1955 (52)

REPRINT No. 52

From THE BRITISH JOURNAL OF ANIMAL BEHAVIOUR *Vol. III, No. 2*
APRIL, 1955

ON THE NEURAL BASIS OF THE CONDITIONED RESPONSE*

BY R. W. SPERRY

Division of Biology, California Institute of Technology, Pasadena, California

A bell sounds and the conditioned dog lifts its forepaw. Although the physiological explanation of this seemingly simple sequence of events has been sought intensively ever since conditioned responses were first discovered more than a half century ago (Pavlov, 1927), we still lack today a satisfactory picture of the underlying neural mechanism. Even the broadest outlines of the neural events remain obscure.

It was first thought that the repeated association of bell and paw movement opened new pathways between the auditory and response centres of the brain. The repeated firing of the two brain centres in association was presumed to leave some kind of residual effect on the interconnecting nerve pathways making it easier for future impulses to travel these same routes. Although now regarded as being much too simple in its original form, this explanation continues to survive with various modifications and qualifications (Konorski, 1948). In particular it is now agreed that the new cerebral associations are more than the direct transcortical linkages earlier supposed, and must involve devious and highly complex pathways.

The physical nature of the residual effects or lasting traces left in the brain remains a matter of pure speculation. Among the wide range of possibilities that continue to receive consideration are the following: growth of new fibre connections, i.e. of axonal and dendritic collaterals and new synaptic endfeet, alteration of transmission resistance at synapses as by expansion and contraction of endfeet or by changes in synaptic membrane permeability, adjustments of excitation threshold within nerve cells or in local portions of their dendritic expansions, specific chemical sensitization of neurons to qualitatively distinct modes of excitation, chemical changes in the neurofibrils and in the neuronal proteins, selective sensitization of neurons to specific spatial patterns of activation, and even structural and chemical changes within the glia.

As an alternative to the notion of the brain traces or engrams as static changes in brain tissue, it has been suggested that the traces might take the form of perseverating eddies of active discharge maintained in self-reexciting circuits (Hilgard & Marquis, 1940; Young, 1938). However, the fact that conditioned responses ordinarily survive deep anesthesia, electro-convulsive shock, epileptic seizures and other disruptions of functional continuity has seemed to rule out any explanation expressed entirely in these dynamic terms.

There remain, nevertheless, certain explaatory advantages in the dynamic form of brain trace, and in a recent advance in learning theory a dual form hypothesis has been proposed (Hebb, 1949), that incorporates the benefits of both the older static and the newer dynamic type of memory trace. In this dual scheme, self-maintained reverberatory excitation is held responsible for the rapid short-term effects in learning and conditioning. In time the continued reverberatory activity is presumed to lead to permanent micro changes in brain morphology that make possible long-term retention of conditioned responses and their survival through states of deep coma, electro-convulsive shock, and the like.

In general the physiological theories have thus far tended to neglect a factor that long has been recognized in psychological studies to be an important element in conditioning namely, the so-called "anticipatory set" or "expectancy."† The incorporation of this factor into the physiological picture leads directly to significant revisions in our thinking about the underlying neural mechanism and to a modified hypothesis which, though somewhat increased in complexity, appears nevertheless to fit better the whole complex of the conditioned response including some of its more paradoxical features.

Following this approach we are lead to postulate as an important aspect of the conditioning process the intermediation of complex high level patterns of central nervous facilitation, i.e. the neural counterparts and derivatives of the psychological expectancies

* From a paper presented in a Symposium on the Physiology of Learning, September 2nd, 1953, at the XIXth International Physiological Congress, Montreal, Canada.

† For a recent discussion of the expectancy factor in conditioning see L. E. Cole (1953) *Human Behavior*, pp. 257-304.

and anticipatory sets. The pattern of central excitation aroused in the brain by the conditioned stimulus can then be assumed to be governed by an intermediate pattern of transient facilitation, rather than being channelled directly by neural engrams or traces as traditionally conceived. Unlike the dynamic traces of earlier theory, the central facilitory pattern or "set" is not a simple vestige or trace of cerebral excitation involved in an earlier pairing of conditioned stimulus and response. It represents rather, a novel and relatively independent organization arising out of high level cerebral activity such as insight, expectancy and the like. In this hypothesis, the permanent structural changes for long-term retention do not help directly to form new associations between sensory and motor centres. Instead they reinforce the intermediary expectancy or facilitory set.

The concept of the central facilitory set and its role in conditioning is best illustrated perhaps with reference to a simple voluntary reaction in man. Let us suppose that electrodes have been taped to the hand of a human subject and that the person has been instructed to lift the hand promptly when a bell sounds, on the penalty of receiving otherwise a severe electric shock. At the sound of the bell, his hand jerks upward instantly, just as quickly as does the paw of the conditioned dog. In the human subject this withdrawal response at the sound of the bell involves a novel pattern of central excitation, which is not the product of any long training, or repeated pairing of bell and shock stimuli. There is no grooving of the fibre pathways between receptor and effector centres of the cortex. In fact, the particular train of central excitation set off by the bell may never have occurred previously in the person's lifetime.

Under these circumstances we say that the human subject gets himself physiologically set to make the proper response at the sound of the bell. The brain pathways for arm flexion become temporarily opened while those for innumerable other possible reactions are temporarily closed. In other words, the prevailing distribution of excitation and inhibition in the brain circuits opens the way for the bell stimulus to release instantly that specific response for which the circuits are adjusted. Once the circuits are properly set, extraneous sensory excitation tends to be absorbed without effect and incompatible reactions are excluded. The novel motor effect of the bell stimulus derives in this case, not from any new pattern of structural pathways but rather, from a new pattern of central excitation and facilitation.

We note that the human subject could just as easily set himself to make the reverse response (i.e. to depress the hand instead of lifting it in which case the excitatory sequence initiated by the bell discharges with equal rapidity into the antagonist muscles. The transient brain set could be patterned likewise for a reaction with the foot, or with the opposite hand. In each case, the specific and immediate motor effect of the incoming auditory impulse is determined by the particular facilitory set that dominates or prevails in the brain at the time the sensory excitations enter.

These governing facilitory sets may be generalized or specific in their organization on both sensory and motor side. For example, a person could set himself to respond only to a bell of low pitch and not to one of high pitch. Or, his set might be quite nonspecific such that any loud noise would set off the reaction; even a sudden tactile stimulus might suffice. Similarly on the motor side, a single response may be rigidly predetermined, or a choice of two responses made possible, or any of a given category.

The main point to be emphasized is that a purely transient, dynamic setting of the brain circuits, in terms of active facilitation and inhibition, is quite capable by itself of causing the bell stimulus to set off a particular limb movement. There is no need here for any new fibre pathways, chemical traces in the neuroplasm, or new synaptic associations. It is evident, furthermore, that *the same response, or other responses, could be coupled functionally in a similar manner to any neutral stimulus (of the type employed in conditioning) simply by an appropriate adjustment in the cerebral facilitory set.* (In referring to these transient adjustments of the brain mechanism as *facilitory* sets, it should be recognized that the central *inhibitory* phase of the process may be fully as important as the facilitory.)

There is much to suggest that the coupling of bell and forepaw response in the dog during conditioning is achieved through the development of a comparable cerebral facilitory set. The bell-shock expectancy that it takes only a moment to convey to the brain of man with the aid of language, may require hours or days to develop in the brain of the dog by the more roundabout method of example and experience.

Eventually as a result of training, the dog comes to expect the shock at the sound of the bell. More than this the dog comes to anticipate both the bell and shock so that by the time the bell sounds, the dog already is actively prepared in advance (like the human subject) to make the correct forelimb response.

We may thus suppose that in the brain of the conditioned animal a well-organized facilitory set appears prior to the incidence of the conditioned stimulus. It is this special pattern of central facilitation—absent before training and developed through training—that is directly responsible for channelling the sensory impulses into the proper motor response.

The activity of these cerebral facilitory sets is not confined to conditioning and voluntary reaction. They operate continually in behaviour and constitute a prime factor in the control of all brain function. Thinking, perceiving, recognizing, imagining, reasoning, reflex activity, learning, and remembering, as well as conditioning and voluntary activity, have all been found in the laboratory to be profoundly affected by the so-called "mental set" (Gibson, 1951). It is by means of differential facilitory sets that the brain is able to function as many machines in one, setting and resetting itself dozens of times in the course of a day, now for one type of operation, now for another. In short, a great deal of the plasticity in vertebrate behaviour, including that of conditioning, is made possible, not through structural remodelling of the fibre pathways but through dynamic readjustments in the background of central facilitation.

To account for the long-term retention of conditioned responses, one must infer further that the facilitory sets and associated dynamic readjustments become reinforced by some permanent type of neural trace, or engram. In considering the problem of the patterning of the brain engrams, it is all-important that one does not lose sight of the primary role played by the above dynamic factors. Even after a long rest interval, or electrochock treatment, the conditioned response is still dependent upon a preliminary rearousal of the anticipatory set. If the proper pattern of central facilitation has not been established by the time the bell sounds, the conditioned forepaw response fails to occur (Cole, 1953).

A number of implications regarding the nature of the static engram are incorporated in this approach to conditioning, some of which may be stated briefly as follows: (*a*) The engrams for a conditioned response do not take the form merely of new associations between the receptor and effector centres of the brain, however complex. A different type of engram pattern is implied, namely, a pattern so designed as to facilitate, at the proper time and place, the establishment of the correct expectancy and facilitory set. (*b*) With much of the burden of the detailed patterning of the conditioned response relegated to the dynamics of cerebral facilitation, the engrams themselves need be much less extensive and complete than is otherwise necessary. Relatively small changes at critical points may be sufficient to reactivate a given anticipatory set in the proper dynamic background. (*c*) The static engrams on the one hand and the dynamic facilitory patterns on the other are regarded as co-functions, i.e. the operation of the two factors in long-term retention becomes mutually dependent with the conditioned response being a product of the combined action. (*d*) Since the engrams reinforce higher level cerebral activity, they are believed to be complex and diffused rather than localized in nature even in the very simplest conditioned responses. (*e*) The engram patterning is further complicated by the fact that the engrams must function in large part to support the transitions between the expectancies and sets as well as the facilitory patterns per se.

There are definite advantages from the engineering standpoint in having the permanent alterations of the brain designed to reinforce the higher level intermediary facilitation, instead of having them affect the sensory-motor associations directly. This makes it possible, to mention just one advantage, for different (even antagonistic) motor responses to be linked simultaneously to the same sensory stimulus, something that may be desirable to suit different circumstances. This scheme also provides much more readily for that class of phenomena included under the terms "sensory and motor equivalence." In general, with this kind of arrangement, the sensory and response mechanisms of the brain do not become tied down nor selectively modified for specific habits, but remain free to be used in various ways and in various combinations in all categories of behaviour.

This description of the neural events also accounts more readily than earlier theories for findings like those of Lashley (1950) and others on the effects of brain extirpation. Consider the difficulty, on this basis of trying to ablate even so simple a habit as the conditioned paw movement mentioned above for illustration. One might eliminate the conditioned response by extirpating enough of the auditory centres of the brain to prevent the stimulus from reaching the anticipatory set. However, if the facilitory set is strongly organized, very little stimulus is needed to trigger off the response. In fact, it is known that the response may occur spontaneously under certain conditions without any outside stimulus. Any small remnant of the primary sensory area that would permit the impulses from the bell to filter through to the central facilitory set would be enough to trigger off the conditioned response.

Similarly on the motor side, any undamaged motor mechanisms by which the animal can achieve the desired effect will be sufficient. The ability to switch quickly from one motor pattern to another to accomplish an end (motor equivalence) is already an intrinsic property of the brain. Once the dog learns that contact with the shock electrode must be broken immediately after the bell sounds, any remaining means available is automatically called into play to accomplish this end. Consequently extensive and widely paralyzing lesions in the motor cortex are required to prevent an animal from making a conditioned response in one way or another. For some learning situations any residual ability to move at all may be enough to carry out the learned performance.

To effectively destroy by surgery the engrams themselves or the expectancies, insights, and sets which they reinforce is equally difficult. According to the foregoing, the constellation of engrams for even the simplest conditioned response tends to be complex in design and diffused through the brain, and thus highly resistant to local ablation. Further, the intermediary central facilitation is reinforced from all sides by a great mass of stimuli, associated not only with the conditioning chamber, but also with the situations leading to it, with the experimenter's person, even with the home cage and with various related events in the diurnal cycle. To eliminate by brain damage the channels for all these reminders of the conditioning experience would be prohibitive.

REFERENCES

Cole, L. E. (1953). *Human behavior.* New York: World Book Co.

Gibson, J. J. (1941). A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, **38**, 781-817.

Hebb, D. O. (1949). *The organisation of behaviour.* New York: Wiley.

Hilgard, E. R., & Marquis, D. G. (1940). *Conditioning and Learning.* New York: Appleton-Century.

Konorski, J. (1948). *Conditioned reflexes and neuron organization.* Cambridge University Press.

Lashley, K. S. (1950). In search of the engram. *Symp. Soc. exp. Biol.*, **4**, 454-482.

Pavlov, I P. (1927). *Conditioned reflexes.* Oxford University Press.

Young, J. Z. (1938). *In de Beer: Evolution.* Oxford University Press, 179-205.