This paper provides a brief and selective overview of several areas of behavior therapy, or applied experimental psychology, with the usual concern for careful measurement, operationization of terms, and dispassionate examination of ideas which can be experimentally tested. The authors review the method of token reinforcement, with its subsequent problem of maintaining token-induced behavioral changes in nontoken environments. Second, they look at systematic desensitization, concluding that researchers are far from even a reasonable tentative answer to why the procedure works. A third area of research explored is the Masters and Johnson therapy, whose treatment package can be construed as behavior therapy. Finally, aversive therapy—avoidance or aversive "conditioning"—with homosexuals is reviewed, with special emphasis on the work of Feldman and MacCulloch. The authors feel that the very essence of behavior modification is its critical and experimental stance towards the whole area of clinical psychology and psychiatry. References are included. (Author/SES)
It seems appropriate to begin with our working definition of behavior therapy. We are strongly opposed to a conception of this rapidly expanding field which restricts conceptualizations to either or both classical and operant conditioning (Davison, Goldfried, and Krasner, 1970). Our reasons are simple. As we will shortly outline, there is insufficient evidence that our therapy procedures work for conditioning reasons. Secondly, as we begin to develop better conceptualizations of why what we do works, we may very well find that constructs other than those couched in conditioning terms are most useful for talking about what we do. As that happens, are we to disregard the techniques simply because they do not seem to be conditioning procedures? Thirdly, if we restrict ourselves to conditioning paradigms, we fear that our work may become sterile and self-limiting (Davison, in press). To be sure, any scientific endeavor takes place within a paradigm (Kuhn, 1962). But it would seem quite premature to act as if behavior therapy is in a period of normal science in which one can have confidence that a given paradigm, especially a narrow one, is the only one to work within.
Effects of Behavior Therapy

Genaid C. Davison and Suzanne Johnson Taffel

State University of New York at Stony Brook

It seems appropriate to begin with our working definition of behavior therapy. We are strongly opposed to a conception of this rapidly expanding field which restricts conceptualizations to either or both classical and operant conditioning (Davison, Goldfried, and Krasner, 1970). Our reasons are simple. As we will shortly outline, there is insufficient evidence that our therapy procedures work for conditioning reasons. Secondly, as we begin to develop better conceptualizations of why what we do works, we may very well find that constructs other than those couched in conditioning terms are most useful for talking about what we do. As that happens, are we to disregard the techniques simply because they do not seem to be conditioning procedures? Thirdly, if we restrict ourselves to conditioning paradigms, we fear that our work may become sterile and self-limiting (Davison, in press). To be sure, any scientific endeavor takes place within a paradigm (Kuhn, 1962). But it would seem quite premature to act as if behavior therapy is in a period of normal science in which one can have confidence that a given paradigm, especially a narrow one, is the only one to work within.
an extended period of time.

The paradigm we would suggest for behavior therapy is admittedly vague but at least is less restrictive than a view of the field which equates it with classical and/or operant conditioning. It is also not novel. We would define the field simply as applied experimental psychology, with the usual concern for careful measurement, operationization of terms, and dispassionate examination of ideas which can be experimentally tested. Whether the explanatory fictions one uses are couched in cognitive terms or conditioning terms or any other terms is, according to this view, quite irrelevant. This means that we would take issue with the view that behavior therapy is action- or behavior-oriented in some fashion that makes it distinguishable from other therapy procedures. Certainly, for example, Gestalt therapy is behavior-oriented, with the therapist taking a very active role in what goes on during each session. To take but one example, Perls (1970) suggests a technique called "reversal", in which a patient who feels a certain way, for example timid, is encouraged to act the opposite way, uninhibited. The similarity is striking to aspects of George Kelly's (1955) exaggerated role-taking and to Lazarus' (1966) behavior rehearsal, even though Gestalt therapy is usually regarded as humanistic-existential.

We turn now to the topic itself. The term "effects" can refer, of course, to either short-term effects -- those that can be observed
either during treatment or immediately following treatment, or long term effects — those that are discernible after some previously specified period of time, usually referred to as follow-up. A third slant on the issue is a search for mechanisms underlying whatever changes can be observed.

The question of long-term effects seems to raise an even more basic and intriguing issue, namely how people function ordinarily in the natural environment. It does not seem unfair to assert that the social and behavioral sciences are a long way from understanding how people actually function outside of experimental settings. Looked at in this way, the question of long-term effects might possibly be one that is only prematurely to be raised at this time. If, as most behaviorists believe, the human being is largely if not entirely a product of the shaping environment, truly long-term effects are possible only as a result of some kind of change in the controlling environment. On the other hand, to the degree that one conceives of the person as an active change agent himself, to that degree might one expect changes of one kind or another within treatment to generalize to a non-treatment situation, perhaps even working against the extratherapy environment.

To illustrate these points, we provide here a brief and selective overview of several areas of behavior therapy.
Token Reinforcement

The first area that we will touch on is token reinforcement. The use of tokens, especially in institutional settings, is prominent in much behavior modification work. Its effectiveness was impressively demonstrated by Aylion and Azrin (1965), who gave psychiatric patients tokens for engaging in certain specified desirable behaviors. These tokens were exchangeable for a variety of reinforcements such as watching TV, sitting in the dayroom, or having access to the canteen. The idea was that by rewarding desirable behavior with tokens that could be exchanged for back-up reinforcers, the frequency of desired behavior would increase. The early work, of course, of Staats and Staats (1963) is also in this tradition. Indeed, the efficacy of such token programs has been repeatedly demonstrated, not only in institutional settings for psychiatric patients, delinquents (e.g., Cohen & Filipczak, 1971; Meichenbaum, Bowers, & Ross, 1968), and the mentally retarded (e.g., Zimmerman, Zimmerman, & Russell, 1969), but in classroom settings as well (e.g., O'Leary, Becker, Evans, & Saudargas, 1969).

Investigators are presently convinced that token programs are dramatically effective in changing behavior while the token system is in effect. They are just beginning to turn their attention to questions of long-term therapeutic effectiveness in settings where tokens are not employed, and they are likewise asking which elements
of a token program are the essential ingredients for its success.

Consider, for example, the token work with psychiatric patients. Such programs have typically been aimed at the chronic, often schizophrenic patient, and have had remarkable success in increasing social interactions and self-care behaviors as well as decreasing apathy and aggression (Kazdin & Bootzin, 1972). Most investigations of this sort, however, have relied on the within-subject ABAAB design to demonstrate that the token procedure is functionally related to the changes in behavior. In such designs, a base line period of behavior is recorded (A), the tokens are introduced (B), then removed (A) and then reintroduced (B). The increases and decreases in behavior following the institution and removal of the tokens seem indeed to demonstrate the dependency of the behavior on the reinforcement contingencies. But, while such a design may show the importance of the token to reinforcement to behavioral change, it also points up a crucial problem. If the desired behavior decreases when tokens are removed, how will these gains be maintained in non-token settings outside the hospital? As yet, there is little hard evidence that such gains are maintained once the patient leaves the hospital. Atthowe & Krasner (1968) do report a doubling of discharge rates as a consequence of their token economy, but 46% of those discharged returned to the hospital within nine months. Readmission rates as low as 14% have been reported by other investigators using token
programs (Schaffer and Martin, 1966), but it is very difficult to attribute change in discharge or admission rates solely to the effects of the therapeutic treatment. Discharge and readmission are largely a function of administrative decisions, and in view of recent changes in emphasis in the mental health field toward increased discharge rates and the development of community resources to prevent prolonged hospitalization, it would be unwise to attribute any changes solely to token economies or to any other treatment employed, a point made by Kazdin and Bootzin (1972) in their recent review. Clearly, only controlled comparative studies which have not yet been done, can clearly decide this issue. (It may be mentioned that Gordon Paul has such a study underway now.)

At present the problem of maintaining token-induced behavioral changes in nontoken environments remains the critical issue for therapists employing these procedures. Similar statements can be made regarding token reinforcement in classroom settings. Again, there is ample demonstration of their effectiveness in decreasing disruptive behavior (e.g., Kaufman & O'Leary, 1972), increasing instruction-following and task attention (e.g., Broden, Hall, Dunlop and Clark, 1970), and improving grades and achievement test scores (e.g., Hewitt, Taylor, and Artuso, 1969). Again, most investigators have used the within subject ABAB design, and they have repeatedly shown the importance of the tokens to the behavior change; in so doing they have also pointed up the difficulty of ensuring maintenance
of the desirable behaviors in nontoken classroom settings. Investigators have, in fact, generally found that behavioral changes induced by token program do not always automatically transfer to nontoken periods within the same classroom (O'Leary and Drabman, 1971; Kazdin and Bootzin, 1972). Furthermore, it is not at all clear from the studies that do report behavioral maintenance (e.g., Walker, Mattson and Buckley, 1969) in nontoken situations exactly what is responsible for the continuation of the desirable behaviors.

Consequently, the thrust of present research is aimed at delineating exactly what procedures can be used to guarantee the transfer of behavioral improvements to non-token settings. To the extent that we see an individual's behavior as a function of the environment, the reprogramming of the natural environment becomes the solution to behavioral maintenance problems. This might lead, for example, to community residences (Fairweather, Sanders, Maynard, & Cressler, 1969), where chronic psychiatric patients can function in a somewhat protected non-institutional setting. Of course, to the extent that we see the individual not as a product of his environment but as a change agent within that environment, the reprogramming of the individual may be part of the solution to behavioral maintenance problems. Such research with children has been primarily centered around self evaluation and self reinforcement (e.g., Kaufman and O'Leary, 1972; Santogrossi, O'Leary, Romanczyk, and Kaufman, in press). As yet there is little evidence that such procedures are
effective over long periods, but work in this area is just in its infancy. One can assume that analogous attempts can be made to reprogram adult hospitalized patients; on the other hand it may be unrealistic to expect that patients such as those of Attlowe and Krasner's, who had been hospitalized for an average of 22 years, can really be reprogrammed or otherwise fitted into a natural environment which reinforces their attempts at adaptive behaviors.

Beyond this question of short- and long-term effects there remains the question of what the essential ingredients of token programs are. On the surface it may appear that changes brought about in token programs are simply due to the tokens and to their backup reinforcers. In actuality, however, most token programs employ many other variables such as social reinforcement and verbal instructions. Since the introduction and removal of the tokens are often confounded with the concomitant introduction and removal of these other variables, it becomes impossible to determine how much of the behavioral change is actually due to the token reinforcement alone. For example, during the token reinforcement phase, the staff typically dispenses considerable contingent attention and praise as well as tokens for good behavior. During the withdrawal of the tokens, social reinforcement is also frequently reduced. This produces a confounding of tokens, praise, attention, and social contact. By having so many variables acting together it becomes impossible to stipulate which of the variables are responsible for
the beneficial changes. One solution is to arrange for control conditions such as pretoken and token withdrawal phases to include all treatment variables except that of token reinforcement. Such a design is employed by O'Leary and his colleagues at Stony Brook in classroom settings with disruptive children. The influence of classroom rules, educational structure, and contingent use of praise is recorded before introducing tokens. Then, in a later stage of the experiment, when the tokens are withdrawn, rule structure and contingent social reinforcement remain in effect. In addition to helping parcel out the effects of the tokens themselves, this procedure permits the investigator to determine whether the token program needs to be instituted at all. With some subject populations, rules, structure and contingent use of praise may be all that is necessary to change behavior (e.g., Becker, Madsen, Arnold, and Thomas, 1967). In such cases there is certainly little point in introducing token systems with their concomitant administrative and monetary costs since the target behaviors can be readily changed by less artificial procedures.

**Systematic Desensitization.**

We turn now to a second area of research, that involving systematic desensitization (Wolpe, 1958). This technique is perhaps the best known and most extensively researched of the various behavior therapy procedures. The number of experiments, both good and bad, published over the past ten years is truly astounding.
People have examined whether the technique itself works for reasons other than contact and placebo (e.g., Paul, 1966; Lang, Lazovich, and Reynolds, 1965), and also have directed their attention to mechanisms that are best inferred to account for whatever beneficial effects the procedure seems to produce (e.g., Davison, 1968). Probably the most widely cited controlled outcome study is that of Gordon Paul (1966), in which it was shown that properly executed desensitization has significant anxiety-reducing effects on interpersonal performance anxiety over and above sheer contact with a therapist with the concomitant expectations of gain. Paul's work is a good example of careful attention being paid to both short term and long term outcome, for it will be recalled that in his follow-up, treatment gains stood up quite well over a period of two years following termination of therapy (Paul, 1967).

We want to mention, however, some recent work that casts doubt on such statements regarding specific desensitization treatment effects. McReynolds and his colleagues in Missouri (McReynolds, Barnes, Brooks, and Rehagen, in press) have questioned the believability of Paul's stress tolerance placebo as well as the relaxation control conditions used in studies like those of Lang, Lazovich and Reynolds (1965) and Davison (1968). He recently employed a stronger placebo control in a comparison with standard desensitization and with Paul's placebo control, and has checked for credibility of each procedure by both subjects and experimenters. He determined that
his new control was more believable than Paul's and as believable as desensitization on the part both of experimenters and subjects. And he found on a number of overt and attitudinal measures of treatment effect that desensitization did not differ from this 'souped-up' control, while it was superior to Paul's control and to the no treatment controls (thereby replicating Paul).

In addition, Bernstein at Illinois has pointed to the importance of demand characteristics in the post-treatment assessment situations; that is, he has suggested desensitization may well demand less fearful performance in post treatment than less impressive controls. Taken together, the research of McReynolds and Bernstein warrants a reexamination of the hitherto accepted finding that desensitization achieves its fear reduction effects via some sort of specific learning process over and above placebo effect, a very sobering possibility indeed.

As far as process-mechanisms are concerned, we have recently examined both animal and human data (Wilson and Davison, 1969; Davison and Wilson, 1972), and have come to the conclusion that we are far from even a reasonable tentative answer to why the procedure works (if, indeed, it works for reasons other than placebo). Is it exposure alone to the anxiety-arousing stimuli (e.g., Nelson, 1966); does relaxation work as a substitute for anxiety (e.g., Davison, 1968) or does it work to reduce arousal to a point where...
long term habituation can take place (Mathews, 1971); is encouragement and reinforcement from the therapist the important factor as the person progresses up the hierarchy (e.g., Leitenberg, Agras, Barlow, and Oliveau, 1969), or does the person relabel the situation as non-threatening (e.g., Valins and Ray, 1967; Rosen, 1972) or himself as one who can tolerate stress (e.g., Goldfried, 1971) -- all these factors and more are currently vying for recognition as the most important variables. It is indeed a fascinating question, and one which will continue to form the basis for both good and bad experimental work. We mention this research here obviously not to provide good answers for reasons of effectiveness, but to indicate that behavior therapists tend to investigate an apparently efficacious procedure in such a way as to tease apart those variables that might account for beneficial outcome. This dismantling approach, introduced by Peter Lang more than 10 years ago, (Lang & Lazovick, 1963) characterizes process research in desensitization.

Masters and Johnson Therapy

A third area of research that we can allude to is the work of Masters and Johnson (1970), if we can assume for the moment that most of their treatment package can be construed usefully as behavior therapy. A very real danger is cultish superstition developing around Masters and Johnson, with an uncritical acceptance of what they assert to be important components of the treatment package.

For example, consider their insistence on a dual sex therapy team,
the assumption being that effective sex therapy can take place only if a male therapist is present to be "friend-in-court" for the male partner, and similarly for the female therapist. The argument is rather compelling and may even be true. However, it seems quite premature to conclude that it has in any way been demonstrated that this dual sex team is indeed necessary. In point of fact, Masters is quite explicit in stating that he initially sought the collaboration of a sensitive female therapist when he, himself, realized that he was having trouble understanding some of the things his female clients were telling him (Belliveau and Richter, 1970). It must also be kept in mind that many of the things Masters and Johnson prescribe have been done for many years by other therapists, for example, Semans (1956) and Wolpe (1958), without a dual sex team, and with apparently good clinical results.

Of course another possibly important factor is the so-called "going to Mecca" phenomenon, where suffering couples spend thousands of dollars to make a pilgrimage to St. Louis to see this renowned therapy team. These observations and others that could be made are certainly not intended as damning criticisms of their pioneering efforts. Rather, these comments are made to illustrate the orientation of us in behavior therapy have toward analyzing the effectiveness of a procedure in terms of the variables that may be responsible for beneficial outcomes.
Aversive therapy with homosexuals

The fourth and final area of research that we'll touch on is avoidance or aversive conditioning with homosexuals. Without question, the most impressive research here is that of Feldman and MacCulloch (1971). In their controlled trial, 30 homosexuals were divided into anticipatory avoidance, classical conditioning, and psychotherapy groups. Feldman and MacCulloch report a general superiority of the two so-called conditioning groups over the psychotherapy group in changing secondary homosexuals. No differences emerged between anticipatory avoidance and classical conditioning. They are unable to say, therefore, whether their outcomes are due to avoidance learning or classical conditioning, but can they say that improvement has anything at all to do with specific conditioning processes? We do not think so, not until some kind of random shock procedure is employed, which can rule out both instrumental and classical conditioning explanations, while holding constant the tremendous placebo and motivational confounds inherent in both techniques. Furthermore, no checks were made of the credibility of the psychotherapy condition, which was run by MacCulloch, who can hardly be seen as impartial vis-à-vis learning procedures.

We find it particularly interesting that at the end of their outstanding monograph, Feldman and MacCulloch introduce dissonance theory as a tool for understanding etiology and treatment. With regard to treatment, they showed themselves aware of the self-justificational which a homosexual might engage in by submitting himself to and remaini
with a painful therapy regimen. This being the case, what does conditioning necessarily have to do with either procedure? For all we know one could use backward or random conditioning and get the same dissonance effects. Feldman and MacCulloch seem not at all unaware of this possibility.

A recent study by Birk and his colleagues in Boston (Birk, Huddleston, Miller, and Cohler, 1971) adds some additional credibility to the learning formulation over and above placebo explanations. They designed an extensive and multifaceted treatment program for male homosexuals and compared treatment that entailed approximately 25 aviodance conditioning and classical conditioning sessions with the same number of placebo treatments. The experimental, or conditioning, group went through a complicated variation of the Feldman and MacCulloch paradigm, but one which utilized more operant response elements as well as a combination of classical conditioning and avoidance learning trials. A noteworthy procedure in the placebo treatment was the use of the same kind of male and female pictures, but without shock. Among the 18 patients in the study, opinion was evenly divided as to which treatment would be more effective, and the expectations did not diverge during the respective treatments. This independent check on the plausibility of placebo treatment is invaluable and only infrequently used in behavior therapy research.

The results show that significantly more of the experimental
group reported striking changes in the direction of increase in heterosexuality, including interest in female bodies. They revealed also a diminution of homosexual urges and ideation. Showing themselves to be good clinicians, Birk et al also report on the importance of available female partners once the avoidance conditioning seems to suppress homosexual urges. Of principal interest here is evidence that contact alone and commitment to therapy did not seem to account for improvement in this particular avoidance conditioning trial.

Conclusion

An important feature of behavior modification is a concern with both the effectiveness and efficiency of various behavior change techniques. As applied experimental psychology, behavior therapy tries to use whatever findings arise from more basic research in the formulation of treatment packages of maximum effectiveness in the clinical situation. With their concern for replicable and reliable observations and results, behavior therapists tend to be (or at least ought to be) very hard task masters on themselves before they draw conclusions about effectiveness of specific therapy interventions and the reasons for such effectiveness. The very essence of behavior modification is its critical and experimental stance towards the whole area of clinical psychology and psychiatry. What we hope we've done here is at least convey some of the flavor of clinical and experimental research in the field, illustrating with a few selected
examples some of the questions that are being asked and how people are going about trying to answer them in what we hope will be a cumulative applied science of behavior change.
References


Davison, G. C. Counter-control in behavior modification. In L. A. Hamerlynck, L. Handy, and E. Mash (Eds.) *Critical issues in*
behavior modification. Champaign, Ill.: Research Press, in press.


Lazarus, A. A. Behavior rehearsal vs. non-directive therapy vs. advice in effecting behavior change. *Behaviour Research and Therapy*, 1966, **4**, 209-212.


