Brit. J. Psychiat. (1963), **109**, 12-18

Behaviour Therapy, Extinction and Relapse in Neurosis

By H. J. EYSENCK

Behaviour therapy is defined as the application of the principles of modern learning theory to the treatment of neurotic disorders (Eysenck, 1960a, 1963). It defines neurotic "symptoms" as unadaptive conditioned autonomic responses, or the skeletal and muscular activities instrumental in moderating these conditioned autonomic responses. Treatment consists essentially in the extinction of autonomic, skeletal and muscular responses of this type. This extinction may be produced in a great variety of ways, but experience has shown that the most useful and important is probably the method of counterconditioning or "reciprocal inhibition" (Wolpe, 1958). This method takes two forms, according to the nature of the symptom. (1) When the symptom is of a dysthymic character (anxieties, phobias, depression, obsessive-compulsive reactions, etc.) it is assumed that the disorder consists of conditioned *sympathetic* reactions, and the treatment consists of reconditioning the stimulus (or stimuli) to produce *parasympathetic* reactions which, being antagonistic to the sympathetic ones, will weaken and finally extinguish them. These disorders we will here call "disorders of the first kind". (2) When the symptom is of a socially disapproved type in which the conditioned stimulus evokes parasympathetic responses (alcoholism, fetishism, homosexuality), or where there is an entire absence of an appropriate conditioned response (enuresis, psychopathic behaviour), treatment (aversion therapy) consists of the pairing of the stimulus in question with strong aversive stimuli producing sympathetic reactions. These disorders we will call "disorders of the second kind". (In putting the distinction between these two types of treatment in this very abbreviated form, we have used the terms "sympathetic" and "parasympathetic" in a rather inexact shorthand notation to refer to hedonically positive and negative experiences respectively; the reader familiar with the complexities of autonomic reactions will no doubt be able to translate these blanket statements into more precise language appropriate to each individual case. We have retained this use of the terms here because it aids in the general description given, and indicates the physiological basis assumed to exist for the hedonic reactions.)

It has always been emphasized that behaviour therapy is *purely symptomatic*; behaviour therapy is based on a theory regarding neurotic behaviour which maintains that "there is no neurosis underlying the symptom, but merely the symptom itself" (Eysenck, 1960a, p. 9). The most frequent objection to purely symptomatic treatment has always been based on an alternative hypothesis, viz. one which regards the symptom merely as the manifestation of some "unconscious" and "repressed" "complex" which is the true base from which symptoms spring; the "complex" is the "illness", and the symptom cannot be cured in any permanent form without "uncovering" the "complex". (Quotation marks are used to indicate those terms and concepts for which the writer has been unable to find any experimental evidence in the literature.) It would follow directly from such an hypothesis that treatment of symptoms only would result in speedy relapse, and one would predict that behaviour therapy, in any of its forms, might produce some initial amelioration, only to be followed by a recrudescence of the same, or emergence of some other symptom. That this is a crucial point has been universally recognized by behaviour therapists and their critics alike; psycho-analysis seldom makes clear-cut and empirically verifiable predictions, and the fact that on this point its many practitioners are unanimous makes it a valuable testing-stone for experimentally-minded psychiatrists and psychologists. The evidence

(Wolpe, 1961; Yates, 1958) suggests strongly that the looked-for recrudescence of symptoms has signally failed to materialize. This is true not only in relation to behaviour therapy; it has also been found in relation to spontaneous remission (which according to psycho-analytic thinking should not occur) that permanence of recovery, rather than relapse, is the rule (Denker, 1946). Nevertheless, the writer does not feel that this problem should be dismissed too easily; the available evidence is suggestive, but by no means conclusive. There undoubtedly are instances in the literature where relapses have been found to take place, and even if the total number should not be very large it may be useful to look at the problem from the theoretical point of view. Protagonists tend to take an all-or-none point of view in these matters; a more objective approach might indicate some less extreme position to be more in accord with the facts, and might also lead to suggestions for improvements in the methods of treatment used.

Let us first consider the fact that behaviour therapy posits the occurrence of classical and instrumental conditioning as crucial elements in the genesis of neurotic disorders of the first kind, and frequently also of the second kind, and that it assumes these two kinds of conditioning to play a central part in the treatment of both kinds of neurosis. Such a statement has many important theoretical consequences, and it seems curious that so little work has been done to investigate theoretically the sets of deductions which can be made from these basic premises. In this paper we shall be concerned in the main with a particular phenomenon universally found in conditioning experiments and entitled "extinction" by Pavlov. This is defined by Kimble (1961) as "the specific procedure of presenting the conditioned stimulus unaccompanied by the usual reinforcement; also the decrement in a conditioned response which results from that procedure". It is our belief that in tracing the fate of those conditioned responses we call neurotic symptoms, too little attention has been paid to the facts of extinction, and the conditions giving rise to them. It is suggested that extinction affects in a profoundly different manner neurotic disorders of the first and second kind respectively, and that the problems of relapse cannot be discussed in any satisfactory manner without paying attention to these differences.

Consider neurotic disorders of the first kind, i.e. the dysthymic disorders. Here it is hypothesized that the original cause of the symptom is a conjunction of a single traumatic event (or several repeated subtraumatic experiences) with the presence of a previously neutral stimulus. Through the process of classical conditioning the previously neutral stimulus (CS) now acquires the properties properly belonging to the traumatic event (UCS), itself, and produces the autonomic disturbances originally produced by the UCS. As I have argued elsewhere (Eysenck, 1962) this account would lead us to expect spontaneous remission to occur in practically all cases of neurosis, due to the fact that on subsequent occasions the presentation of the CS would not be followed by reinforcement, so that in due course extinction should take place. It is, of course, well known that in the great majority of neurotic disorders spontaneous remission does in fact supervene within a relatively short period of time (Eysenck, 1960b), so that the facts are in good accord with our hypothesis.

Those cases where spontaneous remission does *not* occur can be explained in terms of a subsequent *second* stage of instrumental conditioning. If the patient withdraws from the CS upon encountering it there will be a lowering of sympathetic arousal which should act as a reinforcement for the act of avoidance; and this reinforcement should gradually lead to the building up of a new habit of avoiding the CS, thus making extinction impossible (Eysenck, 1962).

In relation to symptoms of the first kind, therefore, we find that the rôle of extinction is a very important one. In a large number of cases it effectively performs the function of therapy without the intervention of any particular therapist, leading to the spontaneous remission (and non-recurrence!) of the symptom. As the events producing extinction may, on a first approximation, be assumed to be randomly distributed over time, one would expect the curve of recovery to be a simple exponential

function of time, and as I have shown this is in fact so (Eysenck, 1960b). It appears that recovery from neurotic disorders roughly follows the course given by the formula:—

$$X = 100(1 - 10^{-0.00435N})$$

where X stands for the amount of improvement and N the number of weeks elapsed. While the exact values in this formula should not, of course, be taken too seriously, its general form is that of the typical extinction curve with which psychologists are familiar from the laboratory. It is clear, then, that in the case of neurosis of the first kind extinction works in favour of the therapist and may even unaided lead to improvement and cure. Where the random events of life, acting in this fashion, do not produce a cure, the therapist can aid the process along the lines laid down by Wolpe (1958) and others. Relapses should not occur in the ordinary way unless a new, repeated traumatic event occurs to produce a new symptom and a new neurotic disorder. This, of course, could not be considered a relapse, just as we would not consider it a relapse if a patient with a broken scapula should years after recovery suffer a Pott's fracture. Cure from one set of symptoms does not confer immunity on the patient.

Now consider the situation in relation to disorders of the second kind. Here the situation is clearly exactly the opposite to that which we have encountered so far. The patient is suffering from a maladaptive habit which is either itself an unconditioned response $\binom{U}{S\ R}$, as in the case of enuresis, or where the conditioned stimulus has become associated with consequences which are immediately pleasurable to the patient, although they may be socially undesirable and highly unpleasant in their long term consequences for the individual himself (fetishism, alcoholism, and the like). Some types of disorder, such as homosexuality, may pertain to either one or the other of these two categories, i.e. homosexual disorders may be entirely due to an accidental conditioning process, or they may be innate response tendencies $\binom{U}{S\ R}$, or they may be a mixture of both. In any case what is true in all these types of disorder is that a strong bond has been created between a previously neutral stimulus and a strong positive reinforcement. Ordinary events of life occurring randomly are not likely to lead to extinction, as they are not likely to associate the conditioned stimulus with lack of reinforcement.

It might be objected that surely punishment in its various forms has been designed specially by society to produce precisely such a dissociative effect and that by imprisonment, beating or torturing homosexuals, fetishists, etc., we are substituting a negative reinforcement for a positive one. That such an objection is not tenable has been shown in practice by the failure of these methods throughout recorded history; no one nowadays assumes that the habits of the homosexual are altered by putting him into prison, even though such punishment may restrain the expression of these habits for a while (in other words punishment may affect $\binom{E}{S\ R}$ but not $\binom{H}{S\ R}$). Even more important, Mowrer has shown, both theoretically and experimentally ". . . that the consequences of a given act determine the future of that act not only in terms of what may be called the quantitative aspects of the consequences but also in terms of their temporal pattern. In other words, if an act has two consequences—the one rewarding and the other punishing—which would be strictly equal if simultaneous, the influence of those consequences upon later performances of that act will vary depending on the *order* in which they occur. If the punishing consequence comes first and the rewarding one later the difference between the inhibiting and the reinforcing effects will be in favour of the inhibition. But if the rewarding consequence comes first and the punishment one later the difference will be in favour of the reinforcement" (Mowrer, 1950).

It is with respect to this temporal sequence that aversion therapy differs from punishment in the ordinary sense of the term. Punishment is a relatively arbitrary and long delayed consequence of action which, according to the principle just considered, should have very little if any influence upon the habit in question.

Aversion therapy attempts to apply the aversive stimulus *immediately* after the conditioned stimulus, and in such a way that it eliminates, or at least precedes, the positive reinforcement resulting from the act. This is often difficult to do, as clearly split-second timing is of the utmost importance; as has been pointed out before, many people who attempt aversion therapy do so without a full appreciation of the complexities of conditioning, and failure easily results from the haphazard manipulation of time relationships.

Consider now a case where aversion therapy has been successful and where the conditioned stimulus has been successfully linked with the aversive stimulus; we will call this link "aversive conditioning". Now clearly aversive conditioning, like all other types of conditioning, is subject to extinction, and we must consider how extinction can arise, and how it would influence the future course of the symptom. The first point to be borne in mind is that aversive conditioning tends to stop when conditioning has only just been achieved, i.e. without any considerable degree of overlearning. As an example, take the treatment of enuresis by means of the bell and blanket method. According to the theoretical analysis of Lovibond (1961), urinating in bed is the conditioned stimulus which becomes linked with the aversive stimulus, the bell, which in turn produces the immediate reflex cessation of urination. Now it is clear that conditioning can only proceed while the patient still produces the conditioned stimulus, i.e. while conditioning is still far from complete. The moment the patient ceases to urinate in bed further conditioning becomes impossible. With modifications the same argument would hold for other types of aversion therapy. Where the conditioned stimulus can be voluntarily applied, as in the case of consumption of alcoholic beverages, consideration of time, expense and the great discomfort produced usually limits the number of conditioning trials to a relatively small proportion of what may be required to produce any considerable degree of overlearning.

After successful aversion therapy, the patient emerges with a central nervous system into which has been built a certain amount of "aversive conditioning" which is subject to what has been called oscillation by Spearman (1927) and Hull (1943); this is usually designated as $\underset{S\ R}{O}$. Oscillation is a feature of all biological systems and produces random variations in the strength of inhibitory and excitatory potential; these oscillations may be quite considerable in relation to the total amount of potential under consideration.

Consider now an individual who has submitted to a course of aversion conditioning, and whose degree of conditioning is just at the point where the original behaviour does not occur in relation to the stimuli which used to set it off before the course of aversive conditioning. Owing to the process of oscillation the effectiveness of aversive conditioning will be much weaker on certain occasions than on others, and if by accident the original stimuli are present at a time when the excitatory potential of the aversive conditioning is low, the individual will be liable to give way to temptation. If he does, then the extinction process phase of the aversive conditioning will have begun, because the conditioned stimulus has been presented without the (negative) reinforcement. It would follow that on subsequent occasions the excitatory potential $\left(\underset{S\ R}{E}\right)$ would already be weaker to begin with, even without the action of oscillation, so that further extinction trials are even more likely to occur. We thus find that in neurotic disorders of the second kind the random events of everyday life, far from leading to spontaneous remission will rather lead to relapse, other things being equal. Thus our prediction, on theoretical grounds, would be that relapse should be rare or even non-existent in disorders of the first kind, but relatively frequent with disorders of the second kind. There are no empirical studies the results of which could be used to support this deduction in any conclusive manner, but it is noteworthy that those who have denied the occurrence of relapse, like Wolpe (1959), have concentrated largely on disorders of the first kind. Writers dealing with disorders of the second kind, like Gwynne Jones (1960), Oswald (1962), Freund

(1960) and others, have drawn attention to the frequency of relapse in patients of this type. It would seem, therefore, that the distinction made is a potentially fruitful one, although the difference in relapse rates may be attributable, in part at least, to other causes as well. Thus the symptoms of disorders of the first kind are usually such as to motivate the patient very strongly to undergo a process of therapy in order to get relief from these symptoms. The symptoms of disorders of the second kind, however, are much less painful to bear as far as the individual is concerned; indeed they may appear quite pleasant and agreeable to him. It is society, through one of its various agencies, which provides the motivation for therapy, and this imposed drive is likely to be much weaker. This is important, because it is well known that the strength of conditioned responses is very much determined by the strength of the drive under which the individual is working. Here we may have, therefore, an additional principle accounting for the high relapse rate predicted for disorders of the second type.

If we are right in assuming that aversion therapy is less effective than it might be because of the reasons given, then we should look to learning theory for suggestions which might lead to more effective methods of conditioning, and to a lessening of the degree of extinction or "relapse". The most obvious suggestion which emerges from laboratory studies appears to be that the process of aversive conditioning should make use of *partial* rather than complete 100 per cent. reinforcement, as is the most usual practice at the moment. The discovery that partial reinforcement, i.e. the random reinforcement and non-reinforcement of conditioned stimuli, leads to a much greater resistance to extinction, was made in 1939 by Humphries, and is sometimes known as "Humphries' paradox". A lengthy discussion of this phenomenon and its theoretical interpretation is given by Kimble (1961). We need not here concern ourselves with the theoretical explanations or with the details of the experimental procedures; it is sufficient to note that the evidence for the phenomenon is surprisingly unanimous and leaves little doubt about its reality. Whilst apparently contrary to common sense, the efficacy of partial reinforcement in delaying extinction is not entirely inexplicable even on the common sense level itself. One hundred per cent. reinforcement during the acquisition phase marks this phase off obviously and completely from the extinction phase, so that not only is the change obvious to the subject, but also there will be a change in a number of the more remote conditioned stimuli, including the drive stimuli, which together set the stage for the occurrence of the conditioned response. In the case of partial reinforcement the two phases are much more difficult to mark off from each other and the pattern of internal stimulation is very much less different as we go from one phase to the other.

However that may be, the crucial test of a theoretical prediction must inevitably lie in the experimental demonstration that the phenomenon predicted does actually occur. Fortunately there is a very carefully controlled experimental study comparing the relapse rates of children treated by means of the bell and blanket method for enuretic disorders by S. H. Lovibond (1961). In half the cases partial reinforcement was used, and with the other half, 100 per cent. reinforcement. He found that a markedly greater tendency to relapse was evident in the latter group as compared with the former. We may, therefore, conclude that this deduction has some empirical support, although it would, of course, be desirable to study it in connection with many other types of disorder of the second kind in addition to the enuretic cases worked on by Lovibond.

It is noteworthy that many investigators in the use of aversion therapy go directly counter to this principle. Thus in the deconditioning treatment of alcoholism patients are often told not to take drinks on their own and outside the treatment situation, as this would interfere with the treatment. Every precaution is taken to ensure 100 per cent. reinforcement—thus also ensuring, if our hypothesis be correct, that relapse should be the more likely to occur! (It is likely, incidentally, that the massing of conditioning trials leads to an increase in inhibition and, therefore, raises the probability of extinction and relapse. Oswald (1962, p. 211) has argued in favour of a method in which alcohol is given even after nausea has begun. This not only goes

counter to the principle of the inefficiency of backward conditioning, but also ensures a quite unusual massing of conditioning trials such as would not be expected to aid in the prevention of extinction.)

Whilst partial reinforcement presents one way of avoiding the difficulties of extinction and relapse, it is of course not the only one. A second method is that of "overlearning", i.e. the continuation of the conditioning process, preferably through the use of widely spaced trials, well beyond the point where conditioning occurs to an apparently satisfactory degree. It is impossible on *a priori* grounds to lay down any general rules as to the precise number of reinforced trials required; this obviously depends on such factors as the subject's drive, his conditionability, the urgency of the situation, facilities present at the time and so on. It might be said in parenthesis, however, that practitioners appear often to have used reinforcements in aversion therapy, the strength of which was much greater than would be required for satisfactory conditioning. The use of strong doses of apomorphine can be defended in that, as is well known, the strength of the unconditioned stimulus in part determines the efficacy of the conditioning process. However, the accrued misery induced puts many people off the treatment altogether, and it also tends to reduce the number of conditioning trials to an absolute minimum. The number of reinforced trials, however, is probably a more important factor in determining the strength of the conditioned response, and more numerous trials with a weaker UCS would probably be more effective than a small number with the strong UCS. The precise effects of varied combinations of conditions must, of course, be determined by experimental investigations; it can only be hoped that these will in the future be more closely geared to existing knowledge than has been the case in the past.

Even with partial reinforcement and overlearning, the dynamics of the situation in the case of disorders of the second type are such that relapse is an ever present danger. It can be guarded against by widely spaced conditioning trials administered throughout the life history of the individual, perhaps once or twice a year; it seems likely that such a programme would prevent relapse quite efficiently, particularly when used in conjunction with the other methods outlined. Again the call must be for experimental investigations, properly designed, executed and evaluated according to the most rigorous criteria of modern statistical and experimental method.

Summary

In this paper a distinction has been drawn on theoretical lines between two types of neurotic disorder, and the consequences have been traced of the extinction process on the symptoms of these two types of disorder. It was shown that while extinctions occurring naturally during the life history of the individual should produce *spontaneous remission* in patients suffering from disorders of the first kind, it should produce a *relapse* after successful treatment in the case of patients suffering from disorders of the second kind. Certain deductions were made from modern learning theory suggesting that the occurrence of extinction and relapse could be lessened by using treatment processes emphasizing partial reinforcement, overlearning, spaced trials and supportive conditioning. It was argued that the distinction between these types of disorder is from the practical point of view a fundamental one, and that generalizations about the recrudescence of symptoms after treatment by behaviour therapy are quite unjustified unless they specify the type of disorder to which they are meant to apply. It is also argued that current methods of aversion therapy tend to violate precisely those principles which would ensure a lessening of the relapse rate, and that widespread experimentation should be encouraged in which these principles would be put to the test of clinical success.

References

Denker, R. (1946). "Results of treatment of psychoneuroses by general practitioners", *New York State J. Med.*, **46**, 2164–2166.

Eysenck, H. J. (1960a). *Behaviour Therapy and Neuroses.* Oxford: Pergamon Press.

——(1960b). *Handbook of Abnormal Psychology.* London: Pitman.

EYSENCK, H. J. (1962). "Behaviour therapy, spontaneous remissions and transference in neurotics", *Amer. J. Psychiat.*, to appear.

——(1963). *Experiments in Behaviour Therapy*. London: Pergamon Press.

FREUND, K. (1960). "Some problems in the treatment of homosexuality". In: Eysenck, H. J. (Ed.), *Behaviour Therapy and the Neuroses*. Oxford: Pergamon Press, 312-326.

HULL, C. L. (1943). *Principles of Behaviour*. New York: Appleton-Century-Crofts.

GWYNNE JONES, H. (1960). "The behavioural treatment of enuresis nocturna". In: Eysenck, H. J. (Ed.), *Behaviour Therapy and the Neuroses*. Oxford: Pergamon Press, 377–403.

KIMBLE, A. A. (1961). In: Hilgard and Marquis' *Conditioning and Learning*. New York: Appleton-Century-Crofts.

LOVIBOND, S. H. (1961). "The theory and method of conditioning in relation to enuresis and its treatment". Adelaide: Unpublished Ph.D. thesis.

MOWRER, O. H. (1950). *Learning Theory and Personality Dynamics*. New York: Ronald.

OSWALD, I. (1962). "Induction of illusory and hallucinatory voices with consideration of behaviour therapy", *J. ment. Sci.*, **108**, 196–212.

SPEARMAN, C. (1927). *Abilities of Man*. London: Macmillan.

WOLPE, J. (1958). *Psychotherapy by Reciprocal Inhibition*. Stanford University Press.

——(1961). "The prognosis in unpsychoanalysed recovery from neurosis", *Amer. J. Psychiat.*, **117**, 35–39.

YATES, A. (1958). "Symptoms and symptom substitution", *Psychol. Rev.*, **65**, 371–374.

H. J. Eysenck, *Professor of Psychology, University of London Institute of Psychiatry, London, S.E.5*

Behaviour Therapy, Extinction and Relapse in Neurosis

H. J. EYSENCK

BJP 1963, 109:12-18.
Access the most recent version at DOI: 10.1192/bjp.109.458.12