of the Eccles/Popper book and I would like to take issue with the point which he raised when he said, "Psi cannot be embodied in the mechanistic scheme." This is, of course, a statement describing the Popper and Eccles' observations. But the fact is that psi or "heteropsychic" experiences or ordinary perceptions and volitions cannot be embodied in a mechanistic scheme. There is an ultimate gap in our ability to account for such transactions in both parapsychology and psychology. Popper and Eccles have tried to fill this gap with a theory which invokes a kind of homunculus, as the last resort. Still, ultimately there remains an epistemological gap. Even if we are dealing with such a homunculus, the next question is: who begot the homunculus?

We are arriving at an infinite regress taking us from the homunculus to the son of homunculus, and so on. Nevertheless, I feel that we parapsychologists can live quite comfortably with this gap as long as we realize that we are not worse off than psychologists or philosophers at large. We can do our experiments and we can make our clinical observations even without being able to account for the very last step in the personal experience or the volitional act. We can do that without suffering from an epistemological hernia, so to speak. Now, you mentioned that certain death-bed experiences could be used as an example for a non-mechanistic event associated with personality or the brain. I pointed out in my book, *The ESP Experience: A Psychiatric Validation*, that all these death-bed observations published by Kübler-Ross, Moody and others are subject to very serious objections. I believe that those people who have reached the threshold of death and come back have never been dead. What they experience is the result of anoxia in the brain and the return of circulation to the damaged brain. When their circulation comes back it produces a sense of euphoria. There are all sorts of visual experiences which have something to do with what happens to the brain as a result of anoxia or recovery from it; for instance the seeing of lights, auras, or the like. You even have it after a migraine attack. So all these very important observations have to be taken with a grain of salt. While the observations are valid, the survival interpretations are, in my view, premature, if not outright wrong.

KELLY: I certainly agree with you that there appear at present to be gaps even in our accounts of ordinary cognitive functions, but I do feel that Eccles and Popper are jumping the gun on that. I don't think we really know yet how large those gaps are. I don't think I conveyed this sufficiently strongly, but I think we have to welcome the advance of knowledge in both brain science and artificial intelligence and related fields because no matter what happens, they're going to illuminate us. Either the gaps will become steadily smaller and eventually disappear, or in time we'll become clearer about just where the irreducible gaps are. I personally don't feel that we can guess with any confidence where it's going to come out right now. That was the first point. Second point, about near-death experiences, I should perhaps have made more clear exactly what kind of an experience it is that I think is relevant. At the recent Parapsychological Association Convention in St. Louis, we heard a seven-minute tape from an interview with a patient who had undergone cardiac arrest while in the hospital, who described in considerable and unpleasant detail various kinds of resuscitation procedures that were attempted even though he had no direct medical knowledge of such things, at a time when by current doctrine he should not have been in any shape to attain that information by the usual means. Now, I agree with you that both the facts and their interpretation remain controversial at this point, but I indicate it as an area in which we can hope to see this potential conflict made much sharper by increasing information in next few years.

HONORTON: Ed, I agree with you about the need to push the mechanistic explanations as far as they will go and with the general loose and speculative nature of Eccles' proposals, but I think you're asking too much of Eccles at this stage to criticize his dualistic formulation because he hasn't shown where in biological development mind arises, or how mind and brain interact. These are questions that are unsettled by any theoretical account of relationship between mind and brain at the present time.

KELLY: I may have belabored poor Eccles a bit too hard. My point is really to stress the potential difficulties with that kind of argument. In fact, a good illustration of their difficulties is provided by the book itself, you know, the dialogues between Popper and Eccles, where they really go on and on about this question as to where in the phylogenetic spectrum different mental properties arise, quite inconclusively, I think. But I'm not averse to proposing theories that go well in advance of data. I just want to make clear that this is not one that is actually compelled by data. It's a story that's told about certain kinds of phenomena on the basis of what's really a pre-existing theory that develops, I think, out of other sources.

PRIBRAM: Eccles and Popper don't present a unified dualism in their book. There are really two theories that are proposed, one by Popper, which is an emergent property theory similar to your own, and the other is Eccles; which relies on a universal mind operating on the association cortex. These are opposite ideas of how the mind-brain interaction takes place.