The function of dream sleep
Francis Crick* & Graeme Mitchison*

We propose that the function of dream sleep (more properly rapid-eye movement or REM sleep) is to remove certain undesirable modes of interaction in networks of cells in the cerebral cortex. We postulate that this is done in REM sleep by a reverse learning mechanism (see also p. 158), so that the trace in the brain of the unconscious dream is weakened, rather than strengthened, by the dream.

MANKIND has always been fascinated by dreams. As might be expected, there have been many attempts to assign a purpose or significance to them. Although we dream for one or two hours every night, we do not remember most of our dreams. Earlier thinkers, such as Freud, did not know this. Modern theories (not reviewed here in detail) have usually proposed that sleep and dreams save energy or have various restorative functions, either to replenish the brain biochemically in some way, or to reclassify or reorder the information stored in it.

Sleep is of several kinds. Dream sleep, or rapid eye movement (REM) sleep, is predominantly found in viviparous mammals and birds. It seems to be associated with homeothermy (a constant internal temperature) and the possession of an appreciable neocortex or its equivalent. It is not unimportant because of the appreciable amount of time we spend in this peculiar state.

We propose here a new explanation for the function of REM sleep. The basis of our theory is the assumption that in viviparous mammals the cortical system (the cerebral cortex and some of its associated subcortical structures) can be regarded as a network of interconnected cells which can support a great variety of modes of mutual excitation. Such a system is likely to be subject to unwanted or 'parasitic' modes of behaviour, which arise as it is disturbed either by the growth of the brain or by the modifications produced by experience. We propose that such modes are detected and suppressed by a special mechanism which operates during REM sleep and has the character of an active process which is, loosely speaking, the opposite of learning. We call this 'reverse learning' or 'unlearning'. This mechanism, which is not the same as normal forgetting,

...
produce an output which combines many of the common features of its normal outputs.

A major difficulty with all nets of this general type is that they become overloading if an attempt is made to store simultaneously too many different patterns or associations of patterns, or if the stored patterns have too large an overlap. This is because of the superimposed nature of the storage. How the net will behave when overloaded depends on the exact structure of the net, but certain patterns of behaviour are likely to emerge: (1) The net may produce many far-fetched or bizarre associations ("fantasy"). (2) The net may tend to produce the same state, or one of a small set of states, whatever the input ("obsession"). (3) Certain kinds of nets, particularly those which feed back on themselves, may respond to inappropriate input signals which would normally evoke no response from the net ("hallucination").

It is against this background of rather tentative and idealized theory that our proposal must be judged.

If the cortex were hard-wired during embryogenesis to an exactly predetermined pattern of synaptic connections, the burden of eliminating parasitic modes in cortical nets would have to be undertaken by the genes alone. Although there is considerable evidence for specificity in the cortical wiring, it is likely that many of the details of the synaptic connections — their exact locations and their strength — are made in a semirandom manner and refined by experience. This is almost a necessity in an organism which is capable of learning very large amounts of novel information. Thus it seems likely that both during cortical growth (when we may say that certain broadly predetermined "associations" are laid down), and also in facing the experiences of adult life, such parasitic modes will be unavoidably generated.

How would one attempt to eliminate these modes? We suggest the following. The major inputs and outputs of the system should be turned off, so that the system is largely isolated. It should then be given successive "random" activations, from internal sources, so that any incipient parasitic modes would be excited, especially if the general balance of excitation to inhibition had been temporarily tilted towards excitation. Some mechanism is then needed to make changes so that these potentially parasitic modes are damped down. The cortex as a whole, in its rough outline description immediately after the forebrain is periodically and widely stimulated by the brain stem. This activity promotes rather nonspecific signals from the brain stem which are likely to be associated with the unconscious equivalent of dreaming, which only reaches normal consciousness if the sleeper awakes.

The mechanism we propose is based on the reverse learning mechanism which will modify the cortex (for example, by altering the strengths of individual synapses) in such a way that this particular activity is less likely in the future. For example, if a synapse needs to be strengthened in order to remember something, then in reverse learning it would be weakened. Put more loosely, we suggest that in REM sleep we unlearn our unconscious dreams. "We dream in order to forget." After this paper had been initially submitted for publication, we learnt from Dr. John Hopfield that he and his colleagues had independently arrived at the idea of reverse learning, though not in connection with dreams. In a parallel communication they have shown that the behaviour of their very idealized neural net is indeed improved by reverse learning. That is, it equalizes the accessibility of stored memories and suppresses most of the conscious ones. We have since repeated their simulations and confirmed their general conclusions. It remains to be seen how well reverse learning acts on other more realistic neural nets. We have revised our paper in the light of their results.

Note that the amount of reverse learning per step in these simulations was very small (only about 1% of the amount needed for complete learning), although several hundred such steps were used. This alerts us to the possibility that the changes produced in REM sleep may individually be very small but cumulative over many PGO spikes and REM periods may of the muscles of the sleeping animal, especially its head and neck muscles, are more relaxed than in non-REM sleep. Its cortex, as judged by the electroencephalogram (EEG) and by the rapid movement of the eyes beneath closed lids, appears to be very active and in a state similar to the waking state. On the other hand, the monoamine neurones in the brain stem, especially those in the locus coeruleus, raphe and peribranchial nuclei, reduce their firing rates in REM sleep to only a few per cent of the corresponding rate in the waking state.

Another major difference between REM and non-REM sleep lies in the dreams associated with them. For most people the few dreams found in non-REM sleep tend to have a rather thought-like character. During REM sleep, on the other hand, dreams occur more frequently and usually have a perceptual vividness and the illogical episodic character with which we are all familiar. A healthy adult usually spends a total of about 4 hours each night in REM sleep, spread over several periods. The evidence suggests that most of the dreams during these REM periods do not reach normal consciousness, dreams being remembered only if the sleeper awakes while dreaming. Even then the memory of a dream is usually very transient, fading quickly if no effort is made to remember it by rehearsing its content.

A most remarkable finding is that newborn humans may have as much as 8 hours of REM sleep per day. There is also evidence to suggest that in the womb, especially in the third trimester, REM sleep occurs even more frequently. This large amount of REM sleep before and after birth is also found in other mammals. All viviparous mammals examined, including primitive marsupials such as the oppossum, show periods of REM sleep. Even an animal like the mole, which can hardly move its eyes, shows the characteristic EEG of REM sleep. Birds have REM sleep, although often only a very small amount of it, occupying perhaps 5% of their sleep. There are no very convincing reports of REM sleep (as judged by the EEG) in reptiles, amphibians or fish.

If an animal is deprived of REM sleep for one or more nights (but allowed non-REM sleep) then it will usually have more REM sleep in subsequent nights.

All this evidence suggests that REM sleep is associated with the unconscious dreaming process — that is, with REM sleep without awakening — rather than with the few dreams which are recalled. It has been shown that during REM sleep the forebrain is periodically and widely stimulated by the brain stem. This activity in the brain stem can happen even in the absence of the cortex. Hobson and McCarley, following the pioneer work of Jouvet, have postulated a 'dream state generator' which lies mainly in the pontine reticular formation (the question of which exact cell groups are involved is controversial). It produces the so-called PGO waves. These waves are likely to be associated with the unconscious equivalent of dreaming, which only reaches normal consciousness if the sleeper awakes.

The postulated mechanism

We need a mechanism which will tune the cortical system, in the sense of removing parasitic modes which arise after the system has been disturbed either by growth of the brain (when new connections are constantly being made) or by the modifications produced by experience. The mechanism we propose is based on the more or less random stimulation of the forebrain by the brain stem that will tend to excite the inappropriate modes of brain activity referred to earlier, and especially those which are too prone to be set off by random noise rather than by highly structured specific signals. We further postulate a reverse learning mechanism which will modify the cortex (for example, by altering the strengths of individual synapses) in such a way that this particular activity is less likely in the future. For example, if a synapse needs to be strengthened in order to remember something, then in reverse learning it would be weakened. Put more loosely, we suggest that in REM sleep we unlearn our unconscious dreams. "We dream in order to forget."

After this paper had been initially submitted for publication, we learnt from Dr. John Hopfield that he and his colleagues had independently arrived at the idea of reverse learning, though not in connection with dreams. In a parallel communication they have shown that the behaviour of their very idealized neural net is indeed improved by reverse learning. That is, it equalizes the accessibility of stored memories and suppresses most of the conscious ones. We have since repeated their simulations and confirmed their general conclusions. It remains to be seen how well reverse learning acts on other more realistic neural nets. We have revised our paper in the light of their results.

Note that the amount of reverse learning per step in these simulations was very small (only about 1% of the amount needed for complete learning), although several hundred such steps were used. This alerts us to the possibility that the changes produced in REM sleep may individually be very small but cumulative over many PGO spikes and
The objection might be raised that some experiments have shown that REM may appear to help the retention of memory, whereas the reverse process of reverse learning would tend to make the memory fade. The results of Hopfield et al.2 show that this need not be the case. After reverse learning the recall of their net was less confused and more uniform.

If there is indeed a mechanism for reverse learning, many questions arise about its character. Does it act via the same mechanism as normal learning (whatever that is) or is a special, quite separate, mechanism involved? Is the mechanism associated with one particular system in the brain stem? Another possibility is that a small amount of reverse learning is always present but is normally overwhelmed by the positive plasticity produced by one or more of the diffuse systems from the brain stem. When their activity is greatly reduced, as it is in REM sleep, the residual reverse learning can then exert its effect unopposed, at least on recently modified synapses.

In its simplest form our theory postulates that there is no intelligent supervisor inside the brain which decides in detail which potential neural activities should be left untouched and which should be dampened down. This choice is made solely by the response of the forebrain to the relatively nonspecific signals from the brain stem. In very general terms, the brain stem gives the forebrain a varied pattern of bangs (the PGO waves). Any resulting activity is then modified so that it is less likely to occur in the future.

It would of course be possible to postulate a more complex mechanism. For example, in REM sleep, especially in early development, there could be innate testing programmes, together with a "supervisor" to decide what to store and what to erase, depending on the result of the tests. Various workers have made proposals along these lines8-24. As far as we know, nobody has previously suggested that the testing procedure involves the removal of potentially parasitic modes.

It has been customary to believe that during an unconscious dream the content of the dream is stored in some form of very short-term memory, that the mechanism for transferring it to longer-term memory is inoperative. We normally become conscious of our dreams only if we wake up while dreaming is in progress. If we then pay attention to our dream, some of its content can be maintained in very short-term memory and may eventually be transferred to longer-term memory as the transfer mechanism becomes activated. Otherwise our dream fades. Thus we can speak of forgetting our dreams, meaning that we know that we had a dream, but are somewhat uncertain of its content.

This forgetting of a dream, which has often been remarked on, does not necessarily involve our postulated reverse learning process. The latter is a passive mechanism which does not merely fail to alter synaptic strengths (or other long-lasting brain parameters) but changes them so that the dream is not just forgotten but actively 'unlearnt'. The result is that the dream (or some of the elements of it) is less likely to recur in the future.

The terms 'reverse learning' or 'unlearning' are not ideal because they rather imply that one has to learn something first in order to unlearn it. What does a fetus 'learn' that has to be unlearned? Our answer is that, during development, the semirandom process of making synaptic connections is likely to produce parasitic modes. It is these which must be 'unlearned' in order to obtain a well-behaved system.

We need some explanation for recurrent dreams. We propose the ad hoc hypothesis that a recurrent dream is one which, for one reason or another, tends to wake up the sleeper, perhaps because of the anxiety often associated with them. This will have the effect that the learning process changes sign, passing from reverse learning to positive learning, so that the underlying spurious associations remain, and so a similar dream is likely to occur on some later occasion. This mechanism does postulate a supervisor of a kind but its sole function is to decide whether the sleeper should wake up or not. Thus for a dream to become recurrent it must have two properties. It must be related to a potentially parasitic mode and it must wake up the dreamer in such a way that he remembers it rather vividly.

Our theory, in its present state, says nothing about the function of non-REM sleep. These stages of sleep usually have less of the hallucinoid type of dream which we associate with our reverse learning mechanism, and they do not involve a significant learning mechanism. Non-REM sleep is likely to have the restorative function often postulated for it but it may also have some informational function. For example, it might be used for the process of 'consolidating' memory in some way. It is worth noting that the first REM period of the night is normally preceded by a substantial period of non-REM sleep.

Testing the theory

As far as we can tell, our theory is broadly compatible with a large amount of experimental data. Starting from a plausible hypothesis about cortical function, it explains in an effortless way both the need for REM sleep in adult life and the large amount of it during the development of the brain. We believe no previous theory explains this distribution of REM sleep in such a simple manner. Any purely psychological theory (such as Freud's) is hard-pressed to explain the large amount of REM sleep in the womb, and any purely developmental theory must account for the quite appreciable amount of REM sleep in adult life. Our theory accounts for both. It is also compatible with the hallucinoid nature of REM dreams.

The effects of REM sleep deprivation are harder to explain. It is well established that REM deprivation often produces a recurrent dream and that REM sleep is less of the hallucinoid type which usually occurs when the subject is eventually allowed to sleep without interruption. We would have expected that REM deprivation, if severe enough, might cause hallucinations — that is, structured visual and auditory responses to 'noise' — and perhaps delusions and obsessions. There is a little evidence for this25, but usually the effects are either small or absent26. This is partly because it is extremely difficult to produce long periods of complete REM deprivation in humans by selective arousal. After a week or two it becomes almost impossible to awaken them promptly at every onset of REM sleep, so that prolonged experiments have not been done. One cannot help but wonder whether similar experiments on food deprivation might lead to the conclusion (if unsupported by other evidence) that food also had no essential function. However, REM deprivation in animals does appear to lower the threshold for cortical instability produced by electroconvulsive shock27-30, which is what we might expect. REM deprivation in humans sometimes produces irritability and an inability to concentrate. One might suggest that these are the effects of the attention mechanism being forced to subdue subthreshold parasitic modes which would otherwise break into consciousness. REM deprivation can also allow feelings and wishes to appear which had previously been kept out of consciousness21, or, in certain subjects, can show changes towards increased internal fantasy during waking22.

A further difficulty is that some drugs, such as certain muscle relaxant inhibitors, appear to prevent REM sleep entirely23 without producing very obvious psychological deficits. This is a difficulty for any theory which assumes REM sleep is important and runs in the face of all the other evidence about it. We can only plead that such drugs may have complicated side effects which make the observations misleading.

A direct test of our postulated reverse learning mechanism seems extremely difficult. It would be necessary to show that our unconscious dreams (dreams we do not remember — a new word for this is really needed, we suggest 'remination') reduce the probability of such thoughts occurring in the future. This is far beyond the methods we have available today. It would be interesting to know if the threshold for hallucination, induced by drugs or other means, is lowered as a result of REM deprivation. Another approach would be to look for the structural and chemical correlates of the postulated reverse learning mechanism, but exactly how to do this is at the moment unclear. Without further evidence of this kind our theory must be regarded as speculative.

It is clear that useful insights can come
from neural modelling. This approach has its limitations, since it is difficult to produce realistic models and even more difficult to simulate them effectively, especially if the hypothetical neural nets approach a realistic size, when the computational time becomes prohibitively long. However, such theoretical studies should at least reveal some of the types of networks which would benefit from our proposed mechanism. They might also help to give more life to our otherwise rather vague characterization of the cortical system.

Another approach would be to undertake comparative studies. There is one mammal which, although possessing a well-developed neocortex, appears not to show any signs of REM sleep (at least in young adults), even though it exhibits normal non-REM sleep. This is the Echidna Tachyglossus aculeatus (the spiny anteater) found in Australia. The Echidnas and the duck-billed platypus are primitive egg-laying mammals (monotremes).

Griffiths has written that "...the gyrencephalic brains of the Tachyglossidae have been and are a source of wonder to neurobiologists". He quotes Elliott Smith who in 1902 wrote "The most obtrusive feature of this brain is the relatively enormous development of the cerebral hemispheres... The meaning of this large neopallium is quite incomprehensible...". Griffiths adds: "Determinants of modern neurophysiology also fail to explain how echidnas come by this cortex!"

We suggest that Tachyglossus needs such a large cortex because it cannot tune it up by the process of reverse learning. Experience with idealized neural nets shows that one can usually avoid overloading a net, and thus the confusion such an overloading creates, by making the net bigger.

Tachyglossus can be studied in captivity. It might be rewarding to examine in more detail its behaviour, neurophysiology and neuroanatomy compared to a primitive placental mammal, such as a hedgehog, which does show REM sleep. If REM sleep serves an important function, this should be reflected in some way in its absence in the spiny anteater.

Possible implications

If it turns out that our ideas are broadly correct, they could help us to understand the evolution of the neocortex which is so typical of mammals. It seems likely that in order for a highly tuned system to perform efficiently at least two requirements are necessary: a fairly constant internal temperature, so that its function is not disturbed by temperature fluctuations, and in addition a cleaning-up mechanism, to remove potentially parasitic modes. In short, without REM dreams evolution could not have produced the highly refined neocortex we have today.

If the reverse learning mechanism we have postulated exists, one might wonder what effects its failure might have. A complete failure might lead to such grave disturbances—a state of almost perpetual obsession or spurious, hallucinatory associations—that it would probably be severely selected against. A partial failure should produce unwanted responses to random noise, perhaps as hallucinations, delusions, and obsessions, and produce a state not unlike some schizophrenias. It has been postulated before that there might be a relation between REM sleep and schizophrenia, but studies have shown that there is little or no connection between the outward signs of REM sleep and schizophrenia. However, a partial failure of the reverse learning mechanism would not necessarily alter the amount of REM sleep, since the control mechanisms for the occurrence of REM sleep might be somewhat distinct from the reverse learning process itself. Thus the possibility that some forms of schizophrenia might be caused by a defect in the reverse learning process should not be overlooked.

In this model, attempting to remember one’s dreams should perhaps not be encouraged, because such remembering may help to retain patterns of thought which are better forgotten. These are the very patterns which the organism was attempting to damp down.

Finally we should remark that even if it turns out that our ideas are wrong and that nature does not employ the reverse learning mechanism we have postulated, the process may well be useful for artificial intelligence machines of the future, especially those having extensive parallel processing, a learning mechanism and a certain amount of randomness in their construction.

We thank our resident and visiting colleagues at The Salk Institute for many useful discussions. We are especially grateful to Drs Allan Hobson and Jim Horne who made detailed comments on our draft manuscript, and to Dr John Hopfield for communicating his work to us before publication and for helpful discussions. This work has been supported by the J. W. Kieckhefer Foundation, Samuel Roberts Noble Foundation, US Air Force Grant number AFOSR-82-0042 and the System Development Foundation.