

Assessment *The Dodo Bird Verdict—Controversial, Inevitable and Important: A Commentary on 30 Years of Meta-Analyses*

Rick Budd¹ and Ian Hughes^{2*}

¹Ludlow Street Healthcare, Llanbedr Court, Newport, Wales, UK

²Psychology Department, Cardiff and Vale NHS Trust, Cardiff, Wales, UK

In this article, the assertion that different psychological therapies are of broadly similar efficacy—often called the ‘Dodo Bird Verdict’—is contrasted with the alternative view that there are specific therapies that are more effective than others for particular diagnoses. We note that, despite thirty years of meta-analytic reviews tending to support the finding of therapy equivalence, this view is still controversial and has not been accepted by many within the psychological therapy community; we explore this from a theory of science perspective. It is further argued that the equivalence of ostensibly different therapies is an inevitable consequence of the methodology that has dominated this field of investigation; namely, randomised controlled trials [RCTs]. The implicit assumptions of RCTs are analysed and it is argued that what we know about psychological therapy indicates that it is not appropriate to treat ‘type of therapy’ and ‘diagnosis’ as if they were independent variables in an experimental design. It is noted that one logical consequence of this is that we would not expect RCTs to be capable of isolating effects that are specific to ‘type of therapy’ and ‘diagnosis’. Rather, RCTs would only be expected to be capable of identifying the non-specific effects of covariates, such as those of therapist allegiance. It is further suggested that those non-specific effects that have been identified *via* meta-analysis are not trivial findings, but rather characterise important features of psychological therapy. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: Meta-Analysis; Dodo Bird Verdict; Psychotherapy, CBT

THE HISTORICAL CONTEXT

The first meta-analysis of studies examining the efficacy of different types of psychological therapy

was reported by Smith and Glass in 1977. At that time, meta-analysis was still a new and relatively controversial methodology (e.g., Gallo, 1978) that was just starting to be used as a way of summarizing a rapidly expanding literature. Rather than subjectively appraise the support that different studies provided for different psychological therapies, as had been done until that time (e.g.,

*Correspondence to: Ian Hughes, Psychology Department, Whitchurch Hospital, Cardiff, Wales, UK. CF14 7XB.
E-mail: Ian.Hughes@CardiffandVale.wales.nhs.uk

Luborsky, Singer, & Luborsky, 1975), Smith and Glass (1977) turned to meta-analysis with the hope that this methodology would provide an objective method for summarizing the literature. The now familiar goal of their meta-analytic review was to aggregate the sizes of the effects found across different studies and thereby draw *objective* conclusions about which psychological therapies were most effective.

With improvements in methodology, meta-analysis has ceased to be regarded as a controversial *methodology* and has become the standard tool for reviewing and summarizing the research literature. Moreover, there have been significant developments both in methodology, and in the specificity of the questions being asked (e.g., Benish, Imel, & Wampold, 2008; Siev & Chambless, 2007). However, while the potential utility of this methodology is now accepted by most, the *conclusions* that can reasonably be drawn from the literature continue to be controversial and hotly debated (e.g., Holmes, 2002; Tarrier, 2002).

THE DODO BIRD VERDICT AS CONTROVERSIAL

The controversy that Smith and Glass (1977) first attempted to solve via the use of meta-analysis was that raised by Luborsky et al.'s (1975) qualitative (non-meta-analytic) review of the literature, from which they concluded that the then extant outcome literature provided no clear support for the superiority of any one type of psychological therapy over any other. Quoting the Dodo Bird in Lewis Carroll's *Alice in Wonderland*, they asked whether the conclusion to be drawn was that 'everyone has won and all must have prizes'. However, to some authorities the superiority of behavioural and cognitive-behavioural techniques was so evident from existing qualitative literature reviews that Eysenck (1978) was prompted to brand meta-analysis, and the conclusions that Smith and Glass (1977) had come to, as an 'exercise in mega-silliness' (emphasis added).

While the subsequent meta-analysis of Shapiro and Shapiro (1982) provided further support for the assertion that most psychological therapies were of broadly equivalent efficacy, this debate has continued to the present day.

In **depression**, some authors have reported meta-analyses showing cognitive-behavioural therapy (CBT) to be superior to other forms of psychological therapy for depression (Gloaguen, Cottraux,

Cucherat, & Blackburn, 1998; Svartberg and Stiles, 1991). On the contrary, Leichsenring (2001) found short-term psychodynamic therapy to be of similar efficacy to CBT for depression, and Ekers, Richards, and Gilbody (2008) have conducted a meta-analysis that suggested that purely behavioural treatments were at least as effective as cognitive-behavioural approaches in depression. Moreover, re-analysing Gloaguen et al.'s (1998) data, Wampold, Minami, Baskin and Tierney (2002) concluded that all the other psychological therapies that they classified as *bona fide* were of similar efficacy to CBT in treating depression. A more recent meta-analysis (Cuijpers, van Straten, Andersson, & Van Oppen, 2008) came to similar overall conclusions, although hinting at a possible slight advantage for Interpersonal Therapy (IPT) over other approaches, in the treatment of mild-moderate severity of depression. Certainly, no clear pattern of superiority for any one treatment modality has yet emerged.

In the area of **anxiety**, the only meta-analysis that has clearly found CBT to be superior to other treatments is that reported by Siev and Chambless (2007), with these authors reporting that although Applied Relaxation (AR) was as effective as CBT for Generalized Anxiety Disorder (GAD), CBT was more effective than AR for Panic Disorder. A major review by Hofmann and Smits (2008) found CBT to be broadly efficacious across the spectrum of anxiety disorders, but the effect sizes in the Intention-to-treat (ITT) studies included were very modest, and Ost (2008) has further noted no clear evidence of significantly improved outcome rates over approximately four decades of CBT in anxiety. Norton and Price (2007), reporting a meta-analysis that summarized 108 treatment studies across all the anxiety disorders, showed that behavioural therapy (BT) was as efficacious as cognitive therapy (CT) and CBT, while Benish et al.'s (2008) meta-analysis in post-traumatic stress disorder (PTSD) also showed broad equivalence of outcomes across modalities.

These findings have led to both reviews of the meta-analytic literature (e.g., Butler, Chapman, Forman, & Beck, 2006) and to meta-analyses of the meta-analyses (e.g., Grissom, 1996). While some authors have come to the conclusion that 'the Dodo Bird Verdict is alive and well – mostly' (Luborsky et al., 2002, p. 2), others suggest that 'the dodo bird is extinct' (Beutler, 2002, p. 30). In the current paper we address the question of why this debate has persisted, and we ask what we can learn from this debate.

OBJECTIVE SCIENCE VERSUS SCIENCE AS DISCOURSE

To understand why this debate has lasted for over 30 years, and has not been resolved despite the numerous meta-analyses that have been conducted over that time, it is helpful to reflect on the nature of scientific debate. It is often thought that science advances by collecting data and objectively evaluating them, with agreed knowledge emerging from these data (Charlton, 2000). From this standpoint, it is easy to understand the hopes of the early meta-analyses, i.e., that aggregating effect sizes across studies would enable a decision to be made about which interventions are most effective, and that this would then act as a basis to promote their widespread adoption and discourage the use of other, less effective therapies (Charlton, 2000). However, the hostile responses (e.g., Eysenck, 1978) that the early meta-analyses (e.g., Smith & Glass, 1977) generated, and the unresolved debate that continues to this day (e.g., Holmes, 2002; Tarrier, 2002) indicate that consensus has still **not** emerged from the data.

But perhaps this continuing controversy is not as surprising as it might appear at first sight. As Charlton (2000) notes: '... science is not an indiscriminating process of relentless, cumulative, impartial observation', (p. 25), but rather '... science is in the questions asked and the means by which they are addressed' (p. 25), with this being very clearly demonstrated in the meta-analytic literature. Over 25 years ago, Shapiro and Shapiro (1982) noted that the outcome literature contained a preponderance of studies examining cognitive-behavioural interventions and that '... the few instances of dynamic and humanistic therapy were interpretable as straw man treatments, not expected to yield good results by the investigators' (p. 597). Similarly, Wampold et al., (2002) reported that, although their meta-analysis initially appeared to indicate that CBT was superior to other talking therapies for depression, when 'straw man' interventions were removed from the meta-analysis CBT was found to be no more efficacious than other *bona fide* therapies. (See also Benish et al., 2008 in relation to PTSD).

The related finding that researchers' therapeutic allegiance has a significant effect on the outcome of studies examining the efficacy of different types of therapy has been well documented across a range of diagnoses (e.g., Berman, Miller, & Massman, 1985; Robinson, Berman, & Neimeyer, 1990), with Luborsky et al. (1999) suggesting that the correlation between the researchers' therapeutic allegiance and the results of comparative outcome

studies may in fact be as high as 0.85. While this might surprise those who view science as merely an objective accumulation of facts, it would come as no surprise to social psychologists who long ago argued that social influence processes make such biases an inherent part of all human activity, scientific activity included (Farr, 1976).

Thus, rather than view science as an objective accumulation of facts, we might better view science as a discourse that occurs in a public domain between proponents of rival theoretical perspectives. As Charlton (2000) notes:

Alternative observations, experiments and interpretations emerge from different directions. Disagreement and dissent (whether over fundamentals or minutiae) is the normal state of affairs within any active subject within science. Despite disagreement, some scientific theories are accepted and built-upon, while others are . . . either ignored or rejected (p. 15).

Therefore, to understand why a consensus has not emerged out of the meta-analytic literature we first need to clarify the nature of the rival theoretical perspectives this debate has centred on.

THE RIVAL PERSPECTIVES AND THE RESULTING SCIENTIFIC DISCOURSE

The theoretical perspectives underpinning this debate can be broadly classified into two rival positions. The first, which is often adopted by researchers aligned to a traditional psychotherapeutic perspective (e.g., Leichsenring, 2001; Luborsky et al., 2002; Shapiro & Shapiro, 1982; Wampold, 2001), is that most psychological therapies are of approximately equal efficacy. Researchers aligned to this perspective have tended to argue that the apparent equivalence of different psychological therapies is due to important common factors that all therapies share. Hence, they argue that it is such elements as creating the therapeutic alliance, remoralizing the patient, instilling hope, normalizing distressing experiences and giving the client a meaningful explanation of her/his difficulties, which account for most of the change that occurs in **any** psychological therapy.

The second, rival perspective, which is generally adopted by researchers aligned to CBT (e.g., Butler et al., 2006; Tarrier, 2002), is that there are specific, usually broadly cognitive-behavioural, interventions that are effective for particular disorders,

with all other interventions being little more than inert placebos (e.g., Chambless & Hollon, 1998). Within this perspective, non-specific common elements of therapy have generally been seen as, at best, necessary but not sufficient elements in the process of change. (In effect, this regards such non-specific elements of therapy as little different in kind from extra-therapeutic factors such as the availability of a comfortable consulting room in which both therapist and client may be seated or the client being greeted by a friendly receptionist.) As Westen, Novotny, and Thompson-Brenner (2004) have noted, this perspective has recently moved from being a purely theoretical perspective to being one that is now closely aligned with the development of treatment guidelines and protocols that emphasize the use of Evidence-Supported Therapies (ESTs) to the exclusion of other (untested) interventions. It has, thus, moved to centre stage in the development of mental health policy in the UK and the USA.

As numerous authors have noted (e.g., Holmes, 2002; Luborsky, McLellan, Diguer, Woody, & Seligman, 1997; Shapiro & Shapiro, 1982), CBT has for many years tended to dominate the research literature. Ever since Eysenck's (1952) groundbreaking literature review, behavioural, and more recently cognitive-behavioural, approaches have generally been viewed as evidence-based interventions founded on scientific principles, while more traditional 'psychotherapeutic' interventions have been viewed at best as not being evidence based, and at worst as being founded on unscientific concepts.

Possibly as a result of the dominance of CBT within the research literature, most supporters of other treatment approaches have been content to demonstrate that these approaches are *as effective as* CBT (e.g., Leichenring, 2001; Leichenring, Rabung, & Leibing, 2004). Researchers aligned to CBT have been keen to demonstrate that it is *more effective than* traditional psychotherapeutic approaches, and that it is *as effective as pharmacotherapy* (e.g., Rachman & Wilson, 2008). Although serious doubts have recently been raised about the effectiveness of medication in mental health care (Kirsch et al., 2008), given the dominance of the medical model the fact that pharmacotherapy continues to be the benchmark against which CBT compares itself is perhaps not surprising.

The desire of CBT proponents to demonstrate that cognitive-behavioural approaches are superior to other psychological interventions is intensified by the fact that the therapeutic interventions in this tradition have been explicitly derived from causal

models. Thus, for proponents of CBT, demonstrating the apparent equivalence of diverse psychological therapies—including CBT—would appear also to undermine the causal models of CBT. In contrast, other psychotherapeutic approaches are less likely to be based on a specific model of the aetiology of the condition they are intended to treat (e.g., IPT), or they are based on complex models that do not relate directly to standard diagnostic categories; see, for example, Strupp and Binder's (1985) or Malan's (1963) brief psychodynamic therapies. Therefore, for researchers coming from a psychotherapeutic perspective, the therapy equivalence finding—or Dodo Bird Verdict—does not present a significant challenge to their models of therapy.

Perhaps more importantly from a CBT perspective, however, demonstrating that all psychological therapies are broadly equivalent in terms of outcome could be taken as evidence that CBT lacks any specific therapeutic ingredients and in reality provides little more than a structure for developing a therapeutic alliance and a treatment rationale that engenders expectations of change in the client. Such a proposition presents a particular challenge to those who are aligned with CBT, since the goal of over 50 years of research has not only been to develop specific interventions for specific conditions, but also to refine and improve these interventions on the basis of developments in our understanding of the psychological processes that are believed to underpin these conditions. By contrast, meta-analytic findings of therapeutic equivalence do not present such a major problem for those aligned with the traditional psychotherapeutic perspective, since these authors are typically concerned with facilitating individual change through promoting insight, clarifying emotions and resolving interpersonal problems within the context of a healing therapeutic relationship, rather than with developing specific interventions to treat specific conditions (see, for example, Lambert, Bergin, & Garfield, 2004).

THE HIDDEN ASSUMPTIONS OF RCTs

Randomized Controlled Trials (RCTs) have long been considered to provide the 'gold standard' for testing the efficacy of new drugs, such that for psychological therapies to challenge the dominance of pharmacotherapy, it was necessary for researchers to use RCTs in order to compare the efficacy of these two treatment modalities (Westen

et al., 2004). Meta-analysis has risen to the challenge of summarizing this literature, having convincingly demonstrated that pharmacotherapy and psychological therapy are of broadly equivalent efficacy (e.g., Gloaguen et al., 1998; Gould, Buckminster, Pollack, Otto, & Yap, 1997; Gould, Otto, & Pollack, 1995). While this finding has been widely applauded by researchers aligned to CBT (Rachman & Wilson, 2008), as Beutler (2000) has noted, it has been more controversial in psychiatry. The success of this enterprise has resulted in RCTs also becoming the dominant methodology for exploring a second question, namely, what are the active ingredients of the psychological therapies? However, while psychological therapies, as a whole domain, have stood up to the 'gold standard' test of RCTs (or at least have proved to be as effective as medication), this methodology has not proved capable of mining these therapies for further gold. That is to say, it has not been possible to use RCTs to determine which of the various psychological therapies is most efficacious, or to determine which particular elements of these therapies account for their efficacy.

Westen et al. (2004) have argued that RCT methodology contains a number of assumptions that are violated when this methodology is applied to psychological interventions. This results in the findings generated by these RCTs not being generalizable to routine clinical practice. In the current paper it is argued that three of the false assumptions identified by Westen et al. (2004) in fact account for the Dodo Bird Verdict. These three false assumptions are as follows: 'most patients have one primary problem or can be treated as if they do' (p. 634); 'psychological symptoms can be understood and treated in isolation from personality dispositions' (p. 636); and '... the elements of efficacious treatment are dissociable and hence subject to dismantling' (p. 640). These assumptions will be described briefly below. We then consider how they may account for the Dodo Bird Verdict.

The present paper also goes beyond the arguments of Westen et al. (2004) to suggest that the core problem lies in the level of analysis that is used in psychological treatment research. We argue that the active ingredients of different psychological treatments are located at the level of individual therapist-client interactions, which, most importantly, are reciprocal in nature. This is in contrast to locating the active ingredients at the level of different psychological treatment packages.

THE PROBLEM OF TREATING 'DIAGNOSIS' AS AN INDEPENDENT VARIABLE

One of the principal criteria used for including participants in RCTs is that all participants have the same diagnosis, with any who present with co-morbid axis I or axis II diagnoses being excluded from the study in order ostensibly to maximize the homogeneity of the group on which the treatment is being tested. Westen et al. (2004) have cogently argued that this selection criterion severely limits the generalizability of RCTs to clinical populations, since large percentages of patients in routine clinical practice typically do have co-morbid axis I and/or axis II diagnoses (Morrison, Bradley, & Westen, 2003; Thompson-Brenner & Westen, 2005a). Nevertheless, restrictive selection criterion constitute a core feature of most RCTs as their principal aim is to explore what intervention is most effective for which tightly defined disorder; or 'what works for whom' as some have termed this question (Roth & Fonagy, 1996). To this end one goal of recent meta-analyses has been to examine the interaction between the independent variables 'diagnosis' and 'type of treatment' (Norton & Price, 2007; Siev & Chambless, 2007).

Even with these restrictive exclusion criteria removing obvious co-morbidity, however, the resulting 'homogeneity' of the now clinically unrepresentative participants in the RCTs is likely to be more apparent than real. As Westen et al. (2004) noted, the assumption that even a tightly-defined diagnosis can be used to classify participants into meaningful groups, and thereby address the question of whether treatment X or Y works best for type of patient Z, is itself likely to be seriously mistaken. As these authors and many others (e.g., Craske & Waters, 2005; Watson, O'Hara, & Stuart, 2008) have argued, there is in fact little reason to believe that the Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnostic categories are valid, psychologically meaningful constructs as they have neither been logically derived from any psychological model, nor derived empirically (e.g., via latent class analysis, factor analysis, etc.). Rather, the diagnostic categories, and the criteria that need to be met for a patient to be given a particular diagnosis, have been agreed by committee with little consideration having been given to issues of construct validity. As many diagnoses consist of groups of overlapping symptoms which have high incidences of co-morbidity (Brown, Campbell, Lehman, Grisham, & Mancill, 2001; Mineka,

Watson, & Clark, 1998; Zimmerman, McGlinchey, Chelminski, & Young, 2008), they would not be expected, on *a priori* grounds, to be conceptually distinct entities. More importantly, however, when the construct validity of these diagnostic categories has been explored empirically they have been found not to be supported by the data (e.g., Krueger, 1999; Zimmerman & Chelminski, 2003). Thus, if patients with different diagnoses present with overlapping symptoms, and the symptoms of those who have the same diagnosis have different (unknown) aetiologies, it would not be expected that there should be specific psychological interventions that are most effective for particular diagnoses.

Most significantly in this regard, structural equation modelling and other related techniques have suggested that anxiety and depression are best described by a triarchic structure (Olinio, Klein, Lewinsohn, Rohde, & Seeley, 2008; Watson et al., 2008). This suggests that although the diagnoses of Major Depressive Disorder (MDD) and GAD each have symptoms that are unique to them, they also share common features, which include personality traits that underlie these presenting problems (Ormel & Wohlfarth, 1991; Ormel, Oldehinkel, & Vollebergh, 2004) and which may account for the large proportion of patients who have alternating episodes of GAD and MDD over time (Kessler et al., 2008).

A further problem with treating diagnosis as an independent variable is the assumption that treatment can be provided to patients in different diagnostic groups independently of the patients' personalities. While participants are typically excluded from RCTs if they have clear axis II personality disorders, no attention is paid to participants' personality characteristics if these are within the 'normal' range. But as Westen and Thompson-Brenner (2004) note, the assumption that patients can be treated independently of their personality is not supported either by clinical experience (Morrison et al., 2003; Thompson-Brenner & Westen, 2005b) or by the research data, which suggest that certain personality traits function as underlying risk factors for experiencing episodes of anxiety and/or depression (Brown, 2007; Ormel et al., 2004).

Interestingly, in the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program, patients' levels of perfectionism were found to moderate the effectiveness of the treatments and, while this personality trait did not interact with the type of treatment provided, there was an interaction between level

of perfectionism and the quality of the therapeutic alliance in predicting outcome (Blatt, Zuroff, Quinlan, & Pilkonis, 1996). Thus, if personality factors that are likely to affect the outcome of an intervention are not assessed, and their effects are not partialled out, this factor is likely to confound the results of the RCT.

THE PROBLEM OF TREATING 'TYPE OF THERAPY' AS AN INDEPENDENT VARIABLE

In experimental designs it is axiomatic that the independent variables are clearly defined, and can be adequately operationalized, so that causal effects on the dependent variable(s) can be examined. As Westen et al. (2004) note, 'type of psychological therapy' is almost unique among independent variables in that the experimental manipulation (a type of therapy) is an intervention provided by different therapists over a long time period, with the experimental manipulation often being provided over as many as 20 sessions. Maintaining the purity of such a complex experimental manipulation therefore presents a significant methodological challenge, even when the different therapies are set out in detailed treatment manuals and their integrity is monitored via the analysis of audio/video recordings of sessions.

In this regard it is instructive to consider the NIMH Treatment of Depression Collaborative Research Program (Elkin et al., 1989), as this is often considered to be one of the most methodologically sound comparative efficacy outcome studies that have been conducted to date. For example, using process analysis to examine the two psychological therapies (CBT and IPT) that were examined in the NIMH study, Ablon and Jones (1999, 2002) concluded that these two therapies shared many common components and that the IPT intervention may have had more in common with a CBT ideal therapy prototype than it had with an IPT ideal therapy prototype (at least across two of the study sites). Reporting on the Sheffield Psychotherapy Studies, Stiles and his colleagues (Stiles & Shapiro, 1994; Stiles, Shapiro, & Firth-Cozens, 1988) noted that although the cognitive-behavioural and interpersonal-psychodynamic therapies used in these studies were clearly distinguishable in terms of therapist response mode, there was nonetheless considerable overlap in terms of therapist behaviour across both treatment types. While this could be viewed simply as a failure to ensure treatment

fidelity, it is interesting to note that in routine clinical practice a similar finding has been noted. That is to say, while the behaviour of therapists aligned to either CBT or psychotherapeutic perspectives is distinguishable, there nonetheless remains significant overlap in the responses provided by therapists from these different orientations (e.g., Ablon & Jones, 1998; Goldfried, Raue, & Castonguay, 1998; Trijsburg, Trent, & Perry, 2004).

Most critically, however, the central problem in maintaining the purity of different therapeutic interventions rests on a feature that is central to, and inherent in all, psychological therapies. That is, all psychological therapies, whatever their theoretical orientation, consist of a social interaction (usually dyadic) between therapist and client where *both* influence each other's behaviour (Stiles, Honos-Webb, & Michael, 1998). Thus the goal of maintaining the 'purity' of a psychological therapy (experimental manipulation) across clients, even when supported by a treatment manual and validity checks, is by its very nature flawed. As Hardy, Stiles, Barkham, and Startup (1998) have shown in the Sheffield Psychotherapy Studies, clients influence the behaviour of therapists in subtle ways despite rigorous attempts to manualize interventions, with this being an inherent feature of all human interactions (Farr, 1976), not just of psychological therapy.

Paradoxically, this phenomenon is as explicitly acknowledged in CBT as it is in the interpersonal-psychodynamic therapies, with CBT defining itself as a *collaborative* enterprise that occurs between client and therapist (Beck, 1976). So it is acknowledged that the therapist must respond to the client in a flexible, collaborative manner in order to develop a therapeutic alliance, which is known to be central to the effectiveness of all psychological interventions (Horvath & Bedi, 2002; Martin, Garske, & Davis, 2000), including CBT (Waddington, 2002). However, despite this fundamental feature of therapy (and all other human interactions) being acknowledged in clinical practice, it is typically sidestepped by RCTs since it presents a serious challenge to their experimental validity.

THE MISLEADING DRUG METAPHOR AS A CAUSE OF THE DODO BIRD VERDICT

Stiles and Shapiro (1989) have noted that a common and often implicit assumption underlying much outcome research is that psychological therapy can be likened to a drug that can be *given* to a

client in various doses (number of sessions), with different drugs (different types of psychological therapy) being indicated for different conditions. This metaphor underpins the use of experimental designs that treat type of therapy and diagnosis as if they were independent variables, with the aim of discovering ESTs that are diagnosis specific. The dominance of the drug metaphor in therapy outcome research reflects the hegemony of the medical model and has prompted the widespread acceptance of RCTs as the 'gold standard' for testing the efficacy of psychological therapies. While the dominance of this approach to outcome research has been successful to the extent that it has resulted in the efficacy of psychological therapies becoming widely accepted by governments and other health care funding organizations (see for example Rachman & Wilson's, 2008 favourable discussion of UK government mental health initiatives), it has at the same time resulted in the Dodo Bird Verdict. The reason for this lies in the fact that the central assumptions of RCTs (and, hence, the drug metaphor) are violated in the case of psychological therapies.

As has been argued above, diagnostic categories do not consist of groups of clients who have distinct problems, with the same psychological processes being responsible for the cause and/or maintenance of 'the problem' that characterizes each diagnosis. Rather, they consist of people presenting with fairly heterogeneous and overlapping problems (symptoms) with different (generally unknown) psychological processes causing and/or maintaining these problems. As such, even if it were possible to ensure that different psychological therapies were 'pure' treatments that contained unique components that did not overlap in any way, we would not expect the same intervention to be effective for every client within a diagnostic group. Rather we would expect the same intervention to be differentially effective for different clients with the same diagnosis, this depending upon the psychological and social factors that are maintaining that client's presenting problems (symptoms). One extension of this analysis is to offer a broadly based, non-specific therapeutic package to a diagnostically mixed patient group, with the expectation that on average any given patient will be able to benefit from enough of the package to enable them to make worthwhile progress. This has been done, for instance, in a recent study by McEvoy and Nathan (2007), with apparently promising results: 10 2-hour CBT groups for a diagnostically *heterogeneous* sample of patients

produced effect sizes at least as good as those that would have been expected from diagnosis-specific interventions. Although these particular authors were using a CBT approach, the overall implication of their study, that non-specific interventions are as useful as are more specific interventions, seems broadly compatible with the wider psychological therapy equivalence findings cited above. A logical extension of this approach is the Internet-based CBT packages now being developed, which are also not diagnosis specific.

Moreover, as has been argued above, different psychological therapies are not well-controlled experimental manipulations, but rather consist of a range of overlapping interventions, which, while broadly distinguishable from each other, share many common components. They are provided by different therapists, some of whom are more effective than others, with the client's understanding and interpretation of the therapist's intervention, and its therapeutic impact, depending upon both the context in which it was provided and upon idiosyncratic features of the client. Worst of all, from a purely experimental perspective, the nature of the therapy (experimental manipulation) that each client (experimental participant) receives is influenced in subtle and uncontrollable ways by the client (experimental participant).

THE DODO BIRD VERDICT AS INEVITABLE

Since the independent variables (diagnosis and type of therapy) that are submitted to meta-analysis can neither be adequately controlled nor operationalized, we would not expect any meta-analysis of RCTs to be capable of successfully partitioning any possible treatment-specific variance in outcome to these two independent variables. It is even more unlikely to identify any interaction effect. That is to say, we would not expect this methodology to be capable of isolating which type of therapy is most effective for which diagnosis. Thus, the repeated failure to find any clear evidence that any one type of therapy is superior to any other, for any given diagnosis (e.g., Benish et al., 2008), is a direct consequence of the fact that key assumptions underpinning RCTs, namely that the experimental manipulations are clearly definable and operationalizable, do not apply to psychological therapy.

A final reason for the Dodo Bird Verdict is that the experimental manipulations (different types of

therapy) on which RCTs are based are not capable of directly and uniquely manipulating the psychological processes that account for the symptom change that occurs in therapy. This point is freely acknowledged by Hofmann (2008) in response to Longmore and Worrell's (2007) criticism of cognitive therapy on the grounds that component studies do not indicate the superiority of interventions that directly target cognitions over purely behavioural interventions that do not do so. In this regard, Hofmann (2008) accepts that cognitions can change in a therapeutically helpful manner without the need to explicitly target them in treatment. Hence, for example, a client's catastrophic cognitions about the consequences of a panic attack might change simply in response to them being invalidated by exposure without the need for them to have been directly challenged via setting up the exposure sessions as behavioural experiments. Or, as a more extreme example, the whole atmosphere of the clinic as a friendly, accepting and 'normal' environment in which the client is valued might begin to challenge patients' negative self-concept even before they get to see their therapist!

Ablon and Jones (1998) have similarly argued that ostensibly different *types* of verbal response from the therapist (e.g., a traditional psychoanalyst offering an interpretation, a CBT practitioner presenting a CBT formulation etc.) can have essentially the same therapeutic *function*, depending on how the therapist's intervention is interpreted and understood. This is consistent with the long known phenomenon that clients' and therapists' reports of what were the significant events that occurred in therapy can often diverge quite markedly (e.g., Llewelyn, 1988): it is not necessarily what the therapist thought s/he was doing in therapy that was the active ingredient, but rather how the client understood/interpreted the therapist's intervention.

WHAT HAVE WE LEARNED? THE DODO BIRD VERDICT AS IMPORTANT

Given that core assumptions underlying RCTs are violated in the case of psychological therapies (Westen et al., 2004) so as to make it impossible to partition any possible treatment specific variance in outcome to either type of therapy and/or the interaction between type of therapy and diagnosis, it is to be expected that most of what we have learned from over 30 years of meta-analytic reviews of RCTs concerns those effects that are

common across different interventions and diagnoses. Although this could be seen as a 'failure', it is worth noting that those effects that have been isolated shed considerable light on the process of therapy, even though these findings are not specific to any individual therapeutic intervention or diagnosis. In summary, meta-analytic reviews of the outcome literature have demonstrated *three* effects that are of both theoretical and practical importance.

First, the literature clearly demonstrates that the therapist's allegiance to a particular intervention is highly predictive of the efficacy of that intervention, with Luborsky et al. (1999) and Wampold (2001) presenting data which indicate that this effect is substantially larger than any other effect (either specific or general) that has so far been isolated from RCTs. Rather than regard this merely as a failure of 'scientific objectivity', it is possible to argue that it in fact highlights an important basic therapeutic processes. In this regard, Wampold and his colleagues (Benish et al., 2008; Wampold et al., 1997; Wampold et al., 2002) have shown, across a range of diagnoses, that therapies with a clear treatment rationale, and which therapists themselves believe to be effective, are more effective than those that are intended to be placebo interventions. Moreover, they have demonstrated that this result holds, regardless of the rationale underpinning the interventions that are intended to be therapeutic.

At a research level, Berman et al. (1985) showed over 20 years ago that this effect has the potential to bias meta-analytic reviews of the literature and, as a result, researcher allegiance needs to be partialled out in meta-analyses in order to control for the effect this may have on the differential outcome of different therapies. At a clinical level this finding implies that an EST is likely to be less effective if it is administered by a therapist who is not convinced of the value of that therapy than if it is administered by one who is. This runs counter to the drug metaphor of psychological therapy, which suggests that therapeutic effectiveness rests mainly on ensuring that therapists only use ESTs and closely adhere to the related treatment manuals. Rather, it may be more appropriate for therapists to use evidence-based *techniques* with their clients, in the context of an *individual* case formulation (which the therapist of necessity has allegiance to), than for therapists slavishly to follow treatment manuals that, as Westen et al. (2004) have cogently argued, may not be appropriate for many of the clients who are treated in routine clinical practice.

Second, the literature clearly demonstrates that the therapeutic alliance is central to the effectiveness of all types of psychological therapy (Horvath & Bedi, 2002; Martin et al., 2000). This implies that one important goal for further research is to explore what factors promote the therapeutic alliance and, in particular, to focus on ways of promoting a strong alliance in those clients whose personality may hinder this process (Blatt et al., 1996). Moreover, at a clinical level it suggests that it may be of value for therapists routinely to monitor clients' progress in therapy (e.g., Lambert, 2007), and the quality of the therapeutic alliance, in order to enable them promptly to address events that may disrupt therapy.

Third, the literature clearly demonstrates that some therapists are more effective than others (Crits-Christoph et al., 1991; Luborsky et al., 1997), with this effect surprisingly having been demonstrated both for clinicians providing psychological therapies (Kim, Wampold, & Bolt, 2006) and for those providing pharmacotherapy (McKay, Imel, & Wampold, 2006). At a research level this suggests, as Westen et al. (2004) propose, that systematically exploring expert clinicians' routine practice might provide useful hypotheses about the nature of successful therapy which can then be tested using more traditional research designs. With regard to clinical practice, this finding suggests that there may be value in trying to construct more systematic methods for coaching and mentoring therapists than is the case with many current clinical supervision arrangements.

CONCLUSIONS AND IMPLICATIONS FOR FUTURE RESEARCH

As has been argued above, the failure of RCTs (and the meta-analyses of RCTs) to partition out any specific effects of psychological therapy to different types of interventions (even when diagnosis has been controlled for) is attributable to the fact that it is exceptionally difficult to operationalize and control this independent variable. That is to say that despite rigorous attempts at manualization the evidence suggests that it is not possible to ensure that different therapies do not share such a large proportion of active therapeutic ingredients as to dilute the power of RCTs to isolate treatment-specific effects. Moreover, and perhaps even more significantly, it is impossible to ensure that different interventions do not affect the same psychological mechanisms of change, thus similarly limiting the

power of RCTs to isolate treatment specific effects. When these limitations are combined with the fact that diagnoses are overlapping, 'fuzzy' concepts, which do not adequately control for psychological factors (e.g., sub-clinical personality features, etc.) that are likely to interact with the type of intervention in determining its efficacy, it is not surprising that the Dodo Bird Verdict has survived for more than 30 years. In short, its survival rests on the fact that while RCTs (and the meta-analysis of data obtained from RCTs) are capable of demonstrating that psychological therapy *as an aggregate* is effective, they are not capable of identifying the *specific* elements of therapy that are effective for different individuals.

From the research perspective the clarion call is clear. RCTs do not represent an objective 'gold standard' of research, but rather, are simply one of many research methodologies, which, like all methodologies, have their strengths and weaknesses. When used to test causal relationships between a dependent variable and a set of independent variables that can be adequately defined and operationalized, there is no alternative to constructing an RCT. However, for addressing other research questions, such as 'how does psychological therapy help different people change?' different methodologies could be more appropriate (e.g., Pachankis & Goldfried, 2007).

Moreover, there is a need to move away from designing psychological interventions targeted at an invalid psychiatric nosology, to developing interventions directed at alleviating distressing experiences (e.g., low mood, intrusive memories) or which are directed at changing problematical behaviours (e.g., avoidance, rumination). These interventions need to be developed in light of our understanding of the psychosocial processes that maintain them. How and when they are used in therapy needs to be informed by a full understanding of the individual therapy process.

To illustrate our favoured approach, let us consider 'depression'. Typically, within the present, diagnostically driven approach, patients diagnosed as having 'depression' are offered a package—often of around 20 sessions—of CBT treatment, which comprises a number of almost invariant elements (e.g., activity scheduling, thought diaries, thought challenging). We consider that such an approach ignores the variability and complexity of the symptoms of people diagnosed as having 'depression', and that it also ignores the psychosocial factors maintaining the symptoms (e.g., family relationships, employment problems, the

broader environmental context); and, that it also ignores the therapeutic relationship. Low mood, for instance, is a very common feature of 'depression'. In order to ameliorate it, we would advocate a detailed analysis of its maintaining factors on an individual basis, and then the application of an appropriate technique (e.g., increasing activity levels *or* resolving relationship conflicts *or* exploring the validity of negative automatic thoughts). But, importantly, this needs to be done within the context of the therapeutic process, that is to say, at an appropriate time, and in a way that is most likely to promote change within the client. (A similar general approach is much more fully described in the important chapter by Norcross and Beutler [2008]).

Most importantly we need to reject the medical model, cease to view therapy as being like a drug that is *given* to clients and, instead, view it for what it is, a social influence process that occurs (most typically) between two people.

REFERENCES

- Ablon, J.S., & Jones, E.E. (1998). How expert clinicians' prototype of an ideal correlate with outcome in psychodynamic and cognitive-behavioural therapy. *Psychotherapy Research*, 8, 71–83.
- Ablon, J.S., & Jones, E.E. (1999). Psychotherapy process in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*, 67, 64–75.
- Ablon, J.S., & Jones, E.E. (2002). Validity of controlled clinical trials of psychotherapy: Findings from the NIMH Treatment of Depression Collaborative Research Program. *American Journal of Psychiatry*, 159, 775–783.
- Beck, A.T. (1976). *Cognitive therapy and the emotional disorders*. New York: International Universities Press.
- Benish, S.G., Imel, Z.E., & Wampold, B.E. (2008). The relative efficacy of bona fide psychotherapies for treating post-traumatic stress disorder: A meta-analysis of direct comparisons. *Clinical Psychology Review*, 28, 746–758.
- Berman, J.S., Miller, R.C., & Massman, P.J. (1985). Cognitive Therapy Versus Systematic Desensitization: Is One Treatment Superior? *Psychological Bulletin*, 97, 451–461.
- Beutler, L.E. (2000). David and Goliath: When empirical and clinical standards of practice meet. *American Psychologist*, 55, 997–1007.
- Beutler, L.E. (2002). The dodo bird is extinct. *Clinical Psychology: Science and Practice*, 9, 30–34.
- Blatt, S.J., Zuroff, D.C., Quinlan, D.M., & Pilkonis, P.A. (1996). Interpersonal factors in brief treatment of depression: further analyses of the National Institute of Mental Health Treatment of Depression Collabora-

- tive Research Program. *Journal of Consulting and Clinical Psychology*, 64, 162–171.
- Brown, T.A. (2007). Temporal course and structural relationships among dimensions of temperament and DSM-IV anxiety and mood disorder constructs. *Journal of Abnormal Psychology*, 116, 313–328.
- Brown, T.A., Campbell, L.A., Lehman, C.L., Grisham, J.R., & Mancill, R.B. (2001). Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. *Journal of Abnormal Psychology*, 110, 585–599.
- Butler, A.C., Chapman, J.E., Forman, E.M., & Beck, A.T. (2006). The empirical status of cognitive-behavioural therapy: A review of meta-analyses. *Clinical Psychology Review*, 26, 17–31.
- Chambless, D.L., & Hollon, S.D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, 7–18.
- Charlton, B.G. (2000). The new management of scientific knowledge: a change in direction with profound implications. In A. Miles, J.R. Hampton, and B. Hurwitz (Eds), *NICE, CHI and the NHS reforms: enabling excellence or imposing control?* (pp. 13–32). London: Aesculapius Medical Press.
- Craske, M.G., & Waters, A.M. (2005). Panic disorder, phobias and Generalized Anxiety Disorder. *Annual Review of Clinical Psychology*, 1, 197–225.
- Crits-Christoph, P., Baranackie, K., Kurcias, J.S., Beck, A., Carroll, K., & Perry, K. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research*, 1, 81–91.
- Cuijpers, P., van Straten, A., Andersson, G., & Van Oppen, P. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology*, 76, 909–922.
- Ekers, D., Richards, D., & Gilbody, S. (2008). A meta-analysis of randomized trials of behavioural treatment of depression. *Psychological Medicine*, 38, 611–623.
- Elkin, I., Shea, T., Watkins, J.T., Imber, S.D., Sotsky, S.M., & Collier, J.F. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Archives of General Psychiatry*, 46, 971–982.
- Eysenck, H.J. (1952). The effects of psychotherapy: an evaluation. *Journal of Consulting Psychology*, 16, 319–324.
- Eysenck, H.J. (1978). An exercise in mega-silliness. *American Psychologist*, 33, 517.
- Farr, R.M. (1976). Experimentation: A social psychological perspective. *British Journal of Social and Clinical Psychology*, 15, 225–238.
- Gallo, P.S. (1978). Meta-analysis – a mixed metaphor. *American Psychologist*, 33, 515–517.
- Gloaguen, V., Cottraux, J., Cucherat, M., & Blackburn, I.M. (1998). A meta-analysis of the effects of cognitive therapy in depressed patients. *Journal of Affective Disorders*, 49, 59–72.
- Goldfried, M., Raue, P.J., & Castonguay, L.G. (1998). The therapeutic focus in significant sessions of master therapists: A comparison of cognitive-behavioural and psychodynamic-interpersonal interventions. *Journal of Consulting and Clinical Psychology*, 66, 803–810.
- Gould, R.A., Buckminster, S., Pollack, M.H., Otto, M.W., & Yap, L. (1997). Cognitive-behavioural and pharmacological treatment of social phobia: A meta-analysis. *Clinical Psychology: Science and Practice*, 4, 291–306.
- Gould, R.A., Otto, M.W., & Pollack, M.H. (1995). A meta-analysis of treatment outcomes for panic disorder. *Clinical Psychology Review*, 15, 819–884.
- Grissom, R.J. (1996). The magic number 7 + –2: Meta-analysis of the probability of superior outcome in comparisons involving therapy, placebo and control. *Journal of Consulting and Clinical Psychology*, 64, 973–982.
- Hardy, G.E., Stiles, W.B., Barkham, M., & Startup, M. (1998). Therapist responsiveness to client interpersonal styles during time-limited treatments for depression. *Journal of Consulting and Clinical Psychology*, 66, 304–312.
- Hofmann, S.G. (2008). Common misconceptions about cognitive mediation of treatment change: A commentary to Longmore and Worrell (2007). *Clinical Psychology Review*, 28, 67–70.
- Hofmann, S.G., & Smits, J.A.J. (2008). Cognitive-behavioral therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *Journal of Clinical Psychiatry*, 69, 621–632.
- Holmes, J. (2002). All you need is cognitive behaviour therapy? *British Medical Journal*, 324, 288–289.
- Horvath, A.O., & Bedi, R.P. (2002). The alliance. In J.C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 37–69). New York: Oxford University Press.
- Kessler, R.C., Gruber, M., Hettema, J.M., Hwang, I., Sampson, N., & Yonkers, K.A. (2008). Co-morbid major depression and generalized anxiety disorders in the National Comorbidity Survey follow-up. *Psychological Medicine*, 38, 365–374.
- Kim, D.A., Wampold, B.E., & Bolt, D.M. (2006). Therapist effects in psychotherapy: A random effects modelling of the NIMH TDCRP data. *Psychotherapy Research*, 16, 161–172.
- Kirsch, I., Brett, J.D., Huedo-Medina, T.B., Scoboria, A., Moore, T.J., & Johnson, B.T. (2008). Initial severity and antidepressant benefits: A meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med*, 5(2), e45.
- Krueger, R.F. (1999). The structure of common mental disorders. *Archives of General Psychiatry*, 56, 921–926.
- Lambert, M. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, 17, 1–14.
- Lambert, M., Bergin, A., & Garfield, S. (2004). *Handbook of psychotherapy and behaviour change* (5th ed.). New York: Wiley.
- Leichsenring, F. (2001). Comparative effects of short-term psychodynamic psychotherapy and cognitive-behavioural therapy in depression: A meta-analytic approach. *Clinical Psychology Review*, 21, 401–419.
- Leichsenring, F., Rabung, S., & Leibling, E. (2004). The efficacy of short-term psychodynamic psychotherapy in specific psychiatric disorders: a meta-analysis. *Archives of General Psychiatry*, 61, 1208–1216.

- Llewellyn, S.P. (1988). Psychological therapy as viewed by clients and therapists. *British Journal of Clinical Psychology*, 27, 223–237.
- Longmore, R.J., & Worrell, M. (2007). Do we need to challenge thoughts in cognitive behaviour therapy? *Clinical Psychology Review*, 27, 173–187.
- Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapy: Is it true that 'everyone has won and all must have prizes?' *Archives of General Psychiatry*, 32, 995–1008.
- Luborsky, L., McLellan, A.T., Diguier, L., Woody, G., & Seligman, D.A. (1997). The psychotherapist matters: comparison of outcomes across twenty-two therapists and seven patient samples. *Clinical Psychology: Science and Practice*, 4, 53–65.
- Luborsky, L., Diguier, L., Seligman, D.A., Rosenthal, R., Johnson, S., Halperin, G., Bishop, M., & Schweizer, E. (1999). The researcher's own therapeutic allegiances: A 'wild card' in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice*, 6, 49–62.
- Luborsky, L., Rosenthal, R., Diguier, L., Andrusyna, T.P., Berman, J.S., Levitt, J.T., Seligman, D.A., & Krause, E.D. (2002). The dodo bird verdict is alive and well – mostly. *Clinical Psychology: Science and Practice*, 9, 2–12.
- Malan, D.H. (1963). *A study of brief psychotherapy*. London: Tavistock Press.
- Martin, D.J., Garske, J.P., & Davis, K.M. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 68, 438–450.
- McEvoy, P.M., & Nathan, P. (2007). Effectiveness of cognitive behavioural therapy for diagnostically heterogeneous groups. *Journal of Consulting and Clinical Psychology*, 75, 344–350.
- McKay, K.M., Imel, Z.E., & Wampold, B.E. (2006). Psychiatrist effects in the psychopharmacological treatment of depression. *Journal of Affective Disorders*, 92, 287–290.
- Mineka, S., Watson, D., & Clark, L.A. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology*, 49, 377–412.
- Morrison, K.H., Bradley, R., & Westen, D. (2003). The external validity of controlled clinical trials of psychotherapy for depression and anxiety: A naturalistic study. *Psychology and Psychotherapy: Theory, Research and Practice*, 76, 109–132.
- Norcross, J.C., & Beutler, L.E. (2008) Integrative psychotherapies. In: R.J. Corsini & D. Wedding (Eds) *Current Psychotherapies* (8th ed., pp. 481–511) Belmont, CA: Brooks/Cole.
- Norton, P.J., & Price, E.C. (2007). A meta-analytic review of adult cognitive-behavioural treatment outcome across the anxiety disorders. *Journal of Nervous and Mental Diseases*, 195, 521–531.
- Olino, T.M., Klein, D.N., Lewinsohn, P.M., Rohde, P., & Seeley, J.R. (2008). Longitudinal associations between depressive and anxiety disorders: a comparison of two trait models. *Psychological Medicine*, 38, 353–363.
- Ormel, J., & Wohlfarth, T. (1991). How neuroticism, long-term difficulties, and life situation change influence psychological distress: A longitudinal model. *Journal of Personality and Social Psychology*, 60, 744–755.
- Ormel, J., Oldehinkel, A.J., & Vollebergh, W. (2004). Vulnerability before, during, and after a major depressive episode: A 3-wave population-based study. *Archives of General Psychiatry*, 61, 990–996.
- Ost, L.-G. (2008). Cognitive behaviour therapy for anxiety disorders: 40 years of progress. *Nordic Journal of Psychiatry*, 62, 5–10.
- Pachankis, J.E., & Goldfried, M.R. (2007). On the next generation of process research. *Clinical Psychology Review*, 27, 760–768.
- Rachman, S.J., & Wilson, G.T. (2008). Expansion in the provision of psychological treatment in the United Kingdom. *Behaviour Research and Therapy*, 46, 293–295.
- Robinson, L.A., Berman, J.S., & Neimeyer, R.A. (1990). Psychotherapy for the treatment of depression: a comprehensive review of controlled outcome research. *Psychological Bulletin*, 108, 30–49.
- Roth, A., & Fonagy, P. (1996). *What works for whom*. New York: Guilford Press.
- Shapiro, D.A., & Shapiro, D. (1982). Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin*, 92, 581–604.
- Siev, J., & Chambless, D.L. (2007). Specificity of treatment effects: Cognitive therapy and relaxation for generalized anxiety and panic. *Journal of Consulting and Clinical Psychology*, 75, 513–522.
- Smith, M.L., & Glass, G.V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752–760.
- Stiles, W.B., & Shapiro, D.A. (1989). Abuse of the drug metaphor in psychotherapy process-outcome research. *Clinical Psychology Review*, 28, 1–16.
- Stiles, W.B., & Shapiro, D.A. (1994). Disabuse of the drug metaphor: Psychotherapy process-outcome correlations. *Journal of Consulting and Clinical Psychology*, 62, 942–948.
- Stiles, W.B., Shapiro, D.A., & Firth-Cozens, J.A. (1988). Verbal response mode use in contrasting psychotherapies: A within-subjects comparison. *Journal of Consulting and Clinical Psychology*, 56, 727–733.
- Stiles, W.B., Honos-Webb, L., & Michael, M. (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice*, 4, 439–458.
- Strupp, H.H., & Binder, J.L. (1985). *Psychotherapy in a new key: A guide to time limited dynamic psychotherapy*. New York: Basic Books.
- Svartberg, M., & Stiles, T.C. (1991). Comparative effects of short-term psychodynamic psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 59, 704–714.
- Tarrier, N. (2002). Yes, cognitive behaviour therapy may be all you need. *British Medical Journal*, 324, 291–292.
- Thompson-Brenner, H., & Westen, D. (2005a). A naturalistic study of psychotherapy for bulimia nervosa, part 1: comorbidity and therapeutic outcome. *Journal of Nervous and Mental Diseases*, 193, 573–584.
- Thompson-Brenner, H., & Westen, D. (2005b). A naturalistic study of psychotherapy for bulimia nervosa, part 2 therapeutic interventions in the community. *Journal of Nervous and Mental Diseases*, 193, 585–595.

- Trijsburg, W.R., Trent, S.T., & Perry, C.J. (2004). An empirical study of the differences in interventions between psychodynamic therapy and cognitive-behavioural therapy for recurrent major depression. *Canadian Journal of Psychoanalysis*, 12, 325–345.
- Waddington, L. (2002). The therapy relationship in cognitive therapy: A review. *Behavioural and Cognitive Psychotherapy*, 30, 179–191.
- Wampold, B.E. (2001). *The great psychotherapy debate: models, methods and findings*. Mahwah, NJ: Erlbaum.
- Wampold, B.E., Mondin, G.W., Moody, M., Stich, F., Benson, K., & Ahn, H.N. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: empirically 'all must have prizes'. *Psychological Bulletin*, 122, 203–215.
- Wampold, B., Minami, T., Baskin, T., & Tierney, S. (2002). A meta-(re)analysis of the effects of cognitive therapy versus 'other therapies' for depression. *Journal of Affective Disorders*, 68, 159–165.
- Watson, D., O'Hara, M.W., & Stuart, S. (2008). Hierarchical structures of affect and psychopathology and their implications for the classification of emotional disorders. *Anxiety and Depression*, 25, 282–288.
- Westen, D., Novotny, C., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin*, 30, 631–663.
- Zimmerman, M., & Chelminski, I. (2003). Generalized anxiety disorder in patients with major depression: Is DSM-IV's hierarchy correct? *American Journal of Psychiatry*, 160, 504–512.
- Zimmerman, M., McGlinchey, J.B., Chelminski, I., & Young, D. (2008). Diagnostic co-morbidity in 2300 psychiatric out-patients presenting for treatment evaluated with a semi-structured diagnostic interview. *Psychological Medicine*, 38(38), 199–210.