META-ANALYSIS: AN ABUSE OF RESEARCH INTEGRATION

H. J. Eysenck, Ph.D., D.Sc. University of London

The paper reviews many objections to metaanalysis, both in principle, and in particular, as practiced by its originators. Particular attention is given to the most widely cited application of meta-analysis: to the facts of psychotherapy. It is suggested that the claims to inclusiveness and objectivity, made for these analyses, are not supported by the evidence, and that the conclusions drawn from the evidence surveyed are incompatible with the results of the analysis on which they are supposed to be based. It is further suggested that no pseudo-objective computerized technique can substitute for the scientific insight and theoretical acumen of the investigator. Indeed, for most purposes the simplistic scoring systems of meta-analysis are not only useless, but may be counterproductive. It is precisely in those areas where there is most disagreement that these methods are least applicable.

Meta-analysis is offered to us as a technique that obviates certain pitfalls of integrating divergent research findings, giving a greater degree of objectivity than is forthcoming in the usual type of research summary. We are told that the objective of the research review is to analyze and present the separate studies in such a way that an overall conclusion can be reached about the nature of the process studied. How is this normally done? Smith, Glass & Miller (1980) state, "Data in the form of findings from research studies are aggregated or accumulated almost in the same way as measurements on individuals are accumulated in primary research to form conclusions about the variable studied" (p. 7). The major problem in the way of coming to a conclusion, as they point out, arises when the findings of the individual studies do not agree, or the characteristics or contexts of the studies are different. The solution is usually sought through impugning the design or analysis employed by some researchers. But Smith, Glass, and Miller inform us, "although clothed in elaborate rationalisations, the process is dangerously vulnerable to the injection of prejudice and bias" (p. 8).

Meta-analysis is offered as a way out of the subjectivity apparently inherent in the traditional approach. The method seems to have originated with Glass (1976, 1977) and various collaborators (Glass & Smith, 1979; Smith & Glass, 1977). Smith, Glass, and Miller (1980) declare that the method was designed to satisfy three basic requirements: (a) Studies should not be excluded from consideration on arbitrary and a priori grounds, the major premise being that some boundaries must be drawn around fields, but it is better to draw them wide than narrow; (b) study findings should be transformed to commensurable expressions of magnitude of experimental effect on correlational relationship; (c) features of studies that might mediate their findings should be defined and measured, and their covaria-

0022-4669/84/1801-0006\$02.00/0

tion with findings should be studied. The way the procedure works is well illustrated in Smith et al. (1980), and their book will be taken as the example of metaanalysis as discussed in this article.

Some knowledge of the methodology of meta-analysis being presumed in the reader, we may go straight on to a critique of the technique itself. Such a critique may take two forms, and both will be exemplified here. The first form would be a theoretical one, pointing out the unscientific nature of the whole enterprise, and the ways in which its assumptions depart from what is commonly agreed by philos-ophers of science to be the nature of scientific research. This will be our first task. The second form would be to take a particular example of meta-analysis (in this case the study of the effects of psychotherapy) and analyze in some detail what the authors have done, and in what ways their use of meta-analysis can be shown to be at least as subjective as are the analyses carried out by others. Such a review of a particular example will also serve to bring out many of the weaknesses of meta-analysis, and to demonstrate fairly conclusively its will-o'-the-wisp nature and its failure to generate acceptable and meaningful scientific conclusions. This will be our second task.

Let us begin with the undoubted fact that in many areas a number of empirical findings regarding a particular issue apparently point to different conclusions, some favorable and some unfavorable to a particular hypothesis. What Glass and his collaborators advocate is essentially to treat all these studies as equals, assess the degree of pro or con tendency of each study, sum over all the studies, and thus arrive at some form of statistical conclusion. No study, however poor in design or analysis, is omitted, such inclusivity apparently constituting objectivity. Where there are parameter differences, these may be summed separately in order to throw some light on the relevance of the parameter to the final conclusion. What, one may say, is wrong with such a treatment?

The answer, or so I would suggest, is simply this. In the first place, some studies are bad in the sense that in their design or analysis they disregard essential parts of the hypothesis to be tested. This means that they ought to be excluded from a consideration of the theory in question; their retention on the basis that exclusion is subjective is nothing more or less than an absurdity. The quality of psychological research in journals is not high, on the whole; this is documented by the high rejection rates as compared with hard science journals (80% vs. 20%, roughly) and by the low interreferee agreement (r = .20, roughly). Summaries that disregard this essential point are not to be taken seriously.

QUALITIES OF DATA

In the second place, we have to consider the quality of the data that go into the literature. Psychologists tend to assume rather lightheartedly that what is published in the literature (and has therefore passed the gatekeepers or referees who are supposed to guard the quality of scientific contributions) must be of sufficiently high standard to be used in summaries of the literature. Nothing could be further from the truth. Again, two examples stemming from our own work illustrate this point.

THE JOURNAL OF SPECIAL EDUCATION VOL. 18/NO. 1/1984

The first comes from our large-scale studies, continued for some 20 years, on pursuit rotor reminiscence (Eysenck & Frith, 1977). We are here dealing with rotary pursuit, a subject on which countless articles have been published, most giving detailed results about the effects of rest pauses on performance on the pursuit rotor. It might be thought that people would use identical apparatus and conditions, but this is not so; almost everyone changed conditions of performance (speed of rotation, length and makeup of stylus, size of target, height of apparatus, etc.) and conditions of testing, to such an extent that any real replication was almost ruled out. It is difficult to see how meta-analysis could deal with such a profusion of data all obtained on different apparatus, in different conditions, and with different parameters.

As part of my interest in these studies, I visited a number of the laboratories using pursuit rotors and publishing research along these lines. I found that conditions were often unsatisfactory, with speed of rotation varying from day to day without being checked; with the target disc and the rotating disc itself not being cleaned scrupulously before it was used; and with the target disc often being pitted, thus producing bounce in the stylus; and so on. These conditions might produce differences in the data output amounting to something like 10% to 15%, making results difficult to compare even if conditions had been identical. None of this could enter into a meta-analysis of the data, because the assumption there is made that everything is perfect. I argue that one's knowledge of the care with which experiments are controlled in different laboratories must play an important part in evaluating research emanating from these laboratories. Although such a judgment might be considered subjective, nevertheless it forms an important part of one's knowledge of the subject.

The same is true of work on eye-blink conditioning. Such study has been conducted in my department for some 30 years, and we have perfected what I think is much the most satisfactory apparatus for the measurement of this response. At various times we have corresponded with other laboratories working in this field, and have received typical print-outs of experiments performed there. In many cases we found this unsatisfactory. Indeed, the data would have been rejected by us had they come from one of our students. Again these faults do not show in the published literature; they become familiar to one through long association with the field, intimate knowledge of the laboratories and people in question, and so forth.

One last example. On one occasion I was invited to visit a psychologist who was working on my theory linking personality and eye-blink conditioning. He showed me an ongoing experiment, during which a train passed along a railway line only about 100 yards from the (not soundproofed) laboratory in which the experiment was being run. The noise was excruciating and upset the subject, but the experiment was continued and the data were included in the final analysis. Indeed, trains kept passing regularly and would have interfered with the measurement of the conditioned response in the majority of subjects! Yet when the psychologist published his account in a reputable journal, no mention was made of the conditions under which the data had been obtained, and the fact that his results seemed to disconfirm my theories was taken seriously by many critics. So long as psychologists adopt such a happy-go-lucky attitude towards the collection of data, so long will it be necessary to adopt a critical stance and to rely on one's knowledge of the adequacy of controls and the known ability of the experimenters. This may be a sad state of affairs, but it is only realistic to acknowledge that it exists. Metaanalysis, disregarding entirely any kind of qualitative analysis of data, is simply following the old adage: Garbage in, garbage out.

THE NEED FOR JUDGMENT

A third point where good judgment enters, and simply averaging fails, relates to the causes of failures to replicate. As my first example, let me take a controversy of some 20 years ago, when Kenneth Spence found that eye-blink conditioning was strongly correlated with emotionality-anxiety-neuroticism but not with extraversion-introversion. Conversely, my colleagues and I found that eye-blink conditioning was correlated with extraversion-introversion but not with neuroticism (Eysenck, 1981). A number of studies were published at the time, including efforts of outsiders to help resolve the argument. One can imagine Glass and colleagues doing a meta-analysis of all these data, and coming to the conclusion that neither extraversion-introversion nor neuroticism-stability had any overall effect on eyeblink conditioning! But that of course (and fortunately!) was not the way the controversy was solved. Gregory Kimble, who had also tried to replicate Spence's work and had failed (leading to the saying at the time that the correlation between anxiety and conditioning existed above the Mason-Dixon line but not below it), went to Spence's laboratory to find out what the experimental differences were.

He concluded that while he himself, our group, and most others had tried to reduce the anxiety of the subjects by keeping the apparatus out of sight, reassuring them that they would not receive any shocks, etc., Spence had done exactly the opposite, maximizing the degree of subject anxiety. Clearly, a trait like anxiety only becomes relevant to a particular experiment when the subjects are exposed to varying degrees of state anxiety; if the situation is not anxiety provoking at all, then differences in tendency to be provoked lose their relevance. Hence the failure of Kimble, Eysenck, and others to find the correlation Spence had repeatedly reported. Conversely, my theory states that a high degree of anxiety would produce, along various routes, a high degree of cortical arousal, thus wiping out the differences in cortical arousal at the resting level that characterize introverts and extraverts. This would lead to the prediction that introverts would condition better than extraverts. Thus the solution was not to do a meta-analysis but rather to investigate more closely the parameters of the experiments, the differences among experimenters, and the details of the theories involved. Only when this was done could the mystery be solved. This, I would suggest, is a more scientific way of discovering why there are different results in a particular field. Meta-analysis has no way of approaching the scientific solution of such a problem (Eysenck, 1981).

Some investigators failed to find a correlation between introversion and conditioning, and reported that they had failed to replicate my findings. When looked at closely, however, it appeared that they had fallen foul of one particular aspect of the theory, namely that the prediction only held for relatively weak unconditioned

THE JOURNAL OF SPECIAL EDUCATION VOL. 18/NO. 1/1984

stimuli (UCS). Using very strong electric shock as a UCS, they automatically disqualified themselves from replicating the original studies. Adopting the Glass type of precept, their studies would have been included, not excluded for what they would no doubt consider subjective reasons. Nevertheless, from a scientific point of view the study is irrelevant to the theory (Eysenck, 1981).

Glass et al. might argue that by averaging parameter values they might have discovered the same thing, but this seems doubtful. How would one compare parameter values of strength of the UCS when some people use electric shock, others a puff of air to the eye, and yet others a loud noise delivered over earphones? All this calls for highly specialized knowledge in specific fields applied to particular investigations, not an overall averaging procedure that in the nature of things would obscure such details.

As another example, consider the work done on intersensory effects in the measurement of thresholds. Since Urbantschitsch's experiment in 1883, there has been evidence that the perception of visual, auditory, tactile, pressure, pain, or olfactory or gustatory stimuli can be facilitated by simultaneous heteromodal stimulation. There have, however, also been reports to the contrary (see reviews by Bartley, 1958; Gilbert, 1941; Harris, 1950; Hartmann, 1935; Kravkov, 1966; London, 1954; Loveless, Brebner, & Hamilton, 1970; Maruyama, 1964; Ryan, 1940; Shigehisa, 1972; Stern, 1935; Symons, 1954). As Shigehisa and Symons (1973a) point out, "Many of the results reported on this subject have been quite divergent and often contradictory, hence no generally tenable conclusions have emerged" (p. 205).

How do Shigehisa and Symons approach the problem of complete inconsistency of reported results? They put forward the hypothesis that in the case of intersensory effects the intensity of the heteromodal stimulus and certain subject variables may be involved in differentiating these divergent effects. They go on to point out that recent work, which attempted to reconcile these inconsistencies, did not sufficiently explore the possibility that the intensity of the heteromodal stimulus could, perhaps depending on the personality variables, exert a direct effect on responses in the primary modality. It is suggested that an increase in intensity of the heteromodal stimulus may lead to effects that are the reverse of those induced by weaker intensities—a hypothesis that has been labeled the law of inversion in the Russian work, paralleling Pavlov's hypothesis of transmarginal inhibition. They further suggest the possibility that personality type is systematically related to the relationship between the intensity of the heteromodal stimulus and sensitivity in the primary modality.

Using Eysenck's (1967) model of extraversion-introversion, and the hypothetical causal element in this personality dimension of cortical arousal and reticular activation, they argued that introverts have lower sensory thresholds and show less inhibition to continued stimulation. Extraverts, on the other hand, have higher thresholds and show greater inhibition of continued stimulation. It follows that a given intensity of heteromodal stimulation may be perceived as higher by introverts than by extraverts. This, taken together with the theory of transmarginal inhibition, "predicts that introverts would show a greater decrease in perform-

ance, or decrease at lower intensities of stimulation, than extraverts" (p. 206). It is predicted, therefore, that introverts will show a greater decrease in sensitivity to the primary stimulus, or a decrease at weaker intensities of the heteromodal stimulus, than will extraverts. This prediction, in rough diagrammatic form, is shown in Figure 1. On the ordinate is shown a scale of auditory thresholds, from high to low; on the abscissa is shown the intensity of heteromodal stimulation (light), ranging from 1 to 10 on an arbitrary scale. I stands for introverts, A for ambiverts, and E for extraverts.

A series of carefully planned and executed experiments was done by Shigehisa and his colleagues, documented in Eysenck (1976). At first the heteromodal stimulus was light, the threshold was determined for auditory stimuli, and results were replicated in a second experiment. Next the investigators changed the two sets of stimuli around, using auditory stimuli as the heteromodal stimuli, and determined thresholds for light. The whole series was repeated in Japan, with Japanese subjects. In all, these studies demonstrated that the hypothesis was strongly supported, along the lines indicated in Figure 1. We now have, therefore, a better



Figure 1. Hypothetical relation between (a) sensory threshold in one modality and (b) strength of heteromodal stimulation and personality.

insight into the general question of intersensory effects and their relations to personality, the law of inversion, and the reasons for the differential effects noted by so many investigators. All this is a consequence of using scientific insight and knowledge of the relevant fields to advance a hypothesis that can integrate the different findings, and to suggest experiments that can support or disprove the theory in question. None of this would have been possible through a simple meta-analysis of the existing data; indeed the mind boggles what such an analysis might have revealed!

To sum up this section, what characterizes a good scientist, and separates him from the bad, is the quality of judgment. A scientist constantly has to appraise experiments, theories, analyses, books, and articles in his field of interest. These judgments are based on his general knowledge of the field and the people concerned, his methodological and statistical ability, and many other less obvious sources. The quality of his final judgment is basic to his success or failure as a scientist; it is based not only on what is reported but also on many other factors that he knows, or suspects, or intuits from the literature. To try and reduce this complex and important aspect of a scientist's work to a simple additive statistical procedure, which even a young and inexperienced student can easily follow, is to make a laughing stock of the whole business of science. No simple mechanical addition of diverse and incommensurate studies can serve the purpose of drawing meaningful conclusions from heterogeneous and complex data. That requires experience, knowledge, and the intangible quality we call good judgment. That there is some subjectivity in this is inevitable, and of course there might be arguments about what is good and bad judgment. Thus the difficulties that arise cannot be overcome by the simplistic methodologies of meta-analysis, which simply sidestep the difficulties and summate the absurdities that sometimes accompany such judgments.

THE PLACE OF THEORY

When Copernicus proclaimed his heliocentric theory, one obvious potential proof was the existence of stellar parallax. This phenomenon was not observed for 250 years after he enunciated his theory, and a meta-analysis of the studies in his time would have indicated clearly the nonexistence of parallax. Nevertheless, the heliocentric theory was accepted almost universally among good judges, and ultimately parallax was indeed observed. These scientists were judging on a much broader basis than a simple statistical addition of data. They took into account many other considerations, some of them intangible and difficult to put in words, but these are crucial in the elaboration of good scientific judgment.

In a similar way, Newton predicted that the earth was not a perfect sphere, but flattened at the poles. Several expeditions were mounted to collect data to support or disprove this hypothesis, with the results (on any form of meta-analysis) coming out rather against the theory. Nevertheless, it was universally accepted and finally photographs taken by orbiting space vehicles demonstrated that Newton indeed had been right. The relationship between theory and experiment is never as simple as it appears in the elementary textbooks. The factors that influence a good scientist's judgment are far more complex than can possibly be taken into account in the meta-analysis type of summary that Glass advocates. At best, meta-analysis allows us to see the truth in a glass, darkly.

I have stressed the importance of theory in science, and readers who feel inclined to debate the point may like to read a rather lengthy defense of my position published elsewhere (Eysenck, in press). My position is almost universally shared by philosophers of science and by working scientists in the hard sciences. Unfortunately it is not often shared by psychologists, who seem to take a rather Baconian view of the importance of induction rather than of inductive-deductive reasoning. Smith et al. (1980) do not seem to share this common view, because they quote Poincaré in their book to the effect that "science is built up with facts, as a house with stones. But a collection of facts is no more a science than a heap of stones is a house" (p. 23). Brave words, but, as will be apparent in our analysis of their actual procedure, it is the heap of stones they look at rather than the architecture of the house.

It would be interesting to pursue why psychologists have seemed rather hostile to theory. They often appear to be intent on a purely pragmatic, often eclectic kind of pursuit of simple facts, facts that they hope may somehow add up to science in the long run. There is no space here to indulge in such a pastime, but it may be useful to point out that the very notion of facts without theory is meaningless in science. As a simple lecture demonstration of this I sometimes ask a student audience what I have in my hand, and they say "a glass of water." I then ask them if that is a simple fact unrelated to any kind of theory, and they usually answer in the affirmative.

As a matter of fact the vessel is not made of glass at all; it is transparent plastic. The liquid is not H_2O ; it is almost pure ethanol in the form of vodka. To decide between the two theories (glass versus plastic, H_2O versus ethanol) one needs considerable chemical knowledge, theories regarding the nature of fluids, etc. To imagine that this is a simple factual problem, without involving any kind of scientific theory, is untutored thinking ignorant of the very basis of science. Whitehead said, "The utmost abstractions are the true weapons with which to control our thought of concrete fact." Darwin pointed out that he could not see how anyone could carry out an investigation except to support some theory or other. As we shall see, in practice meta-analysis disregards theory, and is content simply to count up alleged facts very much in the manner of the investigator who mistakes the heap of stones for a house.

Smith et al. (1980) depart from the path of scientific righteousness by making a distinction (originally made by Glass) between evaluative and elucidatory enquiry. The latter is defined as a process of obtaining generalizable knowledge by contriving and testing claims about relationships among the variables of generalizable phenomena, resulting in functional or statistical relationships, models, and ultimately theories. Such theories, taken together with knowledge of particular circumstances, enable one to obtain *explanations*.

Evaluative enquiry, on the other hand, is defined as "the determination of the worth of a thing" (p. 25). It apparently involves obtaining information to judge the worth of a program, product, or procedure. Apparently explanations are not the goal of evaluation; a proper and useful evaluation can be conducted, so we are

told, without producing an explanation of why a product or program is good or bad or how it operates to produce its effects. Smith et al. list a number of other ways in which evaluative and elucidatory enquiry differ, but it may be said in general that the distinction is a dangerous one, particularly as used by them. Fundamental to their confusion seems to be their belief that these two methods of enquiry are related, the elucidatory to the nomothetic, and the elucidatory to the idiographic type of investigation. The idiographic investigation, which they define as being "descriptive of the particular," is meaningless except in a context of nomothetic knowledge. We can only describe the particular in terms that are made meaningful by general theory and practice in science. When we describe a table as having a certain length and width, being a certain color, being made of a certain metal or wood, etc., we are relying on well-established theories and generalizations without which no description could proceed. It is impossible to divide the evaluative from the elucidatory, or to miss the point that the evaluation done in terms of metaanalysis (i.e., leaving out elucidatory aspects of the enquiry) must inevitably be unscientific, meaningless, and pointless. Unless we regard the outcome of therapy research as a scientific enterprise inevitably combining the elucidatory and evaluative types of enquiry, we will be engaged in what I once labeled "an exercise in mega-silliness" (Eysenck, 1978).

THE EFFECTS OF PSYCHOTHERAPY

Let us begin by looking at the definition adopted by Smith, Glass, and Miller, quoted from Meltzoff and Kornreich (1970):

Psychotherapy is taken to mean the informed and planful application of techniques derived from established psychological principles, by persons qualified through training and experience to understand these principles and to apply these techniques with the intention of assisting individuals to modify such personal characteristics as feelings, values, attitudes, and behaviours which are judged by the therapist to be maladaptive and maladjustive (p. 6).

Although Smith, Glass, and Miller do not say so explicitly, certain implications are concerned in this definition, in the sense that psychotherapists whose views are encapsulated in it would insist on the inclusion of these implications. One would be that the training and experience of the therapist are an important variable. If they were not, it would be meaningless to include a sentence about "persons qualified through training and experience." Another point would be the inclusion of the temporal element; by and large, the longer the involvement of the therapist with the patient, the more successful (within limits) the outcome should be. (This is a translation of the simple dose-effect relationship familiar to researchers into psychopharmacology.)

We are now in a position to look at the results of the large-scale comparisons carried out by Smith et al. using the methods of meta-analysis. The overall effects are presented in the form of a table on page 89, which lists 18 types of therapy, compared with no-treatment control groups. The effects are listed in terms of average effect size (ES) (i.e., listed in standard terms). One notes that the average effect over the 18 different types of treatment is 0.85, which, as the authors argue, is not negligible. One also notices, however, that one of the 18 types of therapy is

placebo treatment! Now this is a curious inclusion. Placebo treatment can hardly be identified as meaning "the informed and planful application of techniques derived from established psychological principles, with the intention of assisting individuals to modify such personal characteristics as feelings, values, attitudes, and behaviours which are judged by the therapist to be maladaptive or maladjustive"! The intenton of placebo treatment, and its raison d'être, are simply to provide a semblance of a treatment that, according to psychological principles, would have no effect in assisting individuals to modify their characteristics! Indeed, placebo treatment is the proper control that should be used in making comparisons with types of therapy. How it came to be listed as a type of therapy is consequently unintelligible. This indicates the subjective way in which the authors have allocated their data—a degree of arbitrariness and subjectivity that goes well beyond that of any of the authors they criticize for these very faults in the course of their book.

Let us consider for a moment the effect of using placebo treatment as a control, as of course it is usually taken to be. Psychodynamic therapy has an average ES of 0.69; placebo treatment has an average ES of 0.56. Subtracting the latter from the former gives an average ES for psychodynamic therapy of 0.13, so small as to be completely negligible. And this despite the fact that we are comparing 200 placebo treatment effects with 108 psychodynamic therapy effects, a number of instances large enough to indicate an appropriate degree of effectiveness, if such existed.

Thus, using placebo treatment as the proper control (which it undoubtedly is), we find that the alleged effectiveness of psychodynamic therapy almost vanishes. The positive outcome reported by Smith, Glass, and Miller is merely a consequence of their sleight of hand in using placebo treatment, not as a control, but as a type of therapy. They do not even defend this inadmissible intrusion; they simply assume that the reader will not be watchful enough to discover it.

Let us now turn to the no-treatment control group they use instead of the proper placebo-treatment control. They have simply averaged all the no-treatment groups in order to obtain a figure with which to compare the treatment effectiveness, but on theoretical grounds this is inadmissible. There are essentially two types of notreatment groups. Those with which I was concerned in my original study of the effectiveness of psychotherapy (Eysenck, 1952) were neurotics who received no psychiatric or psychological treatment during any stage of their illness. Many of the patients included in the Smith, Glass, and Miller book were waiting-list patients; i.e., patients who for the purpose of the experiment were kept on a waiting list (or might be on a waiting list because there were not enough therapists to go round) and who were told explicitly that within a relatively short period of time they would receive treatment.

Psychologically these two groups of patients are in an entirely different situation. The former, having no hope of therapy, are likely to seek help from friends, family members, teachers, priests, or quacks. In other words, they discuss their problems with all sorts of people who although mostly untrained, nevertheless provide a certain amount of support. They help reduce anxiety, foster relaxation, and work to counter the conditioned emotional responses that, according to the conditioning theory of neurosis, make up the symptomatology of the neurotic illness (Eysenck, 1979, 1982). People on a waiting list, assured that their troubles will be dealt with within a relatively short time, are less likely to talk about their troubles with lay people. Consequently, they are less likely to receive the reduction in anxiety, the relaxation, and the conditioned-response extinction that follow such discussions. Thus I would consider it inadmissible to take an average over different types of (no-treatment) controls, as Smith, Glass, and Miller do; they should be strictly separated and assessed according to the type of control.

Smith, Glass, and Miller may of course argue that this is a theoretical point that they do not want to consider; that because it is theoretical there is little evidence in its favor (although it is in agreement with what is known about the fate of patients who cannot, for some reason or other, obtain psychiatric treatment); and that they are entitled to disregard it. Such a reply would illustrate the difference between their approach and mine. It seems meaningless to me to aggregate two groups that differ so profoundly in the way they have been treated, in the hopes that are held out for them, in the position in which they find themselves, and in the expectations they have for the future. Smith, Glass, and Miller simply do not care. Indeed, the point never seems to have struck them, since they make no mention of these vital differences. One would think that this confusion invalidates their whole set of comparisons, particularly as they claim that their work contradicts my own earlier summaries of the literature. The results I was summarizing dealt with *true* no-treatment groups; it is the later studies that included waiting groups, the studies that appear to give more positive results.

What I see emerging from my approach, and missing from the Smith, Glass, and Miller approach, is the growth of a hypothesis that can be empirically tested (waiting-group patients behave differently from no-treatment patients during the no-treatment period), with the corollary that the different behaviors of these groups are relevant to the amount of spontaneous remission they may show. Until and unless this hypothesis is tested, the easy adoption by Smith, Glass, and Miller of the view that it doesn't make any difference undermines one's confidence in their results. This is additional to the point made before, namely that properly speaking no-treatment (spontaneous remission) controls are much less appropriate than are placebo-treatment controls, for the simple reason that if we are concerned with specific effects of a different type of treatment, then placebo effects must be taken into account. None of these important points is discussed by Smith et al., making it apparent that their meta-analysis is not a scientific exercise in any meaningful sense of the term. It is merely evaluative in a way that makes evaluation pointless.

SOME CONSEQUENCES OF META-ANALYSIS

Let us consider some further results of the meta-analysis performed by Smith, Glass, and Miller. As pointed out above, the definition of psychotherapy accepted by them requires, among other things, that the duration of therapy be related to its effectiveness. Averaging over 1,735 ES measures, the authors find a correlation with duration of therapy of -.05! In effect, they reveal there is no relationship between duration and effectiveness of therapy! Since the duration of therapy bore ranged from 1 hour to over 300 hours, the finding that "the effect of therapy bore

no simple or consistent relationship to its duration" (p. 115) is astounding. Smith et al. argue that "the lower effect sizes in the therapies of extremely long duration should be viewed in the light of the diagnosis and severity of the problems that clients of long-term therapies probably present" (p. 116). This can of course be argued, but it represents exactly the type of subjectivity that the same authors try to eradicate from the discussion of results summaries. My own interpretation would be the opposite. Longest in duration are usually psychodynamic and psychoanalytic investigations. These clients are carefully sifted in such a way that the more serious cases are excluded, while the most hopeful (intelligence, education, and high socioeconomic status) are included. Thus one would expect precisely the long-continued cases to be suffering from less serious illnesses and to do correspondingly better. I do not wish to insist on the correctness of this interpretation. It merely illustrates the countless possibilities of introducing subjectivity into the assessment of meta-analyses, the very subjectivity that the authors hope to exclude through their particular method of statistical summary. Such a hope is illusory.

Therapist's experience is considered a critical variable by practically all therapists, regardless of their persuasion or general theory. In the Smith, Glass, and Miller study the correlation over 1,637 ES measures is exactly .00! The authors have little to say about this astonishing finding, which renders absurd their whole claim to have demonstrated the effectiveness of psychotherapy. Psychotherapy of any kind applies techniques based on certain theories. These theories demand not only that there be a correlation between success and length of treatment, but also that the training and experience of the therapist are crucial. To find that neither corollary is in fact borne out is a death blow to any claims of having demonstrated the effectiveness of psychotherapy. These experimenters may have demonstrated something, but that something must be entirely different from what is usually conceived to be psychotherapy of any kind. They argue that there is a confounding between the types of training of the experimenters with the different types of problem addressed by professionals in the various fields. This is possible, but again it is an attempt to argue along subjective lines, not a scientific determination of falsifiable facts. They may be right in what they say, but certainly the meta-analysis, for all its vaunted virtues, does not bring it out. Nor is it likely that these alleged confoundings could by any stretch of the imagination have obliterated any effects of therapist's training and experience (see Strupp & Hadley, 1979).

In assessing the effectiveness of the different kinds of therapy, and of therapy as a whole, Smith, Glass, and Miller pay little attention to the marked decline of effectiveness over time when therapy is completed. This decline (shown in their Figure 5-2) is in fact grossly underestimated, because the evidence on spontaneous remission would suggest an improvement in the condition of the patient, not a decline. Thus we could possibly conclude from their data that there is a relatively slight effect of psychotherapy, as compared with placebo controls, rapidly vanishing over time. This tiny, evanescent effect, which is irrespective of duration of treatment or training and experience of therapist, is far removed from claims made by the authors that meta-analysis has enabled them to show strong and large-scale effects of psychotherapy! As they put it, "The evidence overwhelmingly supports the efficacy of psychotherapy" (p. 183). According to them, psychotherapy is benefi-

cial, consistently so and in many different ways. Its benefits are on a par with other expensive and ambitious interventions, such as schooling and medicine. As regards "the post hoc rationalisations of academic critics of the psychotherapy-out-come literature" (p. 183), these "have nearly been exhausted. They can scarcely advance new excuses without feeling embarrassed, or without raising suspicions about their motives" (p. 183).

Thus on the basis of these fallacious and badly calculated data, the authors find small, evanescent effects that contradict in detail all the assumptions and predictions of practicing psychotherapists. They advance grandiose claims that find no support in their own work, such as: "Psychotherapy benefits people of all ages as reliably as schooling educates them, medicine cures them, or business turns a profit" (p. 183). Indeed, apparently "we are suggesting no less than that psychotherapists have a legitimate, though not exclusive, claim, substantiated by controlled research, in those roles in society, whether privately or publicly endowed, whose responsibility is to restore to health the sick, the suffering, the alienated, and the disaffected" (p. 184). The claims in this sentence are as unsubstantiated by the research reviewed as its grammar is suspect and its syntax faulty. The methodology is inadequate to give rise to any such far-reaching conclusions, and even on their own showing the results are far more modest that would be suggested by these claims.

THE SUBJECTIVITY OF META-ANALYSIS

We have laid particular stress on the subjectivity of the judgments so frequently made by Smith, Glass, and Miller. One item in particular may be singled out to illustrate this reliance on subjective argument: the relative effectiveness of behavioral versus verbal therapies. On global adjustment, the average ES of verbal therapy is 0.78, for behavioral therapy 0.98. Smith, Glass, and Miller admit, "Behavioural therapies reveal larger average effects for measures of global adjustment" (p. 99). The superiority of behavioral over verbal methods was apparent with different types of patients, ranging from the neurotic and true phobic to the psychotic and depressive. Such a conclusion, going counter to many popular prejudices, demands some form of reactivity by the authors, and this is supplied by a particular application of subjective judgment relating, appropriately enough, to a "correction for reactivity." What is reactivity? Smith, Glass, and Miller explain:

In other words, the measures used to assess outcome can be graded in terms of reactivity from 1 (lowest) to 5 (highest). The value 1 is given to physiological measures such as PSI, pulse rate, GSR, or objective tests like grade-point average. At the highest end, Smith, Glass, and Miller group therapist rating of improvement of symptoms, projective tests (nonblind), behavior in the presence of therapist or nonblind evaluator (e.g., Behavioral Approach Test), and instruments that have a direct and obvious relationship with treatment (e.g., where desensitization

Highly *reactive* instruments are those that reveal or closely parallel the obvious goals or valued outcomes of the therapist or experimenter; which are under control of the therapist, who has an acknowledged interest in achieving predetermined goals; or which are subject to the client's needs and ability to alter his scores to show more or less change than what actually took place. (p. 67)

hierarchy items were taken directly from measuring instrument). Using this scale, the experimenters corrected the observed data for reactivity, and found that by doing so they could eliminate the apparent superiority of the behavioral over the verbal methods of treatment!

But of course this effect is achieved very simply by a subjective sleight of hand, namely giving behavioral measures the highest reactivity scores, scores equal to those of the therapist himself rating the improvement he notices in patients. Let us consider what kind of behavioral measure is here given the highest reactivity score. For example, the Behavioral Approach Test is used in the case of monosymptomatic phobics. The phobic object is placed in front of patients and they are invited to approach it as closely as they can; the distance of the object to the point of nearest approach is the score. Another measure is the point reached in a desensitization hierarchy that the patient successfully passes; i.e., the most anxiety-producing stimulus he can cope with after treatment.

Most observers, particularly those (unlike Smith, Glass, and Miller) who have practical experience of the methods used and the reactions of patients, would rate these measures near the bottom of the scale, receiving reactivity values of 1 or 2rather than at the top. The decision of Smith et al. is a purely subjective one; it is not even argued for or defended. The decision seems to have been made in order to explain away the obvious superiority of the behavioral over the verbal methods. I will not here argue the case, as such arguments would be pointless. I do not say that I am necessarily right and Smith, Glass, and Miller wrong in the respective reactivity values we assign to these methods of measurement. I merely point out that their method of analysis does not rule out subjectivity in any way. The decision to give behavioral methods of assessment the highest reactivity value is not based on factual evidence; there is no degree of objectivity involved. It is frankly subjective and yet determines the final result of the allegedly objective meta-analysis. Had the behavioral methods of assessment been given a value of 1, as I would suggest, rather than the 5 of Smith, Glass, and Miller, the difference between behavioral and verbal methods would have been even greater than the raw scores suggest. Thus the *whole* outcome of the analysis is determined by subjective judgment, contrary to their claims for the objectivity of meta-analysis. Here as elsewhere there is nothing objective about the procedure. The findings simply gloss over the subjectivity that would be glaringly apparent had it not been hidden behind a statistical mirage. What is true in their assessment of reactivity values is equally true of many other parts of the book: The authors' subjective assessments and evaluations determine the allegedly objective outcome of the analysis.

THE INCLUSIVENESS OF META-ANALYSIS

Let us now turn to my last point about Smith et al. There are of course many other criticisms that could be made of their book, but here I have focused on their approach to meta-analysis as such, the present topic. I have singled out the aspects that most clearly demonstrate the erroneousness of the claims made by the authors for their methodology. Prime among these is the claim that their method substitutes rigor and objectivity for subjectivity and arbitrariness in the selection of studies to

be included. They point out that most previous reviewers of therapy effects have ruled out certain studies as being poor in design and quality. They continue:

The strategy of ex post facto impeachment of some studies based on design, quality and outcome measurement is unsupportable. This strategy presumes an objectivity and distance from the problem that is rare among acknowledged advocates and adversaries. No study is above criticism. All studies vary on a number of dimensions of quality and rigour. Where any reviewer draws a line—assigning a study the status of acceptable or unacceptable—is purely an exercise in professional judgement. Any judgemental strategy permits the introduction of bias in the conclusions. (p. 19)

They return to this charge again and again, proclaiming the superiority of overall inclusiveness, which they claim to have practiced. We may now with advantage look at the difficulties that arise in such a context, and try to estimate the degree to which meta-analysis has been successful where others, apparently, have failed.

Rachman and Wilson (1980) discuss in detail the success or failure of Smith and Glass (1977) in this respect. They begin by quoting Cooper (1979), who lays down the rules that warrant the application of meta-analytic evaluations of groups of studies. He believes it is only appropriate when the studies in question "(a) share a common conceptual hypothesis or (b) they share operations for the realisation of the independent or dependent variables, regardless of conceptual focus" (p. 133). They comment:

Neither is currently true of the therapy outcome literature. The spreading quality of experimental design, in adequacy of measurement, and in conceptual focus is so great that it tends to integrate findings across such widely divergent research are doomed to muddy the troubled waters still further. Confusion, not clarification is the result. Sacrificing quality to quantity is a misconceived exercise. (p. 250)

Going back to the Smith and Glass (1977) paper, we may consider their coverage of behavior therapy. Of the 75 comparative outcomes studies reviewed by Kazdin and Wilson (1978), only 26, or 35%, are included in the Smith and Glass bibliography. Of the 24 studies on the treatment of delinquency, enuresis, conduct problems, retardation, marital conflict, depression, and hypertension included in the Kazdin and Wilson review, only 1 is included by Smith and Glass!

Rachman and Wilson point out:

The massive omission of key behavioural studies is puzzling. Anyone familiar with the behaviour therapy literature over the past decade will not fail to notice the absence in Smith & Glass's metaanalysis of many studies by prolific and prominent researchers such as Azrin and O'Leary, to name only two. Only one of the several important studies on the treatment of fearful subjects carried out by Bandura and his colleagues is included, despite their importance. Similarly, only one of Lang's several key studies on systematic desensitisation is included. There is scant trace of recent behavioural research on the treatment of obsessional-compulsive disorders. The clinical research studies of British investigators such as Gelder, Mathews, Bancroft and Marks are largely overlooked. (p. 251)

Nor would it be true that the omission of numerous studies of direct relevance to therapy outcome is confined to behavior therapy. Many of the studies Luborsky, Singer, and Luborsky (1975) included in their review find no place in Smith and Glass's meta-analysis. For example, of the 24 studies Luborsky et al. reviewed, 9 are omitted from the Smith and Glass bibliography. Thus the use of meta-analysis does imply selection on a surprising scale, since the authors had before them the work of others whom they accuse of overselectivity! Their own selectivity seems even more arbitrary and unaccountable than that of those they criticize.

Rachman and Wilson go on to say that "a second major problem with the studies in the Smith & Glass meta-analysis is that a number of them are so flawed as to put them beyond the pale of acceptability" (p. 252). They give a number of examples. One is a report by Barendregt (1957), in which the outcome depends on three specially constructed Rorschach measures of unknown status. These measures cannot possibly give a meaningful estimate of the degree of therapeutic change, and the author himself would not be likely to cite the result as evidence on which to base any conclusions, positive or negative, about the effects of psychotherapy. The report by Cooper (1963) is a retrospective comparison among the effects of various forms of behavior therapy and conventional psychiatric care. The "control group" (if that is the correct term) was compiled retrospectively, making impossible any form of random assignment. The assessment of treatment effects was determined by inspection of the case notes. As Rachman and Wilson comment, "So far from using random assignment, the patients in the one group were specially selected. largely on the grounds that they had failed to respond to any form of early treatment" (p. 252). Yet while this poorly designed study is included, the carefully planned and executed study of Gelder et al. (1973) on the same topic is excluded! This is selectivity with a vengeance.

As puzzling as the exclusion of certain excellent studies is the inclusion of certain irrelevant ones. For example, included is a report by Sheldon (1964) that describes the effects of community after-care services in dealing with former psychiatric patients. Nowhere in this report does the word "psychotherapy" appear, nor does the term "counseling"! Equally strange is the inclusion of an article on transactional analysis as a means of teaching writing to high-school pupils (Beckstrand, 1973)! What this latter has to do with the effects of psychotherapy on mental illness is impossible to fathom.

It would be idle to continue listing the sins against inclusivity committed by the proponents of this false god. They are guilty of the same sin in both their book and their article, in spite of their strictures against this very practice. Note that the objection is not to selective exclusion and inclusion; as I have argued before, judgmental processes must play a part in any review of a complex, diversified, and unsatisfactory field such as this. The objection is that their choices are the wrong choices on any reasonable estimate of quality or relevance, and that their selectivity contradicts their own beliefs in the importance of nonselectivity. Thus Smith, Glass, and Miller fail on both counts, demonstrating again that while the rhetoric of meta-analysis may sound appealing, its practice differs in no way from the unsatisfactory practices castigated by the authors themselves.

Many other criticisms can be found in Rachman and Wilson (1980) or a paper by the present author (Eysenck, in press). Here we merely note that promise and practice of meta-analysis are quite different things. Whether the rules for inclusion and exclusion are set tight or loose, subjective decisions must be made. The scientist must use his judgment; the judgment may be correct or incorrect but it is not infallible. The pretense that meta-analysis absolves the scientist from the duty of making judgments, or makes his judgments infallible, is simply untrue. Hence what appears to be the most important and promising aspect of meta-analysis is fallacious.

GENERAL CONCLUSIONS

Having looked in some detail at one example of meta-analysis, we return to a more general discussion of the pros and cons of this methodology. If in this paper we have concentrated on the cons, this is not to say that we are oblivious to the pros. If and when the conditions laid down by Cooper (1979) obtain in a given field, meta-analysis may add a slight degree of objectivity and interest to a review of the research in that field. Unfortunately the technique has been used mainly in fields where these conditions do not obtain. This is particularly unfortunate in a field where, as Kuhn would phrase it, one paradigm is being replaced by another. In the psychotherapeutic field we have for many years followed the Freudian paradigm in one form or another. In the last 20 years or so this has been increasingly replaced by what we may perhaps call the Pavlovian paradigm. This inevitably leads to ardent debates about not only the nature and origin of neurosis but also the roper measurement of the dependent variable. If behaviorists are concerned mainly with behavioral and autonomic consequences of therapy, and the psychoanalyst with cognitive consequences and changes in subconscious dynamics, then it becomes difficult to compare outcome researches that become incommensurate with each other. These are theoretical problems that must be tackled, but they are outside the scope of the Smith, Glass, and Miller type of analysis. They assume, apparently unconsciously, what Kiesler (1966) has called the "uniformity assumption myth" of psychotherapy research and evaluation. As Rachman and Wilson (1980) point out:

This, then, is another problem that will not go away because we have the statistical facility of averaging disparate and unrelated measures in a meaningless fashion. Meta-analysis always tries to run away from the problems that arise in a given field and to overcome them by disregarding them. This is not the path that science advocates. When we have problems, we must try to solve them, theoretically, or experimentally, or preferably along both lines. We cannot blithely assume that the problems will go away simply because we can add statistically figures bearing no relation to one another. Adding apples and oranges may be a a pastime for children learning to count, but unless we are willing to disregard the differences between these two kinds of fruit, the result will be meaningless. This, indeed, is the gravamen of my charge against meta-analysis. It may have its uses in certain

Nowhere is this more damaging than with respect to measures of therapy outcome. The type of outcome measures incorporated in Smith & Glass's all-encompassing meta-analysis runs the gamut, including disparate measures of anxiety, self-esteem, work or school achievement, physiological stress, projective tests, and the lot. Smith & Glass defend the strategy of mixing different outcome measures together by arguing that "all outcome measures are more or less related to 'well-being' and so at a general level are comparable" (p. 753).

So much for the increasing mass of data demonstrating the importance of obtaining multiple measures of outcome due to the well-established fact that different treatments may have different effects on different measures, whether or not one is assessing anxiety (e.g., Rachman & Hodgson, 1974), sexual responsiveness (e.g., Wincze, Hoon, & Hoon, 1978), pain (e.g., Philips, 1978) or what have you (see also Ciminero, Calhoun, & Adams, 1977; Mischel, 1968, 1977). (p. 253)

circumstances (particularly when there is so much agreement that the use of metaanalysis is unnecessary in any case!). In the majority of cases where it is being used, however, it muddles the waters, disregards the problems, and leads to meaningless conclusions that are likely to hamper proper scientific research.

References

- Barendregt, J. T. A psychological investigation of the effect of group psychotherapy in patients with bronchial asthma. *Journal* of Psychosomatic Research, 1957, 2, 115-119.
- Bartley, S. H. Principles of perception. New York: Harper, 1958.
- Beckstrand, P. E. TA as a means of teaching writing in high school. *Transactional Analy*sis Journal, 1973, 3, 161-163.
- Ciminero, A. R., Calhoun, K. S., & Adams, H. E. (Eds.) Handbook of behavioural assessment. New York: Wiley, 1977.
- Cooper, H. M. Statistically combining independent studies: A meta-analysis of sex differences in conformity research. *Journal of Personality and Social Psychology*, 1979, 37, 131-146.
- Cooper, J. E. A study of behaviour therapy in 30 psychiatric patients. *Lancet*, 1963, *1*, 411-415.
- Cooper, J. E., Gelder, M. G., & Marks, I. M. Results of behaviour therapy in 77 psychiatric patients. *British Medical Journal*, 1965, 1, 1222.
- Eysenck, H. J. The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, 1952, *16*, 319-324.
- Eysenck, H. J. The biological basis of personality. Springfield, Ohio: Thomas, 1967.
- Eysenck, H. J. (Ed.). The measurement of personality. Lancaster, England: Medical and Technical Publishers, 1976.
- Eysenck, H. J. An exercise in megasilliness. American Psychologist, 1978, 33, 517.
- Eysenck, H. J. The conditioning model of neurosis. *The Behavioural and Brain Sciences*, 1979, 2, 155-199.
- Eysenck, H. J. (Ed.). A model for personality. New York: Springer, 1981.
- Eysenck, H. J. Neo-behavioristic (S-R) theory. In G. T. Wilson & C. M. Frank

(Eds.), Contemporary behavior therapy. New York: Guilford Press, 1982.

- Eysenck, H. J. Special review: The benefits of psychotherapy. *Behaviour Research & Therapy*, in press.
- Eysenck, H. J., & Frith, C. Reminiscence, motivation and personality. New York: Plenum, 1977.
- Gelder, M. G., Bancroft, J. H. J., Gath, D., Johnston, D. W., Mathews, A. M., & Shaw, P. M. Specific and non-specific factors in behaviour therapy. *British Journal of Psychiatry*, 1973, 123, 445-462.
- Gilbert, G. M. Intersensory facilitation and inhibition. *Journal of General Psychology*, 1941, 24, 381-407.
- Glass, G. V. Primary, secondary, and metaanalysis of research. *Educational Research*, 1976, 5, 3-8.
- Glass, G. V. Integrating findings: The meta-analysis of research. In L. Shulman (Ed.), *Review of Research in Education*, 5, 59-78. Itasca, Ill.: Peacock, 1977.
- Glass, G. V., & Smith, M. L. Meta-analysis of research on the relationship of classsize and achievement. *Educational Evaluation and Policy Analysis*, 1979, 1, 2-16.
- Harris, J. D. Some relations between vision and audition. Springfield, Ill.: Thomas, 1950.
- Hartman, G. W. Gestalt psychology. New York: Ronald Press, 1935.
- Kazdin, A. E., & Wilson, G. T. Evaluation of behaviour therapy: Issues, evidence and research strategies. Cambridge, Mass.: Ballinger, 1978.
- Kiesler, D. J. Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin*, 1966, 65, 110-136.
- Kravkov, S.V. Interaction of the sense organs. In A. Leontyev et al. (Eds.), Psychological research in the USSR, I. Moscow: Progress Publications, 1966.
- London, I. D. Research on sensory interac-

tion in the Soviet Union. Psychological Bulletin, 1954, 51, 531-568.

- Loveless, N. E., Brebner, J., & Hamilton, P. Bisensory presentation of information. *Psychological Bulletin*, 1970, 73, 161-199.
- Luborsky, L., Singer, B., & Luborsky, L. Comparative studies of psychotherapies: Is it true that everyone has won and all must have prizes? *Archives of General Psychiatry*, 1975, *32*, 995-1008.
- Maruyama, K. Intersensory effects in vision and audition. *Japanese Journal of Psy*chology, 1964, 35, 204-216. (In Japanese)
- Meltzoff, J., & Kornreich, M. Research in psychotherapy. New York: Atherton, 1970.
- Mischel, W. Personality and assessment. New York: Wiley, 1968.
- Mischel, W. On the future of personality measurement. *American Psychologist*, 1977, 32, 246-254.
- Philips, C. Tension headache: Theoretical problems. Behaviour Research & Therapy, 1978, 16, 249-262.
- Rachman, S., & Hodgson, R. J. Synchrony and disynchrony in fear and avoidance. *Behaviour Research and Therapy*, 1974, 12, 311-318.
- Rachman, S. J., & Wilson, G. T. The effects of psychological therapy. Oxford & New York: Pergamon, 1980.
- Ryan, T. A. Intercorrelation of the sensory systems in perception. *Psychological Bulletin*, 1940, 37, 659–698.
- Sheldon, A. An evaluation of psychiatric aftercare. British Journal of Psychiatry, 1964, 110, 662-667.
- Shigehisa, T. An investigation of intersensory effects. Unpublished doctoral dissertation, University of Aberdeen, 1972.

- Shigehisa, T., & Symons, J. R. Effect of intensity of visual stimulation on auditory sensitivity in relation to personality. *British Journal of Psychology*, 1973, 64, 205-213. (a)
- Shigehisa, T., & Symons, J. R. Reliability of auditory responses under increasing intensity of visual stimulation in relation to personality. *British Journal of Psychology*, 1973, 64, 375-381. (b)
- Shigehisa, T., & Symons, J. R. Effect of intensity of visual stimulation on auditory sensitivity in relation to personality. In H. J. Eysenck (Ed.), *The measurement of personality*. London: MTP Press, 1976.
- Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. *Ameri*can Psychologist, 1977, 32, 752-760.
- Smith, M. L., Glass, G. V., & Miller, T. I. The benefits of psychotherapy. Baltimore & London: The Johns Hopkins University -Press, 1980.
- Stern, W. Allgemeine Psychologie auf personalistischen Grundlage. The Hague: Nijhoft, 1935.
- Strupp, H. H., & Hadley, S. W. Specific vs. nonspecific factors in psychotherapy. Archives of General Psychiatry, 1979, 36, 1125– 1136.
- Symons, J. R. An investigation of intersensory relationships. Unpublished doctoral dissertation, University of Reading, 1954.
- Torii, N. Sensory interaction between vision and audition. *Bulletin of the Japanese Research Centre*, 1962, 1, 26-37. (In Japanese)
- Wincze, J. P., Hoon, E. F., & Hoon, P. W. Multiple measure analysis of women experiencing low sexual arousal. *Behaviour Research & Therapy*, 1978, 16, 43-50.