

The Scientific Review of Mental Health Practice

Objective Investigations of Controversial and Unorthodox Claims in Clinical Psychology, Psychiatry, and Social Work

SRMHP Home / Spring ~ Summer 2002 Volume 1 Number 1 /

Dodo Bird, Phoenix, or Urban Legend?

The Question of Psychotherapy Equivalence

Authors:

John Hunsley - School of Psychology, University of Ottawa

Gina Di Giulio - School of Psychology, University of Ottawa

Author Note:

Correspondence concerning this article should be addressed to John Hunsley, School of Psychology, University of Ottawa, Ottawa, Ontario, K1N 6N5; E-mail: hunch@uottawa.ca.

Abstract:

For over 60 years, claims have been made about the general equivalence of all forms of psychotherapy. In the past 2 decades, numerous meta-analyses have been published that bear on the question of psychotherapeutic equivalence (often referred to as the Dodo bird verdict). In this article we critically review meta-analytic work most relevant to this question and, based on our review, conclude that there is overwhelming evidence that the Dodo bird verdict is incorrect. Indeed, with few exceptions, all meta-analytic evidence points to substantial differences among psychological treatments, especially when comparing cognitive-behavioral treatments (including cognitive and behavioral interventions) with other forms of therapy. We discuss the implications of this evidence for current efforts to promote evidence-based psychotherapeutic practices.

Based on hundreds of randomized control trials over the past 40 years, the clear indication is that psychotherapy is generally effective in alleviating the distress and dysfunction associated with a wide range of aversive psychological conditions [1] (e.g., Lipsey & Wilson, 1993; Smith, Glass, & Miller, 1980). Although it is important to know this fact for both professional and public-health reasons (i.e., there is merit in training individuals to provide psychotherapeutic services and there are treatments that can be expected to help many people who suffer from problematic psychological conditions), such a statement is relatively unenlightening, for it is akin to saying that surgery works or that antibiotics are effective. For most health professionals this state of affairs would seem to beg the obvious questions of (1) what, precisely, is the "it" that is effective and (2) for what symptoms, diagnoses, disorders, problems, or concerns, specifically, is "it" effective?

One might imagine that, for practitioners and psychotherapy researchers alike, the search for optimal treatments for the various kinds of conditions encountered in clinical practice would be a very high priority. Indeed, a substantial body of research over the past two decades has addressed this

question (cf. Nathan, Stuart, & Dolan, 2000), and has led to the current focus on evidence-based practice in psychotherapy (e.g., Nathan, Gorman, & Salkind, 1999; Roth & Fonagy, 1996) and, also, to one particular version of evidence-based practice known as empirically supported treatments (ESTs; Task Force on Promotion and Dissemination of Psychological Procedures, 1995). Although acknowledging the importance of general therapeutic skills in establishing and maintaining a supportive and collaborative therapeutic environment, the assumptions at the core of evidence-based treatment are that (1) a specific set of techniques and therapeutic skills may be necessary for the optimal treatment of a given condition and (2) research focused on specific treatments for specific conditions is necessary to determine which treatment or treatments may be most effective.

Even as efforts continue to establish evidence-based psychotherapeutic practice in a number of countries (Andrews, 2000; Chambless & Ollendick, 2001; Hunsley & Johnston, 2000; Roth & Fonagy, 1996; Schulte & Hahlweg, 2000), a substantial number of informed psychotherapy researchers and clinicians consistently and confidently proclaim that there is no convincing evidence that different treatments are differentially effective and, furthermore, that the majority of evidence demonstrates the equivalence of all psychotherapies (e.g., Lambert & Bergin, 1994; Weinberger, 1995). Moreover, implicit in this position is that equivalency is true for all possible types of client conditions.

Such a position, if based on research evidence, could be tenable only within the contextual constraints of the literature. This would mean that only those psychotherapies that have been empirically tested could be assumed to be equivalent, as the subset of therapies that have been evaluated are by no means a representative sample of those therapies offered by practitioners (e.g., Kazdin, 1995). The broader claim—that any treatment provided by a psychotherapist, regardless of the nature of the client's problem or life context, is likely to be as effective as any other possible treatment—is untenable because of the limited range of treatments that have been tested to date. Making this claim would be tantamount to suggesting that, for example, because cognitive therapy has been found to be efficacious in treating depression, any treatment a therapist provides for depression, be it Transactional Analysis, biofeedback, Jungian analysis, or Thought Field Therapy, would also be clinically efficacious. Yet similar claims frequently appear in the professional literature. For example, former American Psychological Association (APA) president Fox (1999), in a newsletter of the APA Practice Directorate, stated that because of the voluminous psychotherapy research base there is no need for further research to examine whether psychotherapy is effective. Moreover, he suggested that the calls for evidence-based treatment are little more than thinly disguised attempts to disenfranchise some mental health professionals by disallowing reimbursement for their services. Similarly, based on their interpretation of the research on psychotherapy process and outcome, Bohart, O'Hara, and Leitner (1998) argued that specific therapeutic techniques are relatively unimportant and that, to provide effective services to a client, one need establish only a therapeutic alliance and then mobilize the client's capacity to resolve problems and distress. Presumably these authors would endorse the view that the same general treatment would be effective for all clients, whether the presenting conditions involve agoraphobia, unresolved bereavement, chemotherapy-induced nausea, eating disorders, or marital distress.

Given the growing attention to evidence-based practice in all forms of health care, the question of whether there is greater support for psychotherapy

equivalence or for psychotherapy specificity has become a central concern for many clinical psychologists, as the answer to this question will greatly influence the types of training provided to clinicians and the types of services that may be deemed to be professionally appropriate. Clear evidence for psychotherapy equivalence would call into question the need for the identification of ESTs, as it could be assumed that any treatment, regardless of condition, would be likely to be effective (Bohart, 2000; Bohart et al., 1998). Therefore, the argument goes, the enterprise of establishing ESTs ignores the body of literature suggesting general equivalence of psychotherapies and undermines the position that psychotherapy qua psychotherapy is "empirically supported" (Elliott, 1998).

The primary purpose of this article is to critically examine the evidence for psychotherapy equivalence. Additionally, given the growing evidence for restricted equivalency (i.e., several different treatments exerting comparable effects on a specific condition), we will discuss the possible scientific and professional implications of psychotherapy equivalence for some specific conditions.

The Dodo Bird Verdict: A Brief Evolutionary Review

Rosenzweig (1936) was the first to use the Dodo bird verdict ("At last the Dodo said, 'Everybody has won, and all must have prizes.'") to describe the hypothesized equivalence of psychotherapies. The quotation comes from Lewis Carroll's *Alice in Wonderland* and the verdict, a not so subtle jab at political committees, was delivered in the context of a competition—a caucus race in which competitors started at different points and ran in different directions for half an hour. Aside from applying Carroll's prose to the context of psychotherapy, Rosenzweig's article is noteworthy for at least two reasons: (1) it was the first time a case was made for psychotherapy equivalence based on what is now known as the "common factors" argument (i.e., factors in the structure and process of treatment that are present in most or all therapies) and (2) the argument for equivalence was made without reference to any supporting data.

The Dodo bird verdict lay dormant for many years when Luborsky, Singer, and Luborsky (1975) concluded, based on a review of comparative treatment research, that there was no evidence of differential treatment effects. In subsequent reviews, Luborsky and colleagues (Luborsky et al., 1993; Luborsky et al., 1999) reiterated this position and provided evidence suggesting that any differential treatment effects may be due to biases in results introduced by the researcher's theoretical orientation (i.e., allegiance to one of the tested therapies, which may result in the preferred treatment being delivered in a more sophisticated and informed manner). Importantly, though, Luborsky et al. (1993) did note that there was evidence of differential treatment effects for a small number of psychological conditions (panic disorder, mild phobias, and schizophrenia).

In the years following the Luborsky et al. (1975) article, attention to the Dodo bird verdict grew, with only a few challenges to it (which we summarize later in this article). For example, while calling psychotherapy equivalence a myth, Beutler (1991) suggested that most therapists and psychotherapy researchers had accepted the Dodo bird verdict as true. Bergin and Garfield (1994), in contrast, concluded that there was "massive" evidence that psychotherapeutic

techniques did not have specific effects. Perhaps most strikingly, in the foreword to Roth and Fonagy's (1996) review of psychotherapy research—a review that listed numerous specific examples of treatment specificity—Shapiro (1996) suggested that there was little evidence that the Dodo bird verdict was incorrect.

It is now commonplace to see sweeping statements about the veracity of the Dodo bird verdict in the literature, with little attention paid to the possible conceptual and methodological constraints on this verdict. Some recent examples include Zinbarg (2000), who wrote that ". . . the well-known 'Dodo bird effect' from meta-analyses of psychotherapy outcome studies suggests that common factors such as the establishment of a sound therapeutic alliance are sufficient for producing at least some degree of improvement" (p. 397), and Polkinghorne and Vernon (2000), who concluded that "[c]ontinuing studies (for example, the classic study of Luborsky, Singer, and Luborsky, 1975; and the more recent study of Wampold et al., 1997) show that various kinds of psychotherapies produce a general equivalence in outcomes" (p. 494). Moreover, reminiscent of Rosenzweig's (1936) article, there are contemporary statements concerning the existence of psychotherapy equivalence that fail to even include references to supporting empirical evidence (e.g., Margison et al., 2000). Presumably, these authors felt that equivalence was so commonly accepted that there was no need to supply supporting citations. Given the ubiquity of the claims for psychotherapy equivalence and the limited attention typically given to the actual research relevant to the claim, there is the real possibility that practitioners and students in mental health fields accept the Dodo bird verdict simply because it appears to be generally and uncritically accepted by others. In conclusion, it appears that the Dodo bird verdict has been welcomed by many and accepted, perhaps reluctantly, by many others. Indeed, given the hopes that some are placing on the evidence for psychotherapy equivalence (or, more accurately, against psychotherapy specificity) and the language used to proclaim the virtues of equivalence, it may be more apt to use the simile of a glorious phoenix rising from the ashes rather than the somewhat less inspiring image of a confused proclamation from an extinct bird.

Psychotherapy Equivalence: Cautions and Evidence

Thus far we have suggested that psychotherapy equivalence has been routinely accepted by many mental health professionals. In this section we focus primarily on the evidence relevant to the Dodo bird verdict. Before considering this evidence, though, it is important to note that a number of authors have raised numerous scientific cautions that must be considered in interpreting any evidence that appears to indicate equivalence among treatments. Issues such as sample size, treatment fidelity, and measurement quality must be closely examined before one can tentatively accept that there may be no true differences between treatments in a given study. Table 1 provides information on a number of conceptual, methodological, and statistical considerations that must be considered in examining the treatment literature (Beutler, 1991; Cujipers, 1998; Hsu, 2000; Norcross, 1995; Reid, 1997; Shadish & Sweeney, 1991; Stiles, Shapiro, & Elliott, 1986).

If one examines the psychotherapy literature without imposing some structure on it (analogous, perhaps, to allowing the competitors in a race to start at different points and run in different directions) it is difficult to ascertain whether there are important differences between therapies. As a start, therefore, one must distinguish between treatment outcome studies and comparative

treatment studies. Treatment outcome studies are experiments in which the impact of a treatment is compared with a control condition in which no services are provided (typically a wait-list control group). In contrast, comparative treatment studies are experiments in which the differential impact of at least two treatments are compared, and a no-treatment control group may or may not be included. Obviously the type of research most relevant to the Dodo bird verdict is the comparative treatment study, as it is a "head-to-head" comparison of treatments drawing on the same sample of clients who have been randomly assigned to treatment. Therefore, in the following sections dealing with the results of meta-analyses relevant to the Dodo bird verdict, we distinguish between meta-analyses of treatment outcome studies and meta-analyses of comparative treatment studies.

In order to obtain some sense of general trends in the results of psychotherapy outcome research, a reliance on meta-analytic data (i.e., the examination of average effect sizes) is necessary given the many hundreds of studies that have been conducted. Except when noted, we report effect sizes as a *d* statistic. Although we have been selective in deciding which meta-analyses to review, we included only meta-analyses that focused on a wide range of psychotherapies and client conditions. It is important to be aware of possible methodological confounds that may affect the conclusions one can draw from meta-analyses. Therefore, based on previous issues raised in the literature, we will attend specifically to treatment categorization, measurement reactivity, and researcher allegiance effects (cf., Lambert & Bergin, 1994) in reviewing these meta-analytic results.

Smith, Glass, and Miller (1980)

The first meta-analysis of psychotherapy was conducted by Smith, Glass, and Miller (1980). Based on hundreds of treatment outcome and comparative treatment studies, they found clear evidence for significant differences among the effects of different "subclasses" of therapy (Table 5-4, p. 94): Cognitive and cognitive-behavioral treatments had the largest effect sizes (mean ES values of 1.31 and 1.24, respectively), followed by behavioral and psychodynamic (0.91 and 0.78) treatments, humanistic treatments (0.63), and, finally, developmental treatments (including vocational-personal development counseling and "undifferentiated counseling"; 0.42). Thus, at the general level, there was clear evidence that these subclasses were far from equivalent. Smith et al. moved on to analyze their data based on client conditions (their term was "diagnostic types"). Again, substantial differences were found among treatment subclasses (Table 5-5, p. 96): For example, the mean effect sizes for behavioral and humanistic treatments in the treatment of depression were 1.18 and 0.50, respectively.

At this juncture, it is important to note that these results are not the ones highlighted by advocates of psychotherapy equivalence. Instead, they tend to focus on analyses conducted on therapy "classes," in which behavioral (mean ES = 0.98) and verbal (mean ES = 0.85) treatments were found to produce comparable effects. In order to understand how apparently large psychotherapy subclass differences disappeared when therapy class was considered, it is essential to know how these classes were constructed. For their categorization system, Smith et al. included cognitive-behavioral, behavior modification, systematic desensitization, and other behavioral treatments in the behavioral class; they included psychodynamic, humanistic, and cognitive treatments in

the verbal class. As the researchers themselves noted, this categorization scheme was arbitrary but defensible (e.g., all behavioral treatments focused primarily on attaining behavioral change). However, given the wide range of effect sizes found for therapy "subclasses," it is difficult to see what could possibly be gained by grouping disparate treatments into therapy "classes." Moreover, it is extremely difficult to imagine that any current categorization scheme would include cognitive therapies with psychodynamic therapies, rather than with behavioral ones, especially as (1) the Association for Advancement of Behavior Therapy (AABT) has suggested for many years that behavioral, cognitive, and cognitive-behavioral treatments are all part of a single family of therapies; (2) the AABT annual convention has increasingly included cognitively oriented presentations (Dobson, Beamish, & Taylor, 1992); (3) cognitive and cognitive-behavioral treatments are typically seen by scholars as part of behavior therapy (e.g., Follette & Hayes, 2000); and (4) most cognitive therapies include numerous behavioral interventions (e.g., Beck, Rush, Shaw, & Emery, 1979). The obvious implication of this argument is that the strongest evidence for the Dodo bird verdict from Smith et al. is based on what is almost certainly a classification error!

The researchers also examined the possible impact of a number of methodological factors on obtained mean ES values. The largest correlation they found was between mean effect size and measurement reactivity ($r = .18$). Measurement of reactivity was coded on a 5-point scale, with physiological measures and blinded ratings treated as less reactive than standardized measures of traits having minimal connection with the treatment, which in turn were coded as less reactive than client self-reports, therapist ratings, and measures that bore a direct relation with aspects of the treatment. Regression analyses were conducted in order to obtain statistically adjusted effect sizes in which measurement reactivity was "equated" across treatment classes. As measurement reactivity was found to be highest in behavioral treatments, this adjustment yielded almost identical adjusted mean ES values for behavioral and verbal classes (0.91 and 0.88, respectively). Although there is certainly a case to be made for making this adjustment, it fails to take into account that (1) behavioral approaches have traditionally viewed client symptoms and problems as samples of the problem to be changed rather than as signs of an underlying problem (cf. Goldfried & Kent, 1972) and (2) the choice of which variables to assess in behavioral treatments is directly related to the client condition and the goals of treatment agreed upon by client and therapist (Mash & Hunsley, 1990).

Smith et al. also conducted analyses on data they obtained from 56 comparative outcome studies of the behavioral and verbal classes of treatment. Even with the classification error described previously, behavioral treatments were significantly superior to the verbal treatments (mean ES values of 0.96 and 0.77; Table 5-14, p. 108). The researchers then adjusted their results for measurement "tractability." The rationale for the coding of tractability is hard to discern, as measures of anxiety, self-esteem, and global adjustment were rated as more tractable than were such measures as emotional-somatic complaints and life adjustment. Based on their analysis, the adjusted difference between the two therapy classes was large for tractable measures (mean ES = 0.25, favoring the behavioral class) but small for less tractable measures (mean ES = 0.04, favoring the behavioral class). Given the importance placed on adjusting for measurement reactivity in the previous analyses conducted by Smith et al., it is curious that no mention was made of why this analysis was not conducted with this subset of 56 studies and why the "tractability" adjustment was used instead.

In conclusion, the influential meta-analysis published by Smith et al. yielded numerous results that do not support a verdict of psychotherapy equivalence. Whether examined by therapy subclasses (i.e., cognitive, cognitive-behavioral, behavioral, psychodynamic, humanistic, and developmental) or by client conditions within therapy subclasses, clear differences among treatment effects were evident. Only by first (mis)classifying cognitive therapies with psychodynamic and humanistic therapies (rather than with behavioral therapies), and then statistically adjusting for supposed measurement problems (largely related directly to distinctions among the therapies regarding what should be assessed in psychotherapy) did the results suggest equivalence across forms of psychotherapy.

Meta-Analyses by Weisz, Weiss, and colleagues

Weisz, Weiss, Alicke, and Klotz (1987) reviewed the child and adolescent treatment outcome literature published between 1958 and 1984 and concluded that there was strong evidence for the superiority of behavioral treatments (including cognitive treatments) over nonbehavioral treatments. Following up on this finding, Weisz, Weiss, Han, Granger, and Morton (1995) conducted a meta-analysis of 150 child and adolescent psychotherapy outcome studies published between 1983 and 1993. The mean effect size for the behavioral treatments (including cognitive, cognitive-behavioral, parent training, operant methods, respondent methods, and social skills training) was .54, which was significantly greater than the mean effect size of .30 for the nonbehavioral treatments (including client-centered and insight-oriented therapies). Weisz and colleagues coded outcome measures based on their similarity to treatment activities; however, they also distinguished among situations in which the similarity was necessary (given the goals of treatment) or unnecessary for the purposes of each study. They then excluded from analysis all effect sizes obtained from measures that were unnecessarily similar to the treatment. In this analysis the behavioral treatments (mean ES = 0.52) were superior to the nonbehavioral treatments (mean ES = 0.25). Even when all effect sizes from outcome measures that were similar to the treatment were eliminated from the analysis, behavioral treatments were still significantly more effective than nonbehavioral treatments (ES values of 0.47 and 0.25, respectively). Importantly, these results essentially replicated their earlier findings. However, neither of these two meta-analytic studies reported analyses in which results were controlled for possible researcher allegiance effects.

To further evaluate the possible biasing effects of measurement issues, Weiss and Weisz (1995) evaluated the relative effectiveness of behavioral (including cognitive) versus nonbehavioral (psychodynamic and humanistic) treatments for children and adolescents in a subset of the studies used by Weisz et al. (1987). This meta-analysis examined 105 studies of treatments for conditions including anxiety disorders, depression, and social skills deficits. The researchers coded the studies for a number of methodological features, including random assignment to treatments, attrition from treatment, therapist experience, degree of rater blindness, participant blindness to outcome assessment, and type of measurement data (with lower values assigned to self-report measures and higher values to assessments of objective behavior and life events). Contrary to expectations, they found that the nonbehavioral treatment studies were actually of higher overall methodological quality. Specifically, they found that behavioral treatments yielded a mean effect size of 0.85, whereas nonbehavioral treatments yielded a mean effect size of 0.42. When the authors controlled for the methodological quality of the studies, the mean effect

sizes of the behavioral and nonbehavioral treatments became 0.86 and 0.38 respectively. The difference was even greater in the 10 comparative treatment studies in their sample that directly compared behavioral and nonbehavioral treatments (mean ES values of .76 and .17, respectively). No analyses controlling for possible researcher allegiance effects were conducted for either the full data set or the restricted set of 10 comparative treatment studies.

In sum, the results of these large-scale meta-analyses of the child and adolescent treatment outcome literature are clear: Cognitive, cognitive-behavioral, and behavioral treatments are significantly superior to humanistic and psychodynamic treatments. This difference is evident even when the results are adjusted for possible measurement concerns and study design quality. Moreover, the superiority of the behavioral family of treatments is evident not only in the general literature but also in comparative treatment studies. Overall, there is no evidence in the child and adolescent treatment literature to support the psychotherapy equivalence position.

Reid (1997)

Thus far, we have reviewed broad meta-analyses that focused on general trends evident in the psychotherapy treatment literature. There have been, however, numerous focused meta-analyses that have examined treatments for such specific conditions as depression, insomnia, smoking cessation, and bulimia. Reid (1997) reviewed the findings from 42 separate such meta-analyses and concluded that 74% showed evidence of differential treatment effects. He noted that behavioral (including cognitive and cognitive-behavioral) treatments have shown clear superiority to other forms of treatment for child maladaptation, child abuse, juvenile delinquency, and panic-agoraphobia. For bulimia and depression, there was evidence that behavioral approaches were superior to other approaches, but in some cases this superiority vanished when investigator allegiance was controlled for statistically. As the majority of meta-analyses reported differential treatment effects, Reid concluded that there was little evidence in the meta-analytic literature to support the Dodo bird verdict. Although he raised questions about the possible effects of allegiance and measurement, he did not systematically examine the impact of these factors on the meta-analytic results.

Wampold, Mondin, Moody, Stich, Benson, and Ahn (1997)

In the most comprehensive and direct test of the Dodo bird verdict, Wampold, Mondin, Moody, Stich, Benson, and Ahn (1997) conducted a meta-analysis that included data from studies that compared at least two bona fide treatments and were published between 1970 and 1995. By eschewing any categorization of the treatments included in their sample, these authors attempted to avoid problems arising from questions about categorization validity. As a result, they simply calculated all ES values between pairs of treatment and then calculated their mean ES in two ways. First, they aggregated all the absolute values of the obtained ES, and divided by the number of ESs. However, they argued that this greatly overestimated the true mean ES for their sample, and so also calculated a mean ES value by randomly assigning a positive or negative sign to each obtained ES and dividing the aggregate of these values by the number of obtained ESs. Wampold et al. reported an average ES of .19 for their first estimate (which was significantly different from a value of zero) and an average ES of .0021 for their second (which was not significant). No attempt was made

to control for allegiance or measurement reactivity in their analyses. Although emphasizing that their results strongly supported the Dodo bird verdict, Wampold and colleagues explicitly cautioned that their results should not be taken as evidence that all practiced psychotherapies are equally efficacious or as efficacious as those included in their sample.

On the face of it, these results would seem to provide solid evidence for psychotherapy equivalence. However, closer attention to this meta-analysis reveals that Wampold and colleagues' data actually provide exceptionally strong evidence for treatment specificity. First, as Crits-Christoph (1997) noted, most studies included in this meta-analysis compared one type of cognitive-behavioral treatment to another cognitive-behavioral treatment (69% by his estimate, closer to 80% by ours). Therefore, even without any further examination of their methodology, the Wampold et al. conclusion of psychotherapy equivalence could only be logically applied primarily to cognitive-behavioral treatments, not to bona fide treatments in general. Second, Howard, Krause, Saunders, and Kopta (1997) argued compellingly that Wampold et al. erred greatly in their ES calculations, as their second method for calculating the mean ES could, by definition, *only* yield a mean value of zero regardless of the true mean ES value. [2] As a result, the correct method for calculating the mean ES was the first one used by Wampold et al., which yielded a significant average ES value of .19 among bona fide treatments. This finding strongly contradicts the Dodo bird verdict, as it indicates that in the most relevant research (i.e., comparative outcome studies with bona fide treatments) there is a meaningful difference among treatments.

In their response to Howard et al., Wampold, Mondin, Moody, and Ahn (1997) attempted to label the difference of .19 standard deviations as small. In our view it is actually rather substantial considering that the typical finding involves the superiority of one efficacious treatment over another efficacious treatment. To put this result into context, it is informative to use Rosenthal and Rubin's (1982) binomial effect size display (BESD) technique, in which an effect size (an *r* value) is expressed as the difference in success rates between two conditions. Converting an effect size *d* value of .19 to *r* yields a value of .094 (for the conversion equation, see Rosenthal & DiMatteo, 2001). Using the BESD technique, this means that in the situation where there are two bona fide treatments, 94 out of every 1,000 clients would experience greater improvement by receiving the significantly more efficacious treatment. Seen in this light, it is obvious that in a clinical context a *d* value of .19 is far from small or unimportant.

Shadish, Matt, Navarro, and Phillips (2000)

A number of commentators have suggested that any apparent superiority of behavioral treatments shown in meta-analyses is at least partly due to the fact that there are a number of behaviorally oriented treatment studies in which neither research participants nor treatments are representative of clients and treatments in the "real-world" (i.e., the participants are less distressed than are clients and the treatments are simplified or abbreviated compared with typical therapeutic practice). Recent evidence from a meta-analysis of 90 studies of clinically representative psychotherapy by Shadish et al. (2000) is directly relevant to this question. These researchers selected studies in which clients, treatments, and therapists were representative of typical clinical settings. In addition to coding a number of design features, they coded for measurement

reactivity (using the criteria developed by Smith et al., 1980) and for measurement specificity (i.e., how similar the measure was to what was done in therapy). Interestingly, given our earlier comments about the importance of measurement specificity in behaviorally oriented treatments, Shadish et al. reported that measurement specificity was statistically distinct from measurement reactivity. This finding suggests that it is critical to separate these two concepts in meta-analytic work. Turning to the results of their meta-analysis, Shadish et al. found overall evidence of significant treatment effects in the studies they sampled (mean ES = 0.41). Using a random-effects model to predict treatment effect sizes, both treatment orientation (behavioral versus nonbehavioral) and measurement specificity (but not reactivity) were significant predictors. In other words, treatment effect sizes were larger for behavioral than for nonbehavioral treatments and were larger when specific measures of outcome were used.

Summary of Meta-Analytic Evidence

In all of the meta-analyses we reviewed, the weight of evidence is clearly and consistently on the side of differential treatment effects (i.e., evidence of treatment specificity). When measurement quality is controlled for and when treatments are appropriately categorized, there is consistent evidence in both treatment outcome and comparative treatment research that cognitive and behavioral treatments are superior to other treatments for a wide range of conditions, in both adult and child samples. Given its prominence in the literature, we wish to underscore our view that Smith and colleagues' (1980) oft-cited conclusion that psychotherapies are equivalent is inaccurate. A fuller evaluation of their meta-analysis and an appreciation of their categorization error can only lead one to the conclusion that there is no compelling evidence for psychotherapy equivalence. Likewise, the more recent meta-analysis of Wampold et al. (1997) provided evidence that contradicts the Dodo bird verdict, and the evidence that they suggested supported this verdict stemmed entirely from an inappropriate and misleading algebraic manipulation.

We also wish to stress that, even if these meta-analyses had supported a claim for psychotherapy equivalence (which they don't), it would be unreasonable and irresponsible to claim that all therapies are equal in their clinical effects. Creative clinicians are always endeavoring to develop more effective and efficient forms of treatment, but the vast majority of these treatments have not been subjected to the type of empirical evaluation that is necessary to determine their true impact on clients. Over the years many therapists have used (and continue to use) treatments such as Bioenergetic Therapy, Neurolinguistic Programming, Transactional Analysis, Thought Field Therapy, and other therapies that are actively promoted within the clinical community. As the treatment research on such therapies is minimal, few studies on these therapies are available for inclusion in any meta-analysis examining the effects of psychotherapy. Accordingly, any meta-analytic results cannot be generalized to these esoteric treatments. Simply put, proponents of these therapies cannot claim clinical legitimacy for their treatments by relying on the results of research conducted on other forms of psychotherapy. Without research evidence for a treatment's effects, the only scientifically appropriate conclusion that can be drawn is *not* that the effects of the treatment are equivalent to other forms of treatment that have been shown to work but, rather, that there is no evidence that the effects of the treatment are greater than would be obtained without treatment.

Another issue that is often raised about the generalizability to clinical contexts of the published treatment research is that research on cognitive-behavioral therapies is most likely to be conducted on clients with more focused and less severe conditions, whereas research on psychodynamic and other therapies is more likely to involve clients whose conditions are more diffuse and severe. Our sense is that although this differential association between client conditions and treatment orientation may have held in the earlier treatment research, it does not accurately characterize the research conducted in the past 15 to 20 years. Speculation aside, it is evident from the various meta-analyses that have used comparative treatment studies (in which the specificity and severity of client condition is comparable between treatments) and from Shadish and colleagues' (2000) research on clinically representative treatment that differential treatment effects do routinely occur. Nevertheless, to fully address this issue, it will be important that psychotherapy investigators continue to conduct research that (1) focuses on underresearched treatments (including the humanistic/ experiential and psychodynamic therapies) and (2) adds to our rather limited knowledge about what treatments work best for clients suffering from chronic and severe conditions (especially those clients meeting criteria for personality disorders).

Few psychotherapy meta-analyses have systematically controlled for research allegiance, which Luborsky and colleagues have consistently argued must be considered in comparative treatment research. It should be noted, however, that controlling for allegiance statistically (by removing variance in outcome variables related to ratings of researcher allegiance) is far inferior to controlling allegiance through having researchers with differing orientations collaborate on comparative outcome research. Statistical controls may also eliminate variance inappropriately, thereby overcorrecting for any researcher bias that may have influenced the results. After all, researcher allegiance may be largely determined by prior demonstrations of the effectiveness of a treatment in the research literature (cf., Reid, 1997; Weisz et al., 1995). Partialing out variance related to allegiance may therefore eliminate variance from the prediction of treatment outcome that is more appropriately apportioned to true treatment effects. Moreover, as different methods of rating researcher allegiance are only modestly correlated (correlations range from .10 to .48; Luborsky et al., 1999), incorporating measures of allegiance in statistical analyses may result in data adjustments that are substantially in error.

Despite problems in measuring allegiance, it is extremely important that the possible moderating effects of allegiance be examined in future research. Luborsky and colleagues have made a strong case for the possibility that the size of any differential treatment effect is greatly diminished when allegiance is considered. This possibility should be considered in future meta-analytic work on psychotherapy outcome, by either statistically controlling for allegiance or, preferably, focusing on studies in which alliance is controlled through collaborative efforts with respect to study design, treatment implementation, and data analysis. Notwithstanding the importance of continued examination of possible allegiance effects, it should also be noted that there is evidence that allegiance may not be related to treatment outcome in some instances and that, at least for cognitive therapy for depression, allegiance effects are no longer commonly found (Gaffan, Tsaoasis, & Kemp-Wheeler, 1995).

In conclusion, when the meta-analytic evidence is critically examined, there is no support whatsoever for the Dodo bird verdict. Psychotherapy equivalence, at least in its broadest form of general equivalence across all therapies, is most

definitely a myth (cf. Beutler, 2000). Viewed in this light, this Dodo bird is clearly not akin to a phoenix, but more closely resembles a repeated, unsubstantiated rumor about the sighting of a bird that has long been extinct. In other words, the Dodo bird verdict is more likely to be an urban legend than a scientifically substantiated position.

Psychotherapy Specificity, Restricted Psychotherapy Equivalence, and Evidence-Based Practice

In their review of the Dodo bird verdict, Stiles, Shapiro, and Elliott (1986) concluded that the search for "winners" among treatments was the wrong direction for psychotherapy research to take. Instead, they argued that much more would be gained by examining differences among techniques as they are used in the process of treatment. Rather than determining the relative superiority of two treatments, they suggested that much more would be learned by comparing techniques that have been proposed to be useful in achieving intermediate, small changes in a session of treatment (for example, comparing a two-chair technique with reflective listening for resolving decisional conflicts).

Like Stiles et al., we suggest that the efforts to determine winners (and losers) in comparative treatment research is largely misplaced. Over three decades ago Kiesler (1966) and Paul (1967) presented influential formulations that emphasized the importance of *not* assuming that all treatments would work for all clients. At the risk of introducing yet another metaphor, it appears to us that, over the past two decades, the Dodo bird verdict has been little more than a red herring in our search for optimal treatments. In our opinion, rather than conducting more comparative trials of psychotherapy, there is much more to be gained by focusing on (1) expanding the list of ESTs that work for specific conditions and (2) improving upon the therapeutic impact of currently available ESTs. As comparative treatment studies will undoubtedly continue to be conducted, we recommend that the researchers designing such studies attend to the issue of research allegiance and the need for replicated results. It is abundantly evident that there are differences among treatments for a number of conditions, but if there is anything to be gained by conducting comparative outcome research it is essential that promising results be replicated. Borrowing from the standards developed for ESTs, evidence for the superiority of one treatment over another is strongest when it has been independently replicated.

The goal of evidence-based practice is to base treatment decisions on the best available scientific evidence. Consistent with the meta-analytic evidence we reviewed, current lists of ESTs are dominated by cognitive-behavioral treatments (Chambless & Ollendick, 2001). In these lists it is evident that cognitive-behavioral treatments should be the treatment of choice for dozens of adult and child conditions. However, the claim by some that certain treatments, especially psychodynamic and experiential ones, cannot be tested with clinical trials (and therefore potentially be listed as ESTs) is unsupportable, as a number of such treatments have been evaluated in this manner (see Johnson, Hunsley, Greenberg, & Schindler, 1999, for the example of emotionally focused couples therapy). Indeed, it is quite conceivable that the dominance of cognitive-behavioral treatments may change in coming years as more and more controlled studies of other treatments are conducted. For example, interpersonal psychotherapy, with its focus on current interpersonal mechanisms (e.g., interpersonal deficits, grief, role disputes, and role transitions) that may be responsible for psychological conditions, is proving to be efficacious for a number of mood disorders, anxiety disorders, and bulimia (Gotlib &

Schraedley, 2000).

In a small number of cases, such as adult depression, several different treatments have sufficient evidence to be considered as first-line options for clients, including several forms of cognitive-behavioral treatment, interpersonal therapy, and brief psychodynamic therapy (Chambless & Ollendick, 2001). Although some may, yet again, try to present this as evidence for treatment equivalence, we suggest that this is the wrong perspective to adopt. Psychotherapies are not equivalent in their theories, techniques, and, for most conditions, treatment outcomes. Attempts to force the issue of psychotherapy equivalence, for all conditions or for any subset of conditions, are misplaced. What is necessary is that people receive treatments that have the greatest likelihood of helping them. As a result, less attention should be paid to ensuring that clinicians, regardless of the type of service they provide, are reimbursed for their services. Instead, we should attend more to (1) training clinicians to provide ESTs and (2) developing strategies to ensure that people in distress have access to clinicians who can provide scientifically supported treatments. For too long we have let the Dodo bird render an inappropriate verdict for a relatively pointless race. We owe it to those who receive our services to move beyond the urban legend of psychotherapy equivalence by ensuring the widest possible availability of scientifically supported psychotherapies.

Notes

1. Throughout this article we use the term "conditions" to refer to the wide constellation of client problems, concerns, symptoms, disorders, and/or diagnoses for which psychological treatment is provided.
2. Assuming a symmetrical distribution (which Wampold et al. did), randomly assigning positive and negative signs to ES values must result in a mean ES value of zero, whether the actual mean ES value is centered around zero or *any other value*. To illustrate this point, consider a distribution centered on a mean ES value of .40. By randomly assigning positive and negative signs to all obtained ES values, half of the values above .40 would become negative and half would remain positive (assuming the assignment process was actually random, which it should be given a sufficient sample of ES values). By summing the resulting values, the result is a near zero value. A similar result will occur when considering the obtained values below mean of .40 in the distribution. Accordingly, when all values are summed and divided by the total number of values, the result should be exceedingly close to zero. Of course, this is an erroneous mean ES value, as we "set" the actual mean ES value at .40.

References

Andrews, G. (2000). A focus on empirically supported outcomes: A commentary on the search for empirically supported treatments. *Clinical Psychology: Science and Practice*, 7, 264-268.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford.

Bergin, A. E., & Garfield, S. L. (1994). Overview, trends, and future issues. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 821-830). New

York: Wiley.

Beutler, L. E. (1991). Have all won and must all have prizes? Revisiting Luborsky et al.'s verdict. *Journal of Consulting and Clinical Psychology, 59*, 226-232.

Beutler, L. E. (2000). David and Goliath: When empirical and clinical standards of practice meet. *American Psychologist, 55*, 997-1007.

Bohart, A. C. (2000). Paradigm clash: Empirically supported treatments versus empirically supported psychotherapy practices. *Psychotherapy Research, 10*, 488-493.

Bohart, A. C., O'Hara, M., & Leitner, L. M. (1998). Empirically violated treatments: Disenfranchisement of humanistic and other psychotherapies. *Psychotherapy Research, 8*, 141-157.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52*, 685-716.

Crits-Christoph, P. (1997). Limitations of the Dodo bird verdict and the role of clinical trials in psychotherapy research: Comment on Wampold et al. (1997). *Psychological Bulletin, 122*, 216-220.

Cujipers, P. (1998). Minimising interventions in the treatment and prevention of depression: Taking the consequences of the Dodo bird verdict. *Journal of Mental Health, 7*, 335-365.

Dobson, K. S., Beamish, M., & Taylor, J. (1992). Advances in behavior therapy: The changing face of AABT conventions. *Behavior Therapy, 23*, 483-491.

Elliott, R. (1998). Editor's introduction: A guide to the empirically supported treatments controversy. *Psychotherapy Research, 8*, 115-125.

Follette, W. C., & Hayes, S. C. (2000). Contemporary behavior therapy. In C. R. Snyder & R. E. Ingram (Eds.), *Handbook of psychological change: Psychotherapy processes & practices for the 21st century* (pp. 381-408). New York: Wiley.

Fox, R. E. (1999). The dark side of evidence-based treatment. *APA Practitioner Focus, 12*(2), 5. Gaffan, E. A., Tsaousis, I., & Kemp-Wheeler, S. M. (1995). Researcher allegiance and meta-analysis: The case of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 63*, 966-980.

Goldfried, M. R., & Kent, R. N. (1972). Traditional versus behavioral assessment: A comparison of methodological and theoretical

assumptions. *Psychological Bulletin*, 77, 409-420.

Gotlib, I. H., & Schraedley, P. K. (2000). Interpersonal psychotherapy. In C. R. Snyder & R. E. Ingram (Eds.), *Handbook of psychological change: Psychotherapy processes & practices for the 21st century* (pp. 258-279). New York: Wiley.

Howard, K. I., Krause, M. S., Saunders, S. M., & Kopta, S. M. (1997). Trials and tribulations in the meta-analysis of treatment differences: Comment on Wampold et al. (1997). *Psychological Bulletin*, 122, 221-225.

Hsu, L. M. (2000). Effects of directionality of significance tests on the bias of accessible effect sizes. *Psychological Methods*, 5, 333-342.

Hunsley, J., & Johnston, C. (2000). The role of empirically supported treatments in evidence-based psychological practice: A Canadian perspective. *Clinical Psychology: Science and Practice*, 7, 269-272.

Johnson, S. M., Hunsley, J., Greenberg, L., & Schindler, D. (1999). Emotionally focused couples therapy: Status and challenges. *Clinical Psychology: Science and Practice*, 6, 67-79.

Kazdin, A. E. (1995). Scope of child and adolescent psychotherapy research: Limited sampling of dysfunctions, treatments, and client characteristics. *Journal of Clinical Child Psychology*, 24, 125-140.

Kiesler, D. J. (1966). Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin*, 65, 110-136.

Lambert, M. J., & Bergin, A. E. (1994). The effectiveness of psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 143-189). New York: Wiley.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181-1209.

Luborsky, L., Diguier, L., Luborsky, E., Singer, B., Dickter, D., & Schmidt, K. A. (1993). The efficacy of dynamic psychotherapies: Is it true that "Everyone has won and all must have prizes"? In M. E. Miller, L. Luborsky, J. P. Barber, & J. P. Docherty (Eds.), *Psychodynamic treatment research: A handbook for clinical practice* (pp. 497-516). New York: Basic Books.

Luborsky, L., Diguier, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., Halperin, G., Bishop, M., Berman, J. S., & Schweizer, E. (1999). The researcher's own therapy allegiance: A "wild card" in comparisons of treatment efficacy. *Clinical*

Psychology: Science and Practice, 6, 95-106.

Luborsky, L., Singer, B., & Luborsky, E. (1975). Comparative studies of psychotherapies: Is it true that "Everybody has won and all must have prizes"? *Archives of General Psychiatry*, 32, 995-1008.

Margison, F. R., Barkham, M., Evans, C., McGrath, G., Clark, J. M., Audin, K., & Connell, J. (2000). Measurement and psychotherapy: Evidence-based practice and practice-based evidence. *British Journal of Psychiatry*, 177, 123-130.

Mash, E. J. & Hunsley, J. (1990). Behavioral assessment: A contemporary approach. In A. S. Bellack, M. Hersen, & A. E. Kazdin (Eds.), *International Handbook of Behavior Modification and Therapy* (2nd ed., pp. 87-106). New York: Plenum.

Nathan, P. E., Gorman, J. M., & Salkind, N. J. (1999). *Treating mental disorders: A guide to what works*. New York: Oxford University Press.

Nathan, P. E., Stuart, S. P., & Dolan, S. L. (2000). Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychological Bulletin*, 126, 964-981.

Norcross, J. C. (1995). Dispelling the Dodo bird verdict and the exclusivity myth in psychotherapy. *Psychotherapy*, 32, 500-504.

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, 31, 109-118.

Polkinghorne, D. E., & Vernon, R. F. (2000). [Book review of *The psychotherapy relationship: Theory, research, and practice*]. *Psychotherapy Research*, 10, 494-496.

Reid, W. J. (1997). Evaluating the Dodo's verdict: Do all interventions have equivalent outcomes? *Social Work Research*, 21, 5-16.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59-82.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166-169.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry*, 6, 412-415.

Roth, A., & Fonagy, P. (1996). *What works for whom? A critical review of psychotherapy research*. New York: Guilford.

Schulte, D., & Hahlweg, K. (2000). A new law for governing psychotherapy for psychologists in Germany: Impact on training and mental health policy. *Clinical Psychology: Science and Practice*, 7, 259-263.

Shadish, W. R., Matt, G. E., Navarro, A. M., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin*, 126, 512-529.

Shadish, W. R., & Sweeney, R. B. (1991). Mediators and moderators in meta-analysis: There's a reason why we don't let Dodo birds tell us which psychotherapies should have prizes. *Journal of Consulting and Clinical Psychology*, 59, 883-893.

Shapiro, D. A. (1996). *What works for whom? A critical review of psychotherapy research* [Foreword] (pp. viii-x). New York: Guilford.

Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.

Stiles, W. B., Shapiro, D. A., & Elliott, R. (1986). Are all psychotherapies equivalent? *American Psychologist*, 41, 165-180.

Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated psychological treatments: Report and recommendations. *The Clinical Psychologist*, 48, 3-23.

Wampold, B. E., Mondin, G. W., Moody, M., & Ahn, H. (1997). The flat earth as a metaphor for the evidence for uniform efficacy of bona fide psychotherapies: Reply to Crits-Christoph (1997) and Howard et al. (1997). *Psychological Bulletin*, 122, 226-230.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "All must have prizes." *Psychological Bulletin*, 122, 203-215.

Weinberger, J. (1995). Common factors aren't so common: The common factors dilemma. *Clinical Psychology: Science and Practice*, 2, 45-69.

Weiss, B., & Weisz, J. R. (1995). Relative effectiveness of behavioral versus nonbehavioral child psychotherapy. *Journal of Clinical and Consulting Psychology*, 63, 317-320.

Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A

meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology*, 55, 542-549.

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*, 117, 450-468.

Zinbarg, R. E. (2000). Comment on "Role of emotion in cognitive-behavior therapy": Some quibbles, a call for greater attention to patient motivation for change, and implications of adopting a hierarchical model of emotion. *Clinical Psychology: Science and Practice*, 7, 394-399.

You can read this article in
***The Scientific Review of Mental Health Practice*, vol. 1, no. 1**
(Spring/Summer 2002).
Subscribe now!



[About SRMHP](#)

[In The News](#)

[Subscriptions](#)

[Author Guidelines](#)

[Current Issue](#)

[Back Issues](#)

[Online Articles](#)

[Resource Links](#)

[Contact Us](#)

Affiliate Links:

[Commission for
Scientific Medicine
and Mental Health
\(CSMMH\)](#)

[Scientific Review of
Alternative
Medicine \(SRAM\)](#)

[Center for Inquiry
International \(CFI\)](#)

[Committee for the
Scientific
Investigation of
Claims of the
Paranormal
\(CSICOP\)](#)

[Skeptical Inquirer
Magazine \(SI\)](#)



©2004 Center for Inquiry | [SRMHP Home](#) | [About SRMHP](#) | [Contact Us](#) |

Dodo Bird, Phoenix, or Urban Legend? The question of psychotherapy equivalence. *Scientific Review of Mental Health Practice* 1, 11-22. Google Scholar. Hunsley, J, Lee, CM (2007). Research-informed benchmarks for psychological treatments: efficacy studies, effectiveness studies, and beyond. *Professional Psychology: Research and Practice* 38, 21-33. CrossRef Google Scholar. Jacobson, NS, Truax, P (1991). Phoenix - mythical creation - this bird creates a pyre and rising from its own ashes. Phoenix plays an important role in ikusstve and literature, a symbol of immortality and resurrection. In Greek myths the Phoenix. In Greek myths the phoenix is an eagle-like bird with golden and red feathers. The most striking qualities of the phoenix was unusual life and the ability to self-immolation, after rising from its ashes. There are several options for the myth of Phoenix. In the classic version of every 500 years old Phoenix arrives from India in the temple of the sun at Heliopolis, in Libya. Chief The dodo-bird has been extinct for almost three centuries, thanks to the Dutch east Indian shipping company which made Mauritius a stopping point on their sea travels and altered the living conditions on the island in such a way that the dodo died. In the caring sector, the dodo-bird is however alive and kicking thanks to a short article of Saul Rosenzweig (1907-2004) from 1936. This psychologist researched the different effects of specific forms of psychotherapy and wondered whether method A was in any way better than method B or C. Soon enough, he found that the effectiveness of each method