1975 191

From: The Neurosciences: Paths of Discovery. F. G. Worden, J. P. Swazey and G. Adelman, editors. Cambridge: The MIT Press, 1975.

24
In Search of Psyche

Roger W. Sperry

To a beginner in science back in the mid-1930s, it seemed that there could be no more challenging problem at which to aim—as a long-term, ultimate goal kind of thing—than that of consciousness and the mind-brain relation, more acceptably expressed in those days as the problem of the "neural correlates of conscious experience." A naive beginner, of course, could hardly expect to approach a final solution, but it is always reassuring to feel that one's efforts are at least aimed in the general direction of something that might be of ultimate importance. Meantime, as a "brain researcher," one could find plenty of lesser but entirely respectable and more researchable corollary problems along the way, such as perception, learning, and memory.

In the 1930s it already had begun to appear that science might soon close in on the nature of the changes produced in the brain by learning and experience. With the conditioned reflex as a model, researchers had begun to draw hypothetical diagrams for the kind of new brain pathways that must be formed in conditioning to link the conditioning stimulus to the conditioned response. As time passed and further experiments eliminated one neural hypothesis after another, however, it became evident that the nature of the newly formed stimulus-response connections was much more complex than had been at first supposed. Indeed, it began to look doubtful that new nerve connections of any sort were involved in conditioning, or in any brain function. By the late 1930s the connectivity principle as a basis of central nervous integration had come under fire from many directions and was very much in question.

The theoretical impact of K. S. Lashley's brain-lesion studies and his concepts of mass action and cortical equipotentiality had at that time reached their peak. These and related findings pointing up the nonlocalizability of the engram seemed incompatible with any stimulus-response connectionist formula. So also were the Gestalt views of the 1930s that emphasized the control role of excitatory patterns as wholes and their associated metaneuronal "field" forces. The Gestalt or "figure" properties were conceived to transcend the function of individual fiber connections.

The absence of functional specificity in nerve connections seemed to have been substantially confirmed in an extended series of clinical and experimental studies, from all parts of the world, demonstrating that nerves were functionally interchangeable after surgical cross-union. The same was reported to hold with respect to both the transplantation of muscles to take over new functions and the grafting of skin flaps to new locations. It all seemed to confirm the lack of any fixed functional specificity in neural connections, and emphasized an extreme wholeasale plasticity in neural integration that provided for almost unlimited readaptation capacity in the central mechanisms of the brain and spinal cord. The classic account by Stratton (1897) of his own experience in adjusting to the inverted vision produced by wearing an optical device was widely cited in this same connection. To see the world right side up is something we all had to learn, or so it seemed, and it was readily relearnable.

Additional reinforcement for the plasticity and anticonnectionist views of the 1930s came from the field of nerve growth and development, where the prevailing doctrine for more than a decade had proclaimed the outgrowth and termination of developing and regenerating nerve fibers to be diffuse and nonselective. Chemical and electrical selectivity appeared to have been ruled out, leaving only mechanical guidance in command as the primary orienting influence (Weiss and Taylor, 1944). In terms of the evidence then available, it seemed entirely impossible that the enormously intricate and precisely adjusted wiring circuits for adaptive behavior could be grown into a brain directly—that is, organized through the growth process itself without benefit of experience and learning. On these and other grounds it was widely accepted that the nerve networks for behavior could not be inherited, and the idea of "instinct" as an explanatory construct in behavioral science thus reached an all-time low in disrepute in the late 1930s. Under these conditions acceptance of the new upstart discipline called "ethology" remained quite limited, with resistance particularly strong in the United States and the Soviet Union.

Some of the strongest evidence against nerve-connection specificity came from another long series of experiments demonstrating that surgical rearrangements between nerve centers and periphery in amphibians failed to disrupt orderly coordination under conditions where relearning could be excluded. These experiments, pioneered by Paul Weiss (1936), were taken to prove that central nervous integration could not be based on selectivity in fiber connections. As an alternative to the classic connection-switchboard model of integration, a radio-broadcast model was proposed, based on resonance effects involving diffuse morphological interconnection with impulse specificity and selective neuronal and end-organ attunement. Like radio pick-up, the "resonance principle" provided selective response in the presence of diffuse nonselective synaptic connections. Meanwhile, the idea that synaptic relations within the neuropil were not morphologically selective but formed, rather, in an excessive common profusion, seemed to receive further support from C. J. Herrick's (1948) intensive anatomical analyses of the central neuropil in the brain of the tiger salamander.

These many different lines of convergent evidence, combining and mutually reinforcing each other, had built up by the end of the 1930s into quite a substantial and convincing case against the classical Sherringtonian model of central nervous integration. In many quarters it became fashionable to refer to Sherringtonian connectionism when one wished to exemplify simplistic and outmoded naivety.

Anticonnectionist thinking received a further major boost in the early 1940s when it was reported that the largest system of fiber connections in the human brain, the corpus callosum, containing over 200 million elements, could be completely transected in clinical surgery without producing any definite functional symptoms (Akelaitis, 1943). Here again the brain seemed to possess an almost mystical plasticity in its ability to achieve proper orderly function in spite of radical disruptions in its normal wiring plan. Not since the early beginnings of neuroscience had the brain looked so bafflingly obscure and resistant to physiological analysis. How could one even begin to formulate orderly laws and understanding for a mechanism that continued to operate correctly regardless of rearrangement and disruption of its interconnecting parts?

Against this background of general theoretical uncertainty, it appeared a rather poor risk in the late 1930s to invest time and research efforts directly in neural models or working hypotheses concerning such higher psychological functions as learning, memory, or consciousness, or even the conditioned reflex. What we needed first were some better answers at elemental levels, and hopefully some unifying resolution of all the divergent views and issues involved.

When we began to follow up experimentally, one at a time, various aspects of the foregoing plasticity and anticonnectionist phenomena, the results—to our initial surprise—failed to accord with previous accounts. To make a long story short, it was found that motor nerves and muscles, as well as sensory nerves, were not at all functionally interchangeable after surgical transposition, but instead persistently retained their original functions (Sperry, 1945): inverted vision produced surgically by eye rotation showed a fixed persistence, lasting indefinitely without correction by experience and training; and nerve growth in the brain and spinal centers was anything but diffuse and nonselective (Sperry, 1951a,b). To account for the kind of central nerve regeneration found in the new experiments it became necessary to reinstate the old concept of chemotaxis in an even more extreme form and to postulate a degree of cellular specificity and chemotactic guidance more extensive and refined than that previously imagined even by Ramón y Cajal.

Analysis of some of the behavioral effects of reversed vision led us to postulate the function of "corollary discharge" as a mechanism for maintaining perceptual constancy in the presence of disturbing eye, head, and body movements (Sperry, 1950). We had to conclude more generally from the nerve-growth findings that the brain's wiring diagram must, after all, be largely innate, that it is grown in with extreme precision through an enormously elaborate chemical guidance program that is under genetic control, and that it is therefore in very large part inherited. The orderly function found by Weiss to follow nerve disarrangements in amphibians was reconfirmed experimentally; but with our new findings it proved to be explainable in terms of orthodox connectionist principles, thus obviating the need to invoke resonance phenomena or *Erregungspezifität*.

Tests for the postulated electric-field forces of Gestalt theory, conducted in cats and monkeys (Sperry and Miner, 1955), gave results that pointed mainly to the absence of such influences. The data emphasized instead the remarkable capacity of the brain to preserve orderly function when confronted with gross distortions in its internal electric-field pattern and/or disruptions in its horizontal transcortical interactions.

Furthermore, studies involving surgical section of the corpus callosum showed that brain function was by no means left unimpaired. Using various measures to obtain controlled lateralized input and special tests for functional processing within each hemisphere independently, we were able to demonstrate a whole host of cross-communication deficits, first in animals (Sperry, 1961) and later, using the same principles, in human patients undergoing operations for severe, intractable epilepsy (Sperry, 1974a). This last major stronghold of anticonnectionism was shortly to be turned around into a leading bastion for the opposing views. In the evidence relating to the neocommissures, more than from any other place in the brain, we now come closest to tying higher conscious functions to specific cortical-fiber systems.

The conceptual view of the nervous system that emerged from the evidence of the late 1950s had very different properties from the view with which we had had to deal earlier, in the plasticity-equipotentiality period. Differential connection patterns now meant something in terms of functional control. The design and the operating principles of connection circuits, though enormously and perhaps overwhelmingly complex, were subject, at least in principle, to experimental analysis and to lawful formulation. We were now in a much better position to approach such problems as the nature and locus of the new neural connections established in conditioned-response learning. Curiously, the neural model for conditioning that I eventually settled on involved a rejection of connectivity in a sense. I concluded that it had been an error to search for newly formed sensory-motor connections, that we should think of the new sensory-motor linkages observed behaviorally as being effected instead by means of transient cerebral facilitating sets (that is, passing excitatory physiological states) that only temporarily open or prime the requisite stimulus-response connection paths in the conditioning situation.

The long-term "engram" changes in this model (Sperry, 1955) assumed a very different pattern and location, designed to arouse the requisite excitational facilitating set at the right time in the right context. The long-term changes, accordingly, were allocated to the realm of perceptual learning and expectancy, phenomena that involve the association systems of the cortex rather than direct sensorimotor pathways. Furthermore, the changing facilitating set was conceived to be a basic master switching system that would continually alter the functional wiring plan of the brain. By opening and closing different patterns of neural circuits for different functions, this switching system would give the brain, in effect, many different circuit design systems in one, somewhat like different computer programs and subroutines. Switching mechanisms of this sort, based on transient excitatory sets, were felt to account for a large part of the brain's readjustment capacity and its tremendous versatility.

By this time even the remote problem of consciousness had come to look at least a little less remote—mainly through a gradual process of elimination. Among the suggested interpretations of consciousness that it now seemed safe to eliminate was the one in which conscious experience was conceived to be a correlate of isomorphic electric-field forces and volume current changes in the cerebral cortex. This view, engendered in Gestalt psychology and a major contender among theories of consciousness in the 1940s (Köhler and Held, 1949), appeared to be ruled out in particular by the failure of multiple metallic and dielectric inserts (i.e., electric-field distorters) to produce any major disruption in visual pattern perception.

Another view that reached a peak in the 1950s (Delafresnaye, 1954) had consciousness centered in brainstem reticular and centrencephalic mechanisms. The contention was that a person really lives, so far as conscious feeling and experience are concerned, in these deep mesencephalic centers. The neocortex came to be regarded as a relatively recent and superficial adjunct for enhancing and elaborating the basic qualities of conscious experience already evolved in the mesencephalon. Interpretations along these lines had to be largely abandoned in the face of our new findings on brain bisection in which surgical separation of the cerebral hemispheres alone, leaving the brainstem intact, proved sufficient to divide most of the higher psychological functions in cats and primates.

The split-brain findings also helped to resolve another major dichotomy in the theory of mind. A long-standing question in philosophy asks whether conscious awareness is restricted to brains or is, instead, a universal inner property of all things. Do plants, atoms, cities, ships, and molecules all have some form of inner awareness? If one could show that consciousness is selectively localized even within brains, with some neural systems being endowed with the property of conscious experience while others are not, this would be a strong argument against the idea that inner conscious awareness is something universal. If consciousness is lacking in the cerebellum and in other neural systems, if it is lacking even in the cerebrum during dreamless sleep, in coma, or after death, why should we assume it to be present in plants, mountains, or molecules? The added discovery that conscious awareness could be divided into right and left realms by severing a set of forebrain fiber systems at the neocortical level greatly strengthened the view that consciousness is a special and selectively localized property rather than something universal. The balance of the evidence would now appear to favor a prior inference (Sperry, 1952) that consciousness is an operational derivative of activity in particular cerebral circuit systems designed expressly to produce their own specific conscious effects. The implication here of causal action *upon* as well as *from* neural events was yet to be appreciated.

Section of the corpus callosum appears to divide the unified perception of the visual field down the vertical midline, into two inner visual worlds within the left and right hemispheres respectively (Sperry, 1968, 1970a). This and similar split-brain phenomena begin to carry us rather close to where direct correlations can be made between conscious mental experience and activity in specific neural structures. Incidentally, one may now occasionally come across statements, made with the advantage of hindsight, that this callosal syndrome had already been fully recognized and elucidated much earlier in the writings of Maspes, Dejerine, and the German school of neurologists, and had simply been forgotten or overlooked in the English-language literature. Actually, the confusion during the 1940s and early 1950s regarding the corpus callosum and its functions was worldwide. The extensive review in French by F. Bremer and his colleagues, which appeared in

1956 in the *Archives Suisse de Neurologie et de Psychiatrie*, gives a knowledgeable and fair assessment of the world literature and of the confused picture regarding callosal function as it stood at the time.

In any case the split-brain, dielectric-plate, and related findings seemed, along with other developments, to clear the way for a modified approach to the theory of consciousness (Sperry, 1965). This was an interpretation that I had recently come to favor but had been hesitant to publish, mainly because it represented a swing toward mentalism, presenting a conceptual explanatory model for psychophysical interaction. An alternative to psycho-physical parallelisms and psychophysical identity theory, this modified view involved a break with long-established behaviorist-materialist doctrine, amounting almost to a full reversal of the central premise on which behaviorism had been originally founded: instead of renouncing or ignoring the subjective conscious mind, this interpretation gave full recognition to inner conscious experience as a top-level directive force or property in cerebral function.

In this view, which has held up for more than ten years now, the conscious mind is no longer set aside as a passive correlate of brain activity, but becomes instead an essential working part of the brain process and a causal determinant in cerebral action (Sperry, 1970b; 1974b). Consciousness in this scheme is not looked upon as just an inner aspect of the neural process; nor do we relegate it to some metaphysical, epiphenomenal, or other separate dualistic realm. Nor is it dismissed by semantic gymnastics as being unimportant or nonexistent, or as being identical to the neural events. Conscious mental experience in our present interpretation is conceived to be a holistic emergent of brain activity, different from and more than the neural events of which it is composed, and a real phenomenon in its own right possessing causal potency in brain function. At present a more detailed or exact description in objective terms is hardly possible, but this will presumably be achieved with further advances in brain research.

Our current interpretation can be classified as an "emergent" theory of mind, provided it is distinguished from the earlier emergent views of Gestalt psychology. In the present scheme there is no dependence on electric-field forces or volume conduction effects, nor on an isomorphic or topological correspondence between the events of perceptual experience and the corresponding events in the brain. Furthermore, the mental events are conceived to be not merely *correlates* of brain activity, but also *causes*. The causal relation involves the universal power of the whole over its parts, in this case the dynamic enveloping power and properties of conscious high-order brain processes over their constituent neurophysiological and chemical elements. As dynamic emergent properties of cerebral excitation, conscious mental phenomena are given a working role in brain function and a pragmatic reason for being and for having been evolved. I was unable to find anything quite like this interpretation expressed previously from either the mentalist or materialist side, and it seemed to offer a compromise and resolution for the two divergent approaches to the mind-brain problem.

Back in 1965, when this "mind-over-matter" model was first ventured, one had to search a long way in philosophy and especially in science to find anyone who would put into writing the view that mental forces or events are capable of causing physical changes in an organism's behavior or its neurophysiology. With rare exceptions, writings in behavioral science dealing with perception, imagery, emotion, cognition, and other mental phenomena were very cautiously phrased to conform with prevailing materialist-behaviorist doctrine. Care was taken to be sure that the subjective phenomena should not be implied to be more than passive correlates or inner aspects of brain events, and especially to avoid any implication that the mental phenomena themselves might interact causally with the physical brain process. And fifteen years earlier I would not have dreamed that I would ever accept such a concept myself.

Those few in philosophy who had earlier subscribed to psychophysical interaction had been such extreme dualists that little heed had been paid to them in behavioral science. However, once we were able to show that mental events as emergent properties could causally influence neural events in a compromise formulation without violating the principles of scientific explanation, the long-standing resistance to psychophysical interactionism began to decline. For example, the frequency of use of such terms as "mental imagery" or "visual imagery" as explanatory constructs has, after more than five decades of careful avoidance, literally exploded in the recent scientific literature dealing with perception, cognition, and other higher functions (for a critique of this trend see Pylyshyn, 1973). During this same period related philosophical positions have undergone pertinent but subtle rephrasing to encompass these changes, to a degree where it now becomes important in many instances to distinguish between "pre-1965" and "post-1965" versions of a given philosophic stance.

Among other things, the acceptance of inner mental experience as having a significant causal role in cerebral function has produced a changed picture of scientific determinism as applied to human behavior and social action. The phenomena of subjective experience, including feelings and values of all kinds, must now be recognized as positive causal factors in the brain's decision-making process. The freedom thereby introduced into the causal brain sequence leading to a volitional choice far surpasses, in both degree and kind, the notions envisaged in the more mechanistic and atomistic forms of determinism that have excluded mental events. The present interpretation may be seen to set the human brain apart in respect to free will, placing it at an apex above all other known systems in the deterministic universe of science. Our new approach in mind-brain theory thus goes far to restore to human nature some of the personal dignity, freedom, inner creativity, and responsibility of which it has long been deprived by behaviorism and by materialist science generally.

The issues involved here are basic and central to human value questions at all levels. Value problems tend today to take priority over the problem of consciousness or, indeed, any of the theoretical problems of pure science. What good will it do, one may ask, for mankind finally to crack the mind-brain problem if the whole human species is about to be blown off the globe, or starved or crowded off, or polluted out of any reasonable quality of existence? Even staunch advocates of pure science agree that the most important thing many of us can do for pure science these days is not so much to *practice* it as to try to *preserve* it along with civilization and humanity through the coming decades—or generations if one's optimism allows. It is simply a matter of first things first. The same reasoning applies, of course, to many of the aims and objectives that have held priority in biomedical and other fields. The discovery of a cure for cancer, schizophrenia, or cardiovascular disease, for example, would have a relatively minor impact on human existence, generally, as compared to the effect of a slight shift in social values affecting world policy on abortion, birth control, species rights, conservation, and the like.

When it comes to the practicalities of the world problems that are now reshuffling priorities in science, we find, somewhat paradoxically, that the "ivory tower" problem of consciousness continues to carry a top practical rating (Sperry, 1972). World conditions and the future in general will be determined very largely by concepts and beliefs regarding the properties of conscious mind and the kinds of life goals and values that derive from these. Take, for example, the question of whether conscious mind is mortal or immortal, or reincarnate, or cosmic, or whether it is brain-bound, or universal as in panpsychism, or perhaps "supracoalescent" as suggested by Teilhard de Chardin. Clearly, each of these alternatives suggests a different system of value criteria and social priorities.

The final answer, of course, is not yet in, but advances in the mind-brain sciences during the last few decades have substantially narrowed the latitudes for realistic answers. In modern neuroscience it is no longer a question of whether conscious experience is mortal and tied to the living brain, but rather to which particular parts of the brain and which kinds of neural systems are involved. Current interpretation, strongly favoring a "this-world" concept of mind, dispenses with a large number of the "other-world" value determinants of the past. A unifying view of brain, mind, and man in nature can now be seen that provides a monistic framework for values within which science can operate. Even a science of values can be envisaged that would treat values as objective determinants in decision-making and become a basic core for behavioral and social science.

Once it is agreed that mental experience exerts directive control in brain function it follows that the world of inner experience must be given its due in the scientific description of brain action. This puts neuroscience in a position to encompass, at least in principle, all those higher subjective, humanistic aspects of man's nature that the objective approach of science has traditionally seemed to exclude. For these and related reasons (Sperry, 1972), we can no longer accept the dichotomy that has heretofore kept science and values in separate realms. When subjective values have objective consequences and are viewed as universal determinants in all social decision making, they become part of the content of science. The origins, development, and logical structure of values as powerful causal agents become important scientific concerns. More than this, science on these terms, after exclusion of various metaphysical and mystic alternatives including "other-world" mythologies, becomes man's most important means for determining ultimate value and meaning. Our current scheme (Sperry, 1972) would elevate science into a higher social role as source and arbiter of values and belief systems at the highest level. Science would become the final determinant of what is right and true, the best source and authority available to the human brain for finding ultimate axioms and guideline beliefs to live by, and for reaching an intimate understanding and rapport with the forces that control the universe and created man.

Acknowledgment

The work of the author and his laboratory is supported by Grant No. 03372 of the National Institute of Mental Health and by the F. P. Hixon Fund of the California Institute of Technology.

References

Akelaitis, A.J. (1943): Studies on the corpus callosum. VII: Study of language functions (tactile and visual lexia and graphia) unilaterally following section of corpus callosum. *J. Neuropathol. Exp. Neurol.* 2: 226–262.

Bremer, F., Brihaye, J., and André-Balisaux, G. (1956): Physiologie et pathologie du corps calleux. *Arch. Suisses Neurol. Psychiatr.* 78: 31–79.

Delafresnaye, J. F. (1954): *Brain Mechanisms and Consciousness.* Springfield, Ill.: C.C Thomas.

Herrick, J. C. (1948): *The Brain of the Tiger Salamander.* Chicago: University of Chicago Press.

Köhler, W., and Held, R. (1949): The cortical correlate of pattern vision. *Science* 110: 414–419.

Pylyshyn, Z. W. (1973): What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychol. Bull.* 80: 1–24.

Sperry, R. W. (1945): The problem of central nervous reorganization after nerve regeneration and muscle transposition. *Quart. Rev. Biol.* 20: 311–369.

Sperry, R. W. (1950): Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psychol.* 43: 482–489.

Sperry, R. W. (1951a): Mechanisms of neural maturation. *In: Handbook of Experimental Psychology.* Stevens, S. S., ed. New York: John Wiley, pp. 236–280.

Sperry, R. W. (1951b): Regulative factors in the orderly growth of neural circuits. *Growth Symp.* 10: 63–87.

Sperry, R. W. (1952): Neurology and the mind-brain problem. *Am. Sci.* 40: 291–312.

Sperry, R. W. (1955): On the neural basis of the conditioned response. *Br. J. Anim. Behav.* 3: 41–44.

Sperry, R. W. (1961): Cerebral organization and behavior. *Science* 133: 1749–1757.

Sperry, R. W. (1965): Mind, brain and humanist values. *In: New Views of the Nature of Man.* Platt, J.R., ed. Chicago: University of Chicago Press, pp. 71–92. Reprinted in *Bull. Atom. Sci.* (1966) 22(7): 2–6.

Sperry, R. W. (1968): Mental unity following surgical disconnection of the cerebral hemispheres. *Harvey Lect.* 62: 293–323.

Sperry, R. W. (1970a): Perception in the absence of the neocortical commissures. *Res. Publ. Assoc. Res. Nerv. Ment. Dis.* 48: 123–138.

Sperry, R. W. (1970b): An objective approach to subjective experience: Further explanation of a hypothesis. *Psychol. Rev.* 77: 585–590.

Sperry, R. W. (1972): Science and the problem of values. *Perspect. Biol. Med.* 16: 115–130. Reprinted in *Zygon* 9 (1974): 7–21.

Sperry, R. W. (1974a): Lateral specialization in the surgically separated hemisphere. *In: The Neurosciences: Third Study Program.* Schmitt, F.O., and Worden, F.G., eds. Cambridge, Mass.: The MIT Press, pp. 5–19.

Sperry, R. W. (1974b): Mental phenomena as causal determinants in brain function. *In: Mind and Brain: Philosophic and Scientific Strategies.* Globus, G., Maxwell, G., and Savodnik, I., eds. New York: Plenum (in press).

Sperry, R. W., and Miner, N. (1955): Pattern perception following insertion of mica plates into visual cortex. *J. Comp. Physiol. Psychol.* 48: 463–469.

Stratton, G. M. (1897): Vision without inversion of the retinal image. *Psychol. Rev.* 4: 341–360; 463–481.

Weiss, P. (1936): Selectivity controlling the central-peripheral relations in the nervous system *Biol. Rev.* 11: 494–531.

Weiss, P., and Taylor, A. C. (1944): Further experimental evidence against "Neurotropism" in nerve regeneration. *J. Exp. Zool.* 95: 233–257.