

An Alternative Perspective on Biofeedback Efficacy Studies: A Reply to Steiner and Dince¹

Donald G. Kewman²

University of Michigan

Alan H. Roberts

Scripps Clinic and Research Foundation

Clinical applications of biofeedback have proliferated and considerable lore surrounding the application of these techniques has evolved. Many assertions about the effectiveness of biofeedback training are based on findings of the least well-controlled studies, while many of the better controlled studies have failed to show that biofeedback directly mediates target symptoms or is superior to other treatments. Steiner and Dince (1981) suggest that the failure of these controlled studies is primarily attributable to methodological deficiencies. We believe that the question of whether or not there is a specific effect of biofeedback training is still frequently confused with the question of whether or not the treatment package as a whole has therapeutic value. Biofeedback is often therapeutic; however, evidence is often lacking that its effectiveness is due to biofeedback-trained changes in a target physiological process.

Descriptor Key Words: biofeedback research methodology.

In a commentary by Steiner and Dince (1981), the authors criticized efficacy studies of biofeedback techniques that have failed to show a specific or active effect of biofeedback or superiority over other treatments. It is their contention that methodological deficiencies account for the fact that many

¹We express our appreciation to Joel Roberts for his help in the data analysis and to Ann Hubbard for her help in the preparation of this manuscript.

²Address all correspondence to Donald G. Kewman, Department of Physical Medicine and Rehabilitation, University Hospitals, Ann Arbor, Michigan 48109.

controlled studies have failed to demonstrate that biofeedback treatments directly mediate improvement of target symptoms or are superior to alternative treatments.

While we support their view that better biofeedback efficacy studies are needed, we disagree with Steiner and Dince on many points. We are also perplexed by the theme of their paper, which supports the need for better research design yet levels strongest criticism at the more methodologically sophisticated studies. Ironically, the effect of their article can be interpreted as a defense of the results of methodologically inferior studies that form the basis of many claims about the efficacy of biofeedback.

HISTORICAL PERSPECTIVE

Steiner and Dince begin their paper with an attempt to place biofeedback in historical perspective. It is too early to write a definitive history, and alternatives to their perspective (e.g., Katkin, 1982) are possible. In our opinion, the field of biofeedback did not arise out of a search for methods to treat specific psychophysiological problems. The study of biofeedback applications was stimulated by research aimed at answering fundamental questions about whether or not the autonomic nervous system could be modified by instrumental learning. Miller's (1969) influential report summarized a series of animal studies that seemed to demonstrate the operant learning of several visceral responses. That same year Kamiya (1969) described his experiments on the operant control of EEG alpha in human subjects.

Soon after these and similar reports were published, serious questions were raised about the findings and interpretations. Fourteen years later, we are still not certain that the autonomic nervous system is subject to instrumental learning without mediation by striate musculature. From a clinical point of view, we are still without definitive evidence that the autonomic nervous system can be altered to a clinically significant degree by means of biofeedback without mediation from a variety of confounding variables.

Despite this shaky foundation of unreplicated early studies and unresolved theoretical issues, enthusiastic support and a large biofeedback industry composed of clinicians and equipment manufacturers has evolved and grown. Numerous reviews of the literature on biofeedback treatment have appeared recently (e.g., Jessup, Neufeld, & Merskey, 1979; Turner & Chapman, 1982). Some strongly suggest that biofeedback efficacy has been demonstrated only with respect to functions mediated by striate muscle. Even for these functions, there is uncertainty as to whether the efficacy of

biofeedback exceeds that of relaxation training alone. Further, the operation of Roberts's Third Law (Roberts, 1982) would provide enough intermittent reinforcement for therapists to maintain their behavior with great resistance to extinction even in the absence of a specific treatment effect.³

EXPERIMENTAL METHOD

Steiner and Dince appear to confuse the fundamental aims of scientific study with the fundamental aim of therapy. The purpose of scientific study is to apply the scientific method in order to understand, predict, and control observable phenomena. The primary purpose of therapy is to change the behavior of patients. Experiments need to be designed that delineate the essential ingredients in biofeedback treatment and how they can be manipulated most effectively in order to produce clinical change. Even when biofeedback efficacy studies have positive outcomes, they frequently do not address the issue of necessary and sufficient conditions. Is it feedback about changes in specific physiological processes, the therapeutic relationship, or some interaction between the two of these that mediates change? Would the outcomes be the same if an impressive machine with leads, electrodes, dials, sounds, and meters were substituted for the biofeedback equipment and everything else was held constant, including the therapist's and client's belief in the efficacy of the treatment? In a therapeutic relationship where biofeedback is only part of the treatment, there is no way to be certain if the biofeedback is either necessary or sufficient to accomplish the desired treatment outcome.

As Steiner and Dince point out, "ignorance of a mechanism of action is not equal to lack of clinical efficacy" (p. 275). The demonstration of clinical efficacy, however, does not prove that biofeedback training of a target physiological process is either necessary or sufficient to mediate the clinical changes that may occur. It may not be necessary to understand exactly how changing a target process leads to improvement, but it is important to determine if the trained changes significantly contribute to the therapeutic effect obtained by the treatment package as a whole. Continuing to maintain that a particular physiological mechanism of therapeutic action in biofeedback training even exists, when you are unable

³Roberts' Third Law was derived from studies and reports of treatments that were thought to be specific at the time they were published but that were later shown to be nonspecific. It states that if both the patient and the clinician believe in the efficacy of a treatment for a nonspecific disease, outcomes based upon nonspecific effects will approximate one-third "excellent" results, one-third "good," and one-third "poor" results.

to prove that it has a measureable effect in the first place, is more the province of faith than of science.

The issue is not whether treatment with biofeedback is therapeutic. *It is*. The literature is replete with evidence that some patients treated for various problems with biofeedback improve. However, faith healing, acupuncture, electrical stimulation, psychotherapy, dream analysis, and placebo medications also improve symptoms. The task of the scientist is to understand how biofeedback works when it does work, and what will improve its efficacy.

BIOFEEDBACK RESEARCH METHODOLOGY PROBLEMS

We agree with Steiner and Dince that most biofeedback efficacy studies are not exemplary models of research. We also agree that it is important not to generalize uncritically about the biofeedback field as a whole. However, there are many methodological problems that are common to a number of different techniques.

1. Baseline periods are too short or not used at all. Normal adaptational changes from the patient sitting quietly in a comfortable room are erroneously interpreted as evidence of biofeedback-trained control of the target response.

2. Little or no attempt is made to preserve the diagnostic homogeneity of the study sample.

3. There is no attempt to assess the quantitative relationship between amount of control of the target physiological process and symptom change.

4. A mixture of treatments such as psychotherapy, EMG relaxation, autogenic training, relaxation training, or stress management instructions is administered to the patient in addition to biofeedback either deliberately or inadvertently (possibly as part of a warm, supportive therapist-patient relationship). This obviously makes it impossible to determine which of the procedures introduced is responsible for the therapeutic benefit claimed.

5. Inadequate control groups are used to assess the contributions of general relaxation, self-monitoring by keeping symptom diaries, patient expectation, different therapist behaviors or styles, variations in training procedures, passage of time, weather changes (e.g., see Guglielmi & Roberts, 1983), or other nonspecific effects that are likely to be significant.

Variations in training procedures and the way the therapist functions were two issues highlighted by Steiner and Dince.

Therapist Variables

Steiner and Dince assert that the "expectations of the therapist" (p. 280) or the ways in which the therapist acts (p. 279) affect the outcome of

biofeedback treatment. This is a reasonable assertion but it does not address the issue of the relative contribution of these variables to biofeedback training. Indeed, there are some biofeedback studies suggesting that a role can be played by nonspecific factors that outweigh a specific effect of control of the target physiologic process. For instance, Gauthier, Bois, Allaire, and Drolet (1981) compared cooling versus warming the finger and cooling versus warming the temporal artery as a biofeedback treatment for migraine headache. Andrasik and Holroyd (1980) trained three groups of tension headache patients either to increase, decrease, or hold stable their frontalis EMG. The results of these studies were similar to other well-controlled studies that have generally shown improvement in subjects irrespective of the direction of regulation.

It is apparent that more work needs to be done to elucidate the nonspecific effects contributing to biofeedback efficacy. This is more than the assertion that placebo or nonspecific effects exert real and measurable therapeutic changes. It is a proposal to direct some of our research effort toward making nonspecific factors "more specific" in order to use them even more efficaciously. Plotkin (1980) has suggested experimental designs for studying this issue and has hypothesized various factors that merit careful investigation. These include (1) attribution of responsibility, (2) provision for the opportunities to experience success, (3) face validity, and (4) high level of patient acceptance and interest.

Steiner and Dince believe that untrained therapists are less effective than trained therapists. To support their statement, Steiner and Dince cited Taub and School (1978), which compared an impersonal therapist with a friendly and informal therapist. Friendliness, however, has no clear relationship to training. It is even conceivable that there is an inverse relationship between the degree of warmth or friendliness and the degree of training. The Biofeedback Society of America standards notwithstanding, *there are no experimentally determined standards defining what skills or behaviors constitute adequate training.*

Interestingly, the issue of the possible negative impact of an impersonal or untrained therapist seems to arise only when the results of a study fail to show a specific change caused by biofeedback. Published studies with positive results have also utilized inexperienced personnel.⁴ In fact, the ability of untrained therapists to obtain positive results can be viewed as an advantage of biofeedback treatments. Steiner and Dince's

⁴When Roberts, Kewman, and Macdonald (1973) demonstrated the ability of human subjects to differentially control peripheral skin temperature in their 1971 laboratory studies, the experimenters were untrained as "skin temperature biofeedback therapists." When their results were replicated (Roberts, Schuler, Bacon, Zimmerman, & Patterson, 1975), the subjects were run by two untrained students who nevertheless taught the subjects to control peripheral skin temperature. Cram (1980) also showed beneficial effects of EMG biofeedback on tension headache using essentially untrained undergraduate therapists.

belief that "it is possible for the therapist to act in a manner that overrides the effects of biofeedback treatment" (p. 279) is only one possible rationale for treatment failures and has not yet been substantiated by research.

Some researchers have begun to examine these issues. Blanchard et al. (1983) demonstrated that outcome from EMG biofeedback treatment for headache was not significantly related to the amount of experience of the therapist. They also showed that there was no significant relationship between outcome and the patient's perception of therapist competence, helpfulness, or warmth. Interestingly, they found a negative relationship between perceived warmth of the therapist and improvement, although this correlation did not quite reach significance ($p < .10$).

Hamberger and Lohr (1981) studied the effect of therapists' presence and response-contingent feedback in EMG biofeedback relaxation training. They found no difference between subjects who were trained with a therapist present and those who were trained alone in a more isolated environment. They concluded that a trainer's presence is not a crucial variable in facilitating subject-controlled, response-contingent relaxation training. Interestingly, the isolation of the subject from the "unblind" experimenter has been used as a criticism of the double-blind research design in biofeedback research.

Training Procedures

Steiner and Dince raised the issue of the relationship between adequate biofeedback training of the patient and the effectiveness of biofeedback treatment. They asserted that the administration of training is not equal to the subjects learning a biofeedback skill to criterion. This statement is a logical and common criticism of "unsuccessful" studies, but *there is no experimentally validated training criterion for most biofeedback techniques*. It is important to note that no study has shown that reliable learning occurs either to the arbitrary criteria enunciated by Fahrion (1977) or to any other criterion.

In the case of skin temperature, numerous studies have failed to validate voluntary vasodilation for both normal subjects and patients (Guglielmi, Roberts, & Patterson, 1982). Surwit (1978) concluded that "a review of the experimental literature produces only equivocal evidence for the ability of humans to learn to voluntarily vasodilate with the use of temperature feedback alone" (p. 10).

When "learning" does not take place in the laboratory setting, one may reasonably question whether learning occurs in the clinical setting. If it really occurs in the latter and not in the former, is the learning due to the biofeedback, to the setting, or to some interaction? The assumption that

research studies are not “clinical” and that patients are treated differently in the laboratory by inadequately trained personnel is simply an assumption, and there is no evidence to support this.

Steiner and Dince feel that in the training of a biofeedback task, one must specify the terminal behavior, provide the subject with an example of the task, give the subject a demonstration of the task, and provide instruction as to how someone might achieve that end. In fact, just the opposite may be true. The Task Force of the Biofeedback Society of America (Carlson, 1978) concludes their review of the experimental evidence concerning knowledge of the response by stating that “there appears to be no basis for the claim by many clinicians that awareness of the feedback-relevant response is necessary to achieve self-control over the response. . . . [I]n fact, the weight of the evidence to date indicates that non-awareness produces results equal to or better than awareness” (p. 7).

Steiner and Dince cited work suggesting that “eight to twenty-five hours of training have proven necessary for effective and lasting symptom alleviation” (p. 280). Stimulated by this contention, we decided to examine some previously unpublished follow-up data from a clinical biofeedback laboratory staffed by highly competent therapists meeting the standards for certification by the Biofeedback Society of America.⁵ The subjects were 55 patients with medical diagnoses of migraine headaches and 72 patients with medical diagnoses of muscle contraction headaches. Three months to 1 year following completion of biofeedback treatment, the patients were contacted for follow-up data, which consisted in part of ratings of overall improvement from 0 (“none”) to 4 (“complete”). Separate ratings of improvement were also obtained for changes in intensity, frequency, and duration of headaches, as well as reduction in the use of medication.

These data were analyzed separately for the two kinds of headaches with respect to the number of biofeedback training sessions the patients completed before terminating therapy. For migraine headaches, the number of training sessions correlated $-.02$ with overall ratings of improvement and $-.20$ with the number of categories of improvement. For muscle contraction headaches, the correlation was $.02$ for both overall ratings of improvement and the number of categories of improvement.

The subjects in the two groups were also divided into those who rated their improvement as either “none,” “some,” or “moderate” ($n = 39$ for muscle contraction and $n = 28$ for migraines) and those who rated improvement as “considerable” or “complete” ($n = 33$ for muscle contraction and $n = 27$ for migraines). For muscle contraction headaches the

⁵We are grateful to Richard A. Sternbach for providing these data from patients treated in the clinical biofeedback laboratories of the Scripps Clinic Pain Center, which he directs.

mean number of training sessions was 7.45 ($SD = 1.93$) for moderate or less improvement and 7.77 ($SD = 2.34$) for considerable or better improvement. These means are not significantly different from each other ($t(70) = 1.370$). For migraine headaches, the mean number of training sessions was 7.61 ($SD = 2.44$) and 7.89 ($SD = 2.26$), respectively, for the two outcome groups. These means are also not significantly different ($t(53) = 1.045$).

For this follow-up study, at least, we cannot demonstrate any relationship between hours of training and effective and lasting symptom alleviation. Similarly, we are unaware of any published studies that have conclusively shown a clear quantitative relationship between the magnitude of control of physiological processes using biofeedback and degree of clinical improvement.

INTERPRETATION OF EXPERIMENTAL OUTCOMES

Steiner and Dince have pointed out that the failure to find an effect in a research design is not the same as proving the absence of an effect. However, the *superiority* of biofeedback is not demonstrated if the effects of biofeedback are not significantly better than those obtained by control groups.⁶

Steiner and Dince understand this point because they invent an example of a "new drug morphine," which is compared with a sterile water placebo. In their fictional example, the placebo group shows a large reduction in pain, and morphine is not significantly better than the placebo.⁷ They suggest that one should not conclude that morphine lacks analgesic properties, but rather that "the placebo effect was so powerful that it was impossible to demonstrate morphine's analgesic effects" (p. 277). From their example it should be clear, however, that a difference that makes no difference is no difference at all. If one can achieve results with

⁶Two studies (Kewman & Roberts, 1980; Blanchard, Theobald, Williamson, Silver, & Brown, 1978) were erroneously cited by Steiner and Dince (p. 279) as failing to demonstrate the effectiveness of biofeedback. Blanchard et al., (1978) did show efficacy for biofeedback versus headache monitoring but failed to show its superiority over relaxation. Migraine patients receiving biofeedback in Kewman and Roberts (1980) also improved, but the specific therapeutic value of increasing finger temperature was not shown.

⁷Steiner and Dince's example is not too far-fetched. Recent studies (VanDyke & Bick, 1982) have shown that experienced users of cocaine cannot differentiate it from cheaper, legal lidocaine, procaine, or placebo in a laboratory situation. Ironically, VanDyke and Bick conclude their review with the statement that "many of the questions surrounding the issue of cocaine can be answered by scientific investigation, but the final decisions about cocaine will be political and economic, not scientific" (p. 141). It is profoundly hoped that the same will not be true of biofeedback.

placebo that are close to those one achieves with biofeedback, then the advantages of biofeedback are only greater to the patient if it is more efficient and more economical, or if there are fewer side effects.

Steiner and Dince criticize some of the more powerful procedures for controlling for nonspecific effects in biofeedback treatment such as the double-blind design. It is true that a double-blind experimental design is not the only way to control for nonspecific effects, but it is one procedure that can be used to see if the specific effects of a treatment are sufficiently powerful to show a difference over nonspecific factors. Studies that have used this or similar designs unfortunately have not been able to demonstrate a specific or superior effect of the biofeedback despite clinical improvement of the patients treated.

In discussing Type I and Type II statistical errors, one must realize that these issues are as applicable to clinical reports as they are to studies using group comparisons. The biofeedback literature contains numerous reports of positive outcomes with only a few subjects and no controls. The likelihood of committing a Type I error is exceedingly high in these kinds of reports. The problem is further confounded by the fact that the negative results of clinical or research studies are often not published.

Steiner and Dince's suggestion for the use of power analysis in future studies is a good one. When small numbers of subjects are used, it may also be helpful to use groups of subjects who are matched according to initial baseline values or degree of learning of the target skills. Furthermore, longer and better follow-up is needed. This may be particularly important in trying to determine if a specific effect of biofeedback training may be more efficacious than the nonspecific effects because of duration of action.

CONCLUSIONS

As clinicians, we find it too easy to discount research findings that do not support our beliefs by viewing the studies as poor analogues to everyday clinical situations. Yet the major difference between the clinical and laboratory setting is that the experimenters in the laboratory are carefully applying certain rules to the collection and analysis of data. By doing this, they are avoiding the pitfall of allowing fallible clinical judgments to supplant scientifically derived conclusions.

As clinicians, we feel that there is nothing wrong with trying to help our clients with only partly understood treatment modalities. At the same time, it is important to realize that a treatment may work for reasons other than the ones we believe in. It may also be true that some of the things we believe in may be effective only because we believe in them. It is important to keep in mind that ethical guidelines for psychologists require us to

provide clients with a thorough discussion of the limitations of our data as well as to acknowledge the existence of other hypotheses and explanations (American Psychological Association, 1981).

No single study or set of controls will provide the definitive answer regarding the therapeutic efficacy of biofeedback treatment. It is not necessary to argue that experimental studies are complete, perfect, or definitive. It is only necessary to point out that experimental studies attempt to control variables ignored or neglected by clinical reports or other studies and therefore add a dimension to knowledge and understanding. The challenge to biofeedback researchers is not to abandon controlled studies because of their inherent difficulties and limitations, but rather to carry out more sophisticated and better controlled experiments that will enhance our understanding of the phenomenon of biofeedback.

REFERENCES

- American Psychological Association. (1981). *Ethical principles of psychologists*. Washington, D.C.: Author.
- Andrasik, F., & Holroyd, K. A. (1980). A test of specific and nonspecific effects in the biofeedback treatment of tension headache. *Journal of Consulting and Clinical Psychology, 48*, 575-586.
- Blanchard, E. B., Andrasik, F., Neff, D. F., Saunders, N. L., Arena, J. G., Pallmeyer, T. P., Teders, S. J., Jurish, S. E., & Rodichok, L. D. (1983). Four process studies in the behavioral treatment of chronic headache. *Behaviour Research and Therapy, 21*, 209-220.
- Blanchard, E. B., Theobald, D. E., Williamson, D. A., Silver, B. V., & Brown, D. A. (1978). Temperature biofeedback in the treatment of migraine headaches: A controlled evaluation. *Archives of General Psychiatry, 35*, 581-588.
- Carlson, J. G. (Ed.). (1978). *Biofeedback as a research tool* (Task Force Study Section Report). Denver: Biofeedback Society of America.
- Cram, J. R. (1980). EMG biofeedback and the treatment of tension headaches: A systematic analysis of treatment components. *Behavior Therapy, 11*, 699-710.
- Fahrion, S. L. (1977). Autogenic biofeedback treatment of migraine. *Mayo Clinic Proceedings, 52*, 776-784.
- Gauthier, J., Bois, R., Allaire, D., & Drolet, M. (1981). Evaluation of skin temperature biofeedback training at two different sites for migraine. *Journal of Behavioral Medicine, 4*, 407-419.
- Guglielmi, R. S., & Roberts, A. H. (1983). Of cold-handed patients and warm-hearted therapists: A reply to Surwit and Keefe. *Biofeedback and Self-Regulation, 8*, 3-8.
- Guglielmi, R. S., Roberts, A. H., & Patterson, R. (1982). Skin temperature biofeedback for Raynaud's disease: A double-blind study. *Biofeedback and Self-Regulation, 7*, 99-120.
- Hamberger, L. K., & Lohr, J. M. (1981, April). *The effect of therapist's presence and response-contingent feedback in biofeedback relaxation training*. Paper presented at the meeting of the Midwestern Psychological Association, Detroit.
- Jessup, B. A., Neufeld, R. W., & Merskey, H. (1979). Biofeedback therapy for headache and other pain: An evaluative review. *Pain, 7*, 225-270.
- Kamiya, J. (1969). Operant control of the EEG alpha rhythm and some of its reported effects on consciousness. In C. Tart (Ed.), *Altered states of consciousness*. New York: Wiley.

- Katkin, E. S. (1982). On reliable knowledge and the proliferation of professional schools of psychology. *Clinical Psychologist, 36*, 9-11.
- Kewman, D. G., & Roberts, A. H. (1980). Skin temperature biofeedback and migraine headaches: A double-blind study. *Biofeedback and Self-Regulation, 5*, 327-345.
- Miller, N. E. (1969). Learning of visceral and glandular responses. *Science, 163*, 434-445.
- Plotkin, W. B. (1980). *The placebo effect, self-healing and biofeedback: The role of faith in therapy*. Manuscript submitted for publication.
- Roberts, A. H. (1982). Contingency management methods in the treatment of chronic pain. In J. J. Bonica, U. Lindblom, & A. Iggo (Eds.), *Advances in pain research and therapy* (Vol. 5). New York: Raven.
- Roberts, A. H., Kewman, D. G., & Macdonald, H. (1973). Voluntary control of skin temperature: Unilateral changes using hypnosis and feedback. *Journal of Abnormal Psychology, 82*, 163-168.
- Roberts, A. H., Schuler, J., Bacon, J., Zimmermann, R. L., & Patterson, R. (1975). Individual differences and autonomic control: Absorption, hypnotic susceptibility and the unilateral control of skin temperature. *Journal of Abnormal Psychology, 84*, 272-279.
- Steiner, S. S., & Dince, W. M. (1981). Biofeedback efficacy studies: A critique of critiques. *Biofeedback and Self-Regulation, 6*, 275-288.
- Surwit, R. S. (1978). Peripheral vasomotor: Raynaud's disease. *Behavior Medicine Newsletter, 1*, 9-11.
- Taub, E., & School, P. J. (1978). Some methodological considerations in thermal biofeedback training. *Behavior Research Methods and Instrumentation, 10*, 617-622.
- Turner, J. A., & Chapman, C. R. (1982). Psychological interventions for chronic pain: A critical review. I. Relaxation training and biofeedback. *Pain, 12*, 1-21.
- VanDyke, C., & Bick, R. (1982). Cocaine. *Scientific American, 246*, 128-141.