



Strathprints Institutional Repository

Elliott, Robert (2002) *Hermeneutic single-case efficacy design*. *Psychotherapy Research*, 12 (1). pp. 1-21. ISSN 1050-3307

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<http://strathprints.strath.ac.uk/>) and the content of this paper for research or study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to Strathprints administrator: <mailto:strathprints@strath.ac.uk>

Hermeneutic Single Case Efficacy Design

Robert Elliott

University of Toledo

This article is based on the SPR Presidential Address delivered at the June 2001 meeting of the Society for Psychotherapy Research in Montevideo, Uruguay. I gratefully acknowledge the inspiration of Art Bohart, on whose initial work the method described here is based, as well as the contributions of Denise Defey, Constance Fischer, David Rennie and Kirk Schneider; and my students, Mona Amer, Rob Dobrenski, Helena Jersak, Cristina Magaña, Rhea Partyka, Suzanne Smith, John Wagner, and especially Michelle Cutler. Address correspondence to Robert Elliott, Department of Psychology, University of Toledo, Toledo, OH 43606; e-mail: Robert.Elliott@utoledo.edu; telephone: 419-530-2715; fax: 419-530-8479.

Abstract

In this paper, I outline Hermeneutic Single Case Efficacy Design (HSCED), an interpretive approach to evaluating treatment causality in single therapy cases. This approach uses a mixture of quantitative and qualitative methods to create a network of evidence that first identifies direct demonstrations of causal links between therapy process and outcome, and then evaluates plausible nontherapy explanations for apparent change in therapy. I illustrate the method with data from a depressed client who presented with unresolved loss and anger issues.

Keywords: research design, single case design, hermeneutic research, depression, process-experiential psychotherapy

Hermeneutic Single Case Efficacy Design

All Gaul, wrote Julius Caesar (51 BCE/1960), is divided into three parts. Similarly, psychotherapy research can be organized into three main areas. Unlike ancient Gaul, these domains are defined not by the rivers that separate them but rather by the scientific questions which motivate them, and by the language, customs and principles of the researchers who seek to answer these questions.

These three questions and they research territories they define are: (a) Has this client (or group of clients) actually changed? (psychotherapy outcome research; e.g., Strupp, Horowitz, & Lambert, 1997); (b) Is psychotherapy generally responsible for change? (psychotherapy efficacy and effectiveness research; e.g., Haaga & Stiles, 2000); and (c) What specific factors (within therapy or outside it) are responsible for change? (psychotherapy change process research; e.g., Greenberg, 1986).

In this paper I focus on the second question, that of the causal efficacy or effectiveness of psychotherapy. However, tackling this question requires answering both the first question (whether there is any actual change), and the last question (what processes mediate change). Furthermore, I attempt to meet the challenge of answering these three questions for single therapy clients and nonbehavioral therapies, by proposing the Hermeneutic Single Case Efficacy Design (HSCED)¹.

The Need for a Critical-Interpretive Approach to Causal Research Design

The standard tool for addressing the efficacy of psychotherapy, the randomized clinical trials (RCT) design, is an extremely blunt instrument that suffers from a host of scientific difficulties (see Cook & Campbell, 1979; Haaga & Stiles, 2000; Kazdin, 1998), especially poor statistical power, differential attrition, and poor generalizability due to restricted samples.

Causal emptiness. Not the least of these difficulties are two related problems: First, RCTs rely on a stripped-down operational definition of causality (from J.S. Mill; see Cook & Campbell, 1979), in which inferring a causal relationship requires establishing: (a) temporal precedence (priorness); and (b) necessity & sufficiency (that cause and effect covary). Thus, RCTs are “causally empty,” offering conditions under which inferences can be reasonably made, but providing no method for truly understanding the specific nature of the causal relationship. For this reason, Haynes and O’Brien (2000) and others have argued that inferring a causal relation requires another condition, the provision of a plausible account (“logical mechanism”) for the possible causal relation. Unfortunately, RCTs provide no built-in method for establishing or identifying such plausible causal processes².

Poor generalizability to single cases. Second, RCTs do not warrant causal inferences about single cases. Even when a therapy has been shown to be responsible for change in general, for any specific client, factors other than therapy may actually have been the source of the observed or reported changes, or the client’s apparent change may have been illusory. The existence of this inference gap argues for moving the locus of causal inference from the group to the single case, where each client’s distinctive change process can be traced and understood.

Rescuing the N = 1 Design. The traditionally-sanctioned alternative to group experimental design has been single-participant experimental design (Kazdin, 1998). The logic and potential clinical utility of these designs is compelling (Sidman, 1960), and advocates have long argued for the applicability of these designs to nonbehavioral treatments (Peterson, 1968; Morgan & Morgan, 2001). Nevertheless, the logic of these designs depends on behavioral

assumptions about the change process, especially the situational specificity of behavior and the foundational role of functional analysis in treatment. As a result, these designs have never caught on outside traditional behavior therapy, not even for cognitive-behavior therapies.

In order to address the difficulties of applying single case design to nonbehavioral therapies, methodologists such as Kazdin (1981) and Hayes, Barlow and Nelson-Gray (1999) have proposed a more flexible alternatives which "stretch" the guidelines of standard single case design, in particular, the clinical replication series. These authors have proposed the following characteristics of single case research as useful for increasing internal validity (Kazdin, 1981):

1. Systematic, quantitative data (vs. anecdotal).
2. Multiple assessments of change over time.
3. Multiple cases (a form of multiple baseline design).
4. Change in previously chronic or stable problems.
5. Immediate or marked effects following the intervention.

Note that the first three features are design strategies over which the researcher has some control, while the last two (stability and discontinuous change) are case-specific and emergent.

Sources of HSCED

Kazdin's (1981) general guidelines provided me with one of the sources for HSCED. Another source was Cook and Campbell's (1979) brief description of the "modus operandi" (i.e., one-group post-only design), which they argued can be interpreted when there is rich contextual information and what they called "signed causes" (i.e., influences whose presence is evident in their effects). Mohr (1993) goes even further, arguing that the single case is the best situation for inferring and generalizing causal influences, which are obscured in group designs.

The final and most important source for HSCED was Bohart and Boyd's (1997) description of an interpretive approach to examining client qualitative accounts of change over therapy. Starting from a client's assertion that she has changed and her claim that this is the result of therapy, Bohart and Boyd ask, "What would it take to make a convincing case that therapy caused a reported change?" In general, the answer to this question takes the form of two types of information: (a) other evidence that the change occurred (corroboration); and (b) plausible ruling out of alternative possible sources of the change.

A rich case record of comprehensive information on therapy process and outcome (e.g., using multiple perspectives, sources and types of data) provides a useful starting point. However, critical reflection on the claim of therapy-caused change is also required, through maintaining awareness of one's personal expectations and theoretical presuppositions, while systematically searching for evidence that casts doubt on one's preferred account. To do this, Bohart and Boyd (1997) proposed a set of plausibility criteria for evaluating client causal accounts, including evidence for grounding in the client's experience, deviation from expectations, elaboration, discrimination between positive and negative effects and processes, idiosyncraticness, and coherence.

Essentials of Hermeneutic Single Case Design

In our society, experts make systematic use of practical reasoning systems to make various important judgements, including legal rulings and medical decisions. HSCED is proposed as such a practical reasoning system, with the specific purpose of evaluating the causal role of therapy in bringing about outcome. It builds on Bohart and Boyd's (1997) approach, but examines a larger set of alternative nontherapy explanations, makes greater use of quantitative

outcome and weekly change data, and devotes more attention to systematically determining whether change has occurred.

Case Example. To illustrate HSCED, I use a running case example: a depressed, 49-year old male European-American client whom I will refer to as Paul. Paul's main presenting problems were financial worries, general negativity and cynicism, problems communicating with his son (with whom he felt identified and who had become clinically depressed also), and most importantly, unresolved issues from a rapid succession of the deaths in his family (mother, father, brother), 10 years previously. He was diagnosed with Bipolar II disorder (major depressive episodes plus hypomania) and seen at the Center for the Study of Experiential Therapy for 39 sessions of Process-Experiential therapy, primarily focusing on issues of anger and loss. He was seen by a second-year clinical psychology graduate student over the course of 16 months. I did the research interviews.

Rich Case Record

The first prerequisite for a hermeneutic single case efficacy design is a rich, comprehensive collection of information about a client's therapy. This includes background information, as well as data on therapy process and outcome, using multiple sources or measures. I have found the following data to be useful:

(a) Basic facts about client and therapist, including demographic information, diagnoses, presenting problems, therapeutic approach or orientation (e.g., given above for Paul).

(b) Quantitative outcome measures. Therapy outcome has both descriptive qualitative (how the client changed) and quantitative (how much the client changed) aspects. For Paul, quantitative measures included standard self-report questionnaires such as the Symptom Checklist-90 (Derogatis, 1983), the Inventory of Interpersonal Problems (Horowitz, Rosenberg, Baer, Ureño, & Villaseñor, 1988), and the Simplified Personal Questionnaire (PQ; Elliott, Shapiro & Mack, 1999). At a minimum, these measures should be given at the beginning and end of therapy, but it is also a good idea to give them periodically during therapy, every 8 to 10 sessions. Paul's quantitative outcome data are given in Table 1.

(c) Change Interview. The Change Interview (Elliott, Slatick & Urman, 2001) is a semi-structured interview that provides (a) qualitative outcome data, in the form of client descriptions of changes experienced over the course of therapy; and (b) client descriptions of their attributions for these changes, including helpful aspects of their therapy. (Information on negative aspects of therapy and on medications is also collected.) The Change Interview takes 30 - 45 minutes and can be carried out by a third party, every 8 - 10 sessions, at the end of therapy, and at follow-up.

I asked Paul to tell me what changes he had noticed in himself since therapy started. He listed six pre-to-post changes, including "More calm in the face of challenges." "Giving myself more credit for accomplishments." "Doing better financially." "Being a happier person." "Being more hopeful about my life." and "(I) don't feel young anymore" (=a negative change).

I then asked Paul to rate each change on three rating scales, among them the attributional question, "How likely do you think this change would have been without therapy?" Paul substantiated his strong therapy-attribution with his descriptions of the various ways in which he understood his therapy to have brought about the changes, including this summary of his change process: "I don't think I would have looked at those [feelings] on my own. And obviously there were a lot of those that I didn't look at on my own... I

think the therapy actually in some way... gave me a process of grieving, maybe not all the stages of grief, but some."

(d) Weekly outcome measure. A key element in HSCED is the administration of a weekly measure of the client's main therapy-related problems or goals. We used the Simplified Personal Questionnaire (Elliott, Shapiro & Mack, 1999), an individualized target complaint measure consisting of roughly ten 7-point distress rating scales.

Paul's weekly mean PQ scores are given in Figure 1, which reveals numerous statistically reliable ($> .53$) week-to-week shifts in PQ scores.

(e) Helpful Aspects of Therapy (HAT) Form (Llewelyn, 1988) is a frequently-employed qualitative measure of client perceptions of significant therapy events. This open-ended seven-item questionnaire is administered to clients after therapy sessions. In HSCED, HAT data are used to pin-point significant therapeutic processes which may be associated with change on the weekly outcome measure or to corroborate change processes referred to in the Change Interview. In his HAT descriptions, Paul rated 12 significant events with ratings of 8 ("Greatly helpful") or higher. These descriptions provide a summary narrative of what the client considered at the time to be the most helpful events in his therapy. Paul gave two events ratings of 8.5, one in session 10 ("The need to work on resolving my anger toward my father. The realization that the anger I have carried around might be directed at him."), and the other in session 36 ("Realizing the anger and/or distrust of my wife. I believe I have been suppressing this.").

(f) Records of therapy sessions. Therapist process notes and videotapes of therapy sessions are collected in case they are needed to pinpoint, corroborate, or clarify issues or contradictions elsewhere in the data.

For example, in order to make sense out of the largest shifts in Paul weekly PQ scores, I used his therapist's process notes.

Direct Evidence: Clear Links between Therapy Process and Outcome

In HSCED, the starting point is direct evidence pointing to therapy as a major cause of client change. In order to be confident about proceeding further with the analysis, it is best to have at least two separate pieces of evidence supporting the therapy-change link.

(1) Retrospective attribution. First, the client may attribute a reported change to therapy, without specifying the nature of the felt connection.

Clear support for the therapy efficacy hypothesis can be found in Paul's "likelihood-without-therapy" ratings and his description of the role his therapy played in helping him feel more calm in the face of challenges.

(2) Process-outcome mapping. The content of the client's posttherapy changes correspond to specific events, aspects or processes within therapy.

For example, five of Paul's 12 high-rated significant events (e.g., Session 12: "Feeling the hurt, fear, and sadness related to the loss of my family. It enabled me to realize that I can feel what might be under my anger.") refer to work on unresolved loss/grief issues regarding his family of origin, his major posttherapy change.

(3) Within-therapy process-outcome correlation. In addition, theoretically central in-therapy process variables (e.g., adherence to treatment principles) may be found to covary with week-to-week shifts in client problems.

To examine this possibility for Paul's therapy, I correlated his therapist's postsession ratings of her use of Process-Experiential treatment principles, tasks and response modes

with difference scores on the Personal Questionnaire (n=34 pairs of data points). Only two of the 63 correlations were statistically significant ($p < .05$), less than would be expected by chance. Therefore, at least on this basis, I found no evidence of a therapy-change link.

(4) Early change in stable problems. Therapeutic influence can be inferred when therapy coincides with change in long-standing or chronic client problems, contrasting with an explicit or implicit baseline.

Paul's mean PQ scores (Table 1) do appear to show a reliable, 2-point drop from pre- to posttreatment. Although we do not know how long Paul's problems had continued at roughly the same level, it is clear that some of them were of many years' standing. Furthermore, his two pretreatment PQ mean scores are consistent with each other (4.44 and 4.11) and in the clinical range (i.e., well above the cut-off of 3). It is true that his weekly PQs (Figure 1) show some instability, but this appears to be a consequence of three "outlier" sessions (4, 24, 39). If these are ignored, the largest improvement occurs after session 1, moving the client into the nonclinical range.

(5) Event-shift sequences. An important therapy event may immediately precede a stable shift in client problems, particularly if the nature of the therapy process and the change are logically related to one another (e.g., therapeutic exploration of an issue, followed the next week by change on that issue).

Although Paul's PQ ratings contained many substantial shifts (Figure 1), the largest shifts appeared to reflect temporary changes associated with the three outlier sessions. However, there was no more than weak evidence for event-shift sequences in Paul's therapy, because negative shifts followed significant helpful events almost as often as positive shifts did.

Summary of direct evidence. In this evaluation of possible direct evidence for the efficacy of Paul's therapy, I found supportive evidence on three out of five possible indicators, enough to provide corroboration of his claim in the Change Interview.

Indirect Evidence: Competing Explanations for Apparent Client Change

HSCED also requires a good-faith effort to find nontherapy processes that could account for an observed or reported client change. Table 2 summarizes each of eight nontherapy explanations for apparent client change, along with methods for evaluating its presence. The practical reasoning process involved in evaluating these alternatives is like detective work, with contradictory evidence sought and available evidence weighed carefully. As a result, some nontherapy explanations may be ruled out entirely, while others may be found to partially or even completely explain the observed change. In addition, it is important to weigh both positive and negative evidence. Discrepancies point to complexities or restrictions on the scope of change or the possible role of therapy.

A further consideration is the degree of uncertainty considered tolerable. The circumstances under which therapists and their clients operate preclude near certainty ($p < .05$), suggests "reasonable assurance" or "beyond a reasonable doubt" ($p < .2$) as a more realistic and useful standard of proof.

1. Trivial or Negative Change

The first four nontherapy explanations assume that apparent client change is illusory or artifactual. To begin with, the apparent changes may be negative or trivial.

Trivial change. On the one hand, a client might describe a change in such highly qualified or ambivalent terms as to cast doubt on its importance ("I think maybe I'm beginning to see that change might be possible." "Um, I guess feeling better about myself, maybe just a little bit."). In addition, clients sometimes describe changes in other people ("My husband has finally started fixing the house.") or in their life circumstances, (one of Paul's changes was "doing better financially"). In the same way, changes on quantitative outcome measures may also fall into the trivial range (e.g., one point on the BDI).

Negative change. Alternatively, changes might be negative, casting doubt on the overall effectiveness of the therapy. For example, at an earlier assessment, Paul noted that he and his son were now fighting more than when he began therapy; given the importance of this issue for him, this negative change could be taken as evidence of clinical deterioration which negated or at least compromised positive changes. Similarly, changes on quantitative outcome measures may also occur in the negative direction (see Ogles, Lambert & Sawyer, 1995).

Strategies for dealing with trivial or negative change. It is useful to define intervals or threshold values that can be used to define change as nontrivial. Jacobson and Truax (1989?) proposed two criteria for evaluating change: (a) statistically reliable change and (b) movement past clinical caseness (i.e., clinical significance) cut-offs. Table 1 includes these criteria for three key measures (SCL-90-R, IIP, and PQ values taken from Barkham et al., 1996; Ogles et al., 1995).

Paul's outcome data indicate that from pre- to posttherapy he moved past the caseness threshold on 2 out of the 3 measures (SCL-90 and PQ) but that the amount of change was reliable on only 1 of these measures (PQ).

Next, in order to assess for negative changes, the researcher can ask the client to describe any negative changes which might have occurred over the course of therapy.

For example, at posttreatment, when Paul was asked about negative changes, he noted that he did not feel young anymore.

Finally, clients can be asked to evaluate the importance of changes, perhaps using rating scales (cf. Kazdin (1999)). In the Change Interview, the client rates the importance of each change, using a 5-point scale. In addition, the manner of the client's description can be examined for qualifiers and other forms of ambivalence.

Thus, in his posttherapy Change Interview, Paul rated all of his positive changes as either "very" or "extremely" important; by contrast, he rated his one negative change, feeling old, as "slightly" important. His descriptions of his changes were directly stated without qualifiers and even included occasional intensifiers ("I am *in fact* more calm than I've been in a *long* time").

Although examining whether changes are trivial or negative is fairly straightforward, it is important to consider both positive and negative evidence and to interrogate discrepancies among measures and types of evidence.

Conclusions. On the one hand, Paul's Change Interview data clearly support the view that he improved in several important ways. On the other hand, of his quantitative outcome measures, only the PQ shows clear improvement at posttreatment. Thus, the evidence for substantial, positive change is mixed.

2. Statistical Artifacts

Related to the possibility of trivial change is statistical error, including measurement error, regression to the mean, and experimentwise error.

Measurement error involves random inconsistency on quantitative measures, stemming from inattention in completing forms, ambiguous wording of items, misunderstanding of the meaning of items, and rating tasks that exceed the ability of raters to accurately characterize their experiences. The standardized error of the difference (S_{diff}) provides an appropriate estimate of error in measuring client change (Jacobson & Truax, 1991, where r_{xx} is test-retest reliability and s_1 is the standard deviation of a comparable normative population):

$$S_{diff} = s_1 \sqrt{2(1 - r_{xx})}$$

This formula allows us to establish a confidence interval for defining a minimum "Reliable Change Index" (RCI) value for client change, at either the traditional 95% level ($1.96 S_{diff}$), as originally proposed by Jacobson and colleagues (1984), or the 80% level ($1.29 S_{diff}$) proposed here. Client change which is less than the minimum RCI value is judged to reflect measurement error. Table 1 contains RCI minimum values for three common outcome measures: SCL-90, IIP, and PQ.

Paul's pre-to-post change on 2 of these three measures is less than the prescribed values, although his change on the PQ is statistically reliable and would easily survive a bonferroni correction. However, the frequent drastic shifts in Paul's weekly PQ scores raised issues about greater than expected temporal instability (consistent with his atypical bipolar diagnosis) and suggested that it would be a good idea to use the median of his first three and last three PQ scores (Table 1 values in parentheses). The difference between these pre- and posttherapy median PQ scores greatly exceeded the minimum RCI value of .53.

Regression to the mean by outliers. Regression to the mean occurs when measurements with less than perfect reliability are selected on the basis of their extreme values. This introduces bias that is not present when the measurement is later repeated, resulting in the second measurement taking a less extreme value, thus producing illusory change.

For example, one possible explanation for the numerous sharp spikes in Figure 1 is measurement error followed by regression to the mean. (Bipolar cycling between depressive and hypomanic states is another explanation, given the client's diagnosis).

If regression to the mean is operating, then repeating the measurement prior to beginning therapy is likely to reveal a sharp drop; if this occurs, the second measurement can be used as the pretest. If scores are consistent or increase across the two pretests, they can be combined (e.g., using the median of the first 3 and last 3 scores of a weekly change measure). A more qualitative approach to assessing regression to the mean is to carry out careful pretreatment assessment to determine the duration of the client's problems.

As Figure 1 indicates, Paul's two pretreatment PQ scores were fairly stable (both above 4 or "moderately distressed"), indicating that they were representative of his usual responses; thus, the substantial changes observed in Paul's PQ scores are probably not a function of regression to the mean. Unfortunately, we did not obtain multiple pretests on the SCL-90 or IIP, and do not have systematic data on problem duration.

Experimentwise error a function of carrying out multiple significance tests on change measures. When examining many measures for evidence of change, some apparently reliable differences may occur due to chance alone. For example, when three measures are used to evaluate the reliability of pre-post change, with the relaxed standard proposed, each measure has .2 probability of indicating change when none existed (Type I error). Compounded across three

measures, the probability of one or more measures out of three showing reliable change by chance is .49! The solution here is to require reliable change on two out of three measures (this corresponds to a probability of .10), or on one measure at a more conservative probability level, such as $p < .05$. Requiring replication of reliable change across different outcome measures allows us to designate a client as demonstrating "Global Reliable Change."

Using these criteria, Paul showed reliable change at posttreatment on only one out of the three measures, thus failing to demonstrate Global Reliable Change. However, he did satisfy this standard at 6-month follow-up (see last two columns in Table 1), and his posttherapy PQ change exceeded the $p < .05$ significance level.

Conclusions. Regarding statistical artifacts, Paul's results are mixed. Regression to the mean is unlikely to have accounted for pre-post change on the PQ. However, the data do not support global reliable improvement at posttherapy, although they do support it at 6-month follow-up. At posttherapy, it would be most accurate to say that Paul shows reliable but limited change.

3. Relational Artifacts

Apparent client improvement may also reflect interpersonal dynamics between client and therapist or researcher, in particular attempts to please the latter. The classic relational artifact is the legendary (but impossible to attribute) "hello-goodbye" effect, in which the client enters therapy emphasizing distress in order to impress the research staff to accept him/her. However, at the other end of therapy, the client emphasizes positive functioning, either to express gratitude to the therapist and research staff or to justify ending therapy. I suspect that the use of fixed time limits in most therapy research works to strengthen this effect: If therapy is going to end anyway, there is little to be gained by trying to look worse than one is, and one might as well make the best of it!

Evaluating the plausibility of reported therapy attributions. In order to determine the role of self-presentational interpersonal artifacts, client narrative descriptions are invaluable. These accounts are probably most credible when they emerge spontaneously in therapy sessions or research interviews; however, researchers may prefer to obtain these accounts systematically via questionnaire or interview. Because interviews are a highly reactive form of data collection, client quantitative accounts of the effects of therapy need to be read very carefully for their nuance and style. Here is where several of Bohart and Boyd's (1997) plausibility criteria come into play, in particular, elaboration and discrimination. In particular, a credible client account of therapy's influence is elaborated: It contains specific details about what has changed and how the change came about; general descriptions are backed up by supportive detail. In addition, there is a mixture of positive, negative and neutral descriptions (differentiation). On the other hand, highly tentative or overly positive descriptions of the therapy, as well as positive reports which are lacking in detail or cannot be elaborated even under questioning, are likely to indicate interpersonally-driven self-report artifacts.

Interview strategy. The validity of client accounts is also enhanced if a researcher (rather than the therapist) interviews the client, and if the researcher conducts an extended, in-depth interview in which he or she encourages thoughtful self-reflection and openness on the part of the client.

Measuring relational response tendencies. A final strategy for dealing with relational artifacts is to use quantitative outcome measures (e.g., Tennessee Self-Concept Scale; Fitts &

Warren, 1996) that contain indices of the client's tendency to present in ways that emphasize or downplay problems.

Although I did not give Paul a social desirability or other quantitative validity scale, his Change Interview data contained substantial detail and at least some negative descriptions. Nevertheless, his manner and choice of language suggested that he may have deferred to me as an apparently successful authority figure of roughly the same age. This raises the possibility that he may have held back negative views of his therapy in order to avoid the possibility of offending me. This is one possible explanation for the discrepancy between his quantitative outcome measures and his very positive descriptions in the Change Interview. Because I was aware of the possibility of his trying to please me, I tried to communicate the attitude that his critical comments -- would be especially appreciated because they would help improve the therapy.

Conclusions. Overall, the detailed, differentiated nature of the qualitative data make it unlikely that relational artifacts are enough to explain the positive changes Paul described.

4. Expectancy Artifacts

Cultural or personal expectations ("scripts") or wishful thinking may give rise to apparent client change. That is, clients may convince themselves and others that since they have been through therapy they must therefore have changed. We expect post-therapy accounts to be particularly vulnerable to this sort of retrospective expectancy bias. However, longitudinal measurement of change is no guarantee against clients expecting themselves to do better at the end of therapy and therefore giving themselves the benefit of the doubt when recalling, integrating, and rating subtle or ambiguous phenomena such as mood symptoms, relationships, or self-evaluations.

Fortunately, the distinction between expectations and experience can be made partly by examining the language clients use to describe their experience. This is because expectation-driven descriptions must rely on shared cultural schemas about the effects of therapy; therefore, such "scripted" descriptions will typically make use of standard or clichéd phrases, such as "someone to talk to," or "insight into my problems." (See Elliott & James, 1989, for a review.) Client accounts of changes which conform entirely to cultural stereotypes are less credible than those that reflect more unusual experiences. By contrast, descriptions that are Idiosyncratic in their content or word choice are more believable.

In addition, expectation-driven expressions typically sound vague, intellectualized or distant from the client's experience. This is quite different from descriptions that are delivered in a detailed, careful, and self-reflective manner that indicates their grounding in the client's immediate experience (cf. Bohart & Boyd, 1997).

For example, Paul's descriptions generally contained a mixture of stock elements (the idea that releasing blocked feelings is therapeutic), but often qualified in idiosyncratic ways (e.g., typifying this release as a gradual process occurring over the course of a year). Some of Paul's descriptions of his change process did have an intellectualized, self-persuasive quality, for example (*italics added*): "I *think* I could see the progress, and *that can only help* build self-esteem and self-confidence. So as that goes up, *maybe* proportionately, *maybe* the anxiety goes down." Faced with this self-speculative account, I asked Paul to check the accuracy of his description: "Is that what it feels like, that

somehow you have this sense of your own having made progress, and that somehow makes the anxiety less?" Probes such as this enabled him to elaborate a more experientially-based account of extended, painful grieving for deceased family members. In addition, if a client reports being surprised by a change, it is unlikely to reflect generalized expectancies or stereotyped scripts for therapy. Researchers can determine this more systematically by asking clients to rate the degree to which they expected reported changes. For example, on four of his six changes, Paul rated himself as "somewhat surprised."

Conclusions. Paul's descriptions provide some evidence in the influence of therapy "scripts." However, I believe that the weight of the evidence points clearly towards a view of his descriptions as primarily experience-based. In particular, the existence of novel recasting of stock phrases, his ability to elaborate in experience-near terms, as well as his claim to have been somewhat surprised by most of the changes he experienced, all point to this conclusion.

5. Self-correction Processes: Self-help and Self-generated Return to Baseline Functioning

The remaining nontherapy explanations all assume that change has occurred, but that factors other than therapy are responsible. First, client internally-generated maturational processes or self-help efforts may be entirely responsible for observed changes. For example, the client may have entered therapy in a temporary state of distress which has reverted to normal functioning via the self-limiting nature of temporary crises or the person's own problem-solving processes. Alternatively, the change could be a continuation of an ongoing developmental trend. In these instances, client self-healing activities operate prior to or independently of therapy.

Direct and indirect self-report strategies. A general strategy for evaluating the final four nontherapy explanations is to ask the client:

When Paul was asked what brought about his changes, the first thing he said was "Being honest with myself, and being open to change, to trying new things." By itself, this statement would qualify as a report of self-generated change. However, without prompting, Paul then went on to indicate that this self-generated change process was related to therapy: "Since the therapy, I think I've had a lot more courage to really try new things. It's been exciting."

Similarly, the client can also be asked to assess how likely he or she feels the change would have been without therapy:

Paul rated three of his six changes (including the most important one of becoming more calm in the face of challenges) as "very unlikely without therapy," indicating his view that these changes clearly would not have happened without therapy. By contrast, he rated not feeling young any more (a negative change) as "somewhat likely" without therapy and the improvement of his financial situation as "neither likely nor unlikely" without therapy.

Therapist process notes provide an efficient source of information about client self-help efforts, and can be used in conjunction with shifts in PQ score.

Paul showed a large drop on his PQ after session 1; in her process notes, Paul's therapist noted that Paul had recently made the effort to speak to a friend with similar loss issues, and that this conversation had made him feel less alone.

Baseline and multiple pretest strategies. Self-correction in particular can also be evaluated by comparing client change to a temporal or expectational baseline. A temporal

baseline requires measuring the duration or stability of the client's main problems or diagnoses. In lieu of repeated pretreatment measurement, clinicians generally measure the baseline of a client's problem retrospectively, simply by asking the client how long he or she has had the problem. This is typically accomplished in a clinical interview, but it can also be done via a questionnaire or extracted from therapy sessions or therapist process notes.

We do not have systematic data from Paul on the duration of his problems; however, a review of session tapes and therapist process notes made it clear that two of his main problems -- anger/cynicism and unresolved grief -- were difficulties of at least 10 years standing, while his financial problems and anxiety about his son were of relatively recent vintage (i.e., on the order of months). The duration of his central problems make self-correction an unlikely explanation for his change on the PQ.

It may also prove valuable to listen for client narratives of self-help efforts begun before therapy, as when a depressed client applies for therapy services as part of a larger self-help effort that includes joining a health club, starting to take St. John's Wort, and making an effort to spend more time with friends. Such a self-generated process is likely to instigate a cascade of nontherapy change processes, including extra-therapy life events and even psychobiological factors.

Conclusions. The long duration of Paul's problems make it very unlikely that self-correcting processes are primary, independent causal processes here. Although Paul refers to self-correction processes in his interview responses, he emphasizes the causal role of therapy. On the other hand, the large drop in PQ score after session 1 comes in conjunction with reported self-help activities and takes place before therapy could be reasonably expected to have an effect. Thus, there is clear support for Self-Correction as a partial influence on Paul's changes, but the evidence indicates that it is unlikely that Self-Correction was primarily responsible, without themselves reflecting the influence of therapy.

6. Extra-therapy Events

Extra-therapy life events include changes in relationships such as deaths, divorces, initiation of new relationships, marriages, births, and other relational crises, as well as the renegotiation of existing relationships. In addition, clients may change jobs, get fired from jobs, get promoted or take on new work responsibilities, change recreational activities, and so on. Extra-therapy events may be discrete or they may involve chronic situations such as an abusive relationship or the consequences of substance abuse or other problematic behavior patterns. They may also include changes in health status due to physical injuries or illnesses or medical treatments, where these do not directly impinge on psychological functioning. Further, extra-therapy events can contribute both positively and negatively to therapy outcome, and have the potential to obscure the benefits of a successful therapy as well as to make an unsuccessful therapy appear to have been effective. Finally, it is important to consider the bidirectional influence of therapy and life events on one another.

The most obvious method for evaluating the causal influence of extra-therapy events is to ask the client. In the Change Interview, clients are asked what they think brought about changes. If a client does not volunteer extra-therapy events, the interviewer inquires about them. In addition, therapist process notes and session recordings are useful sources of information about extra-therapy events, because clients almost always provide in-session narratives about important positive or negative extra-therapy events. A useful method for locating important extra-therapy

events is to look at weeks associated with reliable shifts in weekly change measures such as the PQ. In addition, as noted in the section on Self-Correction, the Change Interview asks the client to estimate the likelihood that the change would have occurred without therapy.

Extra-therapy events are the major nontherapy counter-explanation in Paul's treatment. When Paul was asked to talk about what he thought had brought about his changes, he spontaneously described the following extra-therapy factors: "Support from my family.... Reading... I have to say my exercise, that's important... New activities. Mainly the jobs." His PQ data reveal one large, clinically significant drop at session 2, and three "spikes," at sessions 4, 24 and 39. Consistent with the drop before session 2, the therapist's process notes describe the client as feeling better, linking this to positive developments in his job and family, as well as a discussion with a friend with similar problems. On the other hand, extra-therapy events had a clear negative influence in the weeks prior to sessions 4 and 25; in both cases, Paul complained of feeling depressed and angry about problems with his severely depressed teenage son, and reported rebuffs from unsympathetic others (his wife and mental health professionals). There was no clear extra-therapy event associated with the spike before session 39, his last session, leaving me to speculate that this was a response to an intra-therapy event -- termination.

Conclusions. Paul's data (including attribution ratings described in the previous section) indicate that extra-therapy factors played a role in his changes, but not to the exclusion of therapy; moreover, based on the weekly PQ data, extra-therapy events appear to have played more of a negative role than a positive one.

7. Psychobiological Causes

The next possibility is that credible improvement is present, but is due primarily to direct, unidirectional psychophysiological or hormonal processes, including psychotropic medications or herbal remedies; the hormonal effects of recovery or stabilization following a major medical illness (e.g., stroke) or after childbirth; or seasonal and endogenously-driven mood cycles. This nontherapy explanation is a major issue given that many clients seeking therapy today are currently taking medications for their mood or anxiety problems. This is a particular problem for psychotherapy research when clients begin or change their medications within a month of beginning psychotherapy, or during the course of therapy.

Assessing medication. The most obvious approach to evaluating psychobiological factors is to keep track of medications, including changes and dose adjustments. It is also important to ask about herbal remedies. (The Change Interview includes questions about both of these.)

Thus, at his posttreatment Change Interview, I learned that one month before the end of therapy, Paul had increased his dose of Celexa to 20 mg/day (he had been taking it for 6 months, after switching from Zoloft). He was also continuing to take Clonazepam for anxiety (2 mg/day) and had been doing so for the past two years. Thus, Paul had been stable on his anti-anxiety medication since well before the beginning of therapy, and had been taking SSRIs for almost as long. Therefore, there appeared to be no connection between changes in his medication and his weekly PQ ratings.

Using in-session narratives. In addition, client interview data and therapist process notes provide useful sources of information about medication and the effects of other medical and biological processes.

For example, at his 6-month follow-up interview, Paul disclosed that he had suffered from a major, life-threatening illness during the intervening time, and had experienced a greater sense of focus and appreciation for what was important as a result.

Conclusions. The evidence for medication or other biological processes on the level of Paul's problems was weak at best, at least during the time he was attending therapy.

8. Reactive Effects of Research

The final nontherapy explanation involves the reactive effects of taking part in research. According to this hypothesis, client outcome is affected mostly as a function of being in research. These include reactive research activities (e.g., PTSD assessment, tape-assisted recall methods) that enhance (or interfere with) therapy; relation with the research staff, which is sometimes better than with the therapist; and enhanced sense of altruism, which allows clients to transmute their suffering by viewing themselves as helping others. On the other hand, research activities can have negative effects on clients, especially if they are particularly evocative or time-consuming.

Teasing out the reactive effects of research on client outcome can be difficult, but qualitative interviewing can help here as well, if clients are asked about the effects the research has on them. Another possibility is to use nonrecruited clients and unobtrusive data collection. Spontaneous comments during sessions, summarized in therapist process notes, are also worth pursuing.

For example, in session 4, Paul expressed concerns at not being able to be totally open in therapy, because of his concerns about the recording equipment. (Several times during therapy, he referred to "All you assholes watching this.") In addition, he sometimes wrote snide comments on his postsession questionnaire.

Conclusions. Paul seemed to take being in the research as more of an inconvenience than a benefit, making it highly unlikely that the research was responsible for the changes he reported.

Summary and Conclusions of HSCED Analysis of Paul's Therapy

Reviewing the results of applying HSCED to Paul's treatment, there is clear or moderate support for 3 out of 5 types of direct evidence, retrospective attribution, immediate perception, and change in stable problems. Since the standard is replication across two or more types of direct evidence, this is more than adequate.

In terms of negative evidence, the standard is that no nontherapy explanation can by itself or in combination with other nontherapy explanations fully explain the client's change, although nontherapy explanations can and usually do play some role in accounting for change. For Paul, there was clear or moderate support against a primary role for all nontherapy explanations, except experimentwise error. The analysis indicates that the change reported on the PQ was unlikely to be due to chance, but identifies Paul's change as narrowly limited to his presenting problems (indicated by lack of change on the SCL-90 and IIP). Self-help, extra-therapy events are also important supporting influences but not to the exclusion of therapy.

Beyond this, however, what have we learned about psychotherapy from this intensive analysis? First, most simplistically, the analysis supports the claim that Process-Experiential therapy can be effective with clients like Paul, i.e., clients with major depressive disorder plus

hypomania ("Bipolar II"), particularly when they present with issues of anger and unresolved grief. Second, although effective, there was still room for improvement, especially with regard to a broader range of problems and areas of functioning. Third, the analysis makes it clear that therapy exerted its helpful effects within a context of other, supporting change processes, especially extra-therapy events and self-help efforts.

Specific Change Processes

Finally, in the process of sorting out the role of therapy in Paul's change process, I came across descriptions of what he found helpful in his therapy, descriptions with substantial practical utility. Some of these took the form of post-session descriptions of significant events, which clearly indicate the central importance that exploring unresolved feelings of anger toward family members. However, Paul's descriptions in his post-therapy Change Interview were more revealing. From examining his discourse, it became clear that Paul did not have a clear "story" about the connection he felt between his therapy and his key change of feeling more calm in the face of challenges. Nevertheless, his account provided enough detail about the therapeutic elements involved and their connections to allow me to construct the following model of his change process:

(a) Paul credited his therapist for "bring[ing] me back to certain areas that she thought I needed to work on, which I might have overlooked"; resulting in (b) "a consistent process of sharing my problems, my frustrations, my heartbreaks"; which (c) "gave me a process of grieving, maybe not all the stages of grief, but some." This grieving process was one of being "able to gradually release it over a year or however long."

As a result of this, he said, (d) "then you see a tangible result. And even before [my nephew's] funeral I went out to my family's graves and I was able to cry." (e) After this, Paul said, he "start[ed] maybe for the first time in a long time, to recognize my progress"; and (f) "that can only help build self-esteem and self-confidence." (g) Finally, Paul implied that this extended grieving/release process had begun to undo his earlier problematic functioning ("I kept a lot of things bottled up [before], and I think that just adds pressure, adds to the anger, adds to the anxiety"); leading to (h) reduced anger and anxiety about hurting other people with his anger ("feeling more calm, and not blowing challenges out of proportion").

This rich account highlights the therapist's main contributions, in the first three steps of the model, as helping the client stay focused on difficult issues; facilitating grieving (trauma retelling and empty chair were used for this); and patiently persisting in this process for an extended period of time (39 sessions). The last five steps primarily show how the client built on the therapy through his own self-help efforts as these interacted with life events such as his nephew's death. The account also supports the conclusions of the hermeneutic analysis by providing a plausible account of the chain of events from cause (therapy) to effect (outcome) (Haynes & O'Brien, 2000).

Issues in Hermeneutic Single Case Efficacy Design

In order to carry out an HSCED study, one needs to (a) find an interesting and agreeable client; (b) collect appropriate measures; (c) apply them to construct a rich case record; (d) analyze the information to see if change occurred; (e) establish if direct evidence linking therapy

to client change is present and replicated; (f) analyze the evidence for each of the eight nontherapy explanations; (g) interpret and weigh the various sets of sometimes conflicting information in order to determine the overall strength and credibility of each nontherapy explanation; and (h) come to an overall conclusion about the likelihood that therapy was a key influence on client change.

HSCED is a new development and clearly needs further testing and elaboration. My team and I have applied HSCED to Paul and other clients seen in our research and training clinic (Elliott et al., 2000; Partyka et al., 2001). What we have learned so far can be summarized as follows:

First, the "Did the client improve?" question has turned out to be more complex than we first thought. Our clients often present us with a mixed picture, showing improvement on some measures and not others, or telling us that they had made great strides when the quantitative data contradicted this (see Partyka et al., 2001). It is important not to underestimate the complexity of this initial step.

Second, this experience has convinced us that more work is needed on how to integrate contradictory information. We need better strategies for determining where the "weight of the evidence" lies (see Schneider, 1999).

Third, we find ourselves in need of additional creative strategies for evaluating nontherapy explanations. For example, to bolster the self-reflective/critical process of examining nontherapy processes, Bohart (2000) has proposed a form of HSCED that relies on an adjudication process involving separate teams of researchers arguing for and against therapy as a primary influence on client change, with final determination made by a "research jury." However, a less involved process might simply make use of two researchers, one (perhaps the therapist) supporting therapy as an important influence, the other playing "devil's advocate" by trying to support alternative explanations. The researchers might present both sides, leaving the final decision to a scientific review process (cf. Fishman, 1999). We are currently testing a form of adjudicated HSCED (Partyka et al., 2001).

Fourth, in comparing HSCED to traditional RCT design, we have found that HSCED requires fewer resources but is in some ways more difficult and demanding in that it requires researchers to address complexities, ambiguities, and contradictions ignored in traditional designs. These complexities are present in all therapy research, but RCTs are able to ignore them by simplifying their data collection and analysis. In my experience, every group design is made up of individual clients whose change process is as rich and contradictory as the clients we have studied. The fact that these complexities are invisible in RCTs is yet another reason to distrust them and to continue working toward viable alternatives which do justice to each client's uniqueness, while still allowing us to determine whether (a) the client has changed, (b) whether these changes have anything to do with our work as therapists, and (c) what specific processes in therapy and in the client's life are responsible for these changes.

Beyond these relatively delimited research applications, HSCED raises broader issues, including the appropriate grounds for causal inference in applied settings, external validity, and the nature of causality in psychotherapy.

Causal inference in the absence of RCTs. It is worth noting that standard suspicions about systematic case studies ignore the fact that skilled practitioners and lay people in a variety of settings continually use generally effective but implicit practical reasoning strategies to make causal judgements about single events, ranging from medical illnesses to lawsuits to acts of

terrorism (see Schön, 1983). For example, legal and medical practice are both fundamentally systems for developing and testing causal inferences in naturalistic situations.

Thus, the task for HSCED is to develop procedures which address various possible alternative explanations for client change. Mechanistic data collection and analysis procedures will not work. Instead, the researcher must use a combination of informant (client and therapist) and observer data collection strategies, both qualitative and quantitative. These strategies confront the researcher with multiple possible indicators of which he or she must make sense, typically by looking for points of convergence and interpreting points of contradiction.

External validity with single cases. Logically, what can be demonstrated by a single case such as the one I have presented is the possibility that this kind of therapy (Process-Experiential, specifically, using primarily empathic exploration and empty chair work over the course of about 40 sessions) can be effective with this kind of client (male, middle-aged, European-American, intellectualizing, psychologically reactant) with this kind of problem (e.g., recurrent depression with hypomanic episodes, unresolved multiple losses, current family conflicts). Predicting how effective a similar therapy would be with a similar client would require a program of systematic replication (Sidman, 1960), and, ultimately, a summary of a collection of similar cases, analogous to precedents established by a body of case law (Fishman, 1999).

Nature of causation in psychotherapy. Another broad issue concerns the kinds of causal processes that are relevant to understanding change in psychotherapy. The following three propositions seem most consistent with how clients change over the course of therapy:

First, change in psychotherapy involves opportunity causes (bringing about change by opening up possibilities to the client), rather than coercive causes (forcing or requiring change). Psychotherapy appears to work by offering clients occasions to engage in new or neglected ways of thinking, feeling and acting; by promoting the desirability of possible changes; and by helping clients remove obstacles to desired behaviors or experiences.

Second, if opportunity causes are the rule in therapy, then, by definition, change in therapy involves multiple contributing causes (“weak” or “soft” causation) rather than sole causes (“strong” or sufficient causation). After all, opportunities are not commands and can always be rejected or simply ignored. Therapist responses in therapy sessions, and even client-therapist interactions in sessions, can provide at best only a partial explanation of client change. Other factors must be assumed to play important roles as well, including extra-therapy life events, biological processes, and especially client internal self-help processes. A complete interpretation of the change process probably requires weaving together the different therapy and nontherapy strands into a narrative such as the one I presented at the end of the analysis section of this paper.

Finally, the development of explanations of therapy outcome is a fundamentally interpretive process, involving a “double hermeneutic” (Rennie, 1999) of client (engaged in a process of self-interpretation) and researcher (engaged in a process of interpreting the interpreter). The double hermeneutic suggests that the client is a really a co-investigator, who acts always as an active self-interpreter and self-changer. As researchers, we follow along behind, performing a second, belated act of interpretation, carefully sifting through the multitude of sometimes contradictory signs and indicators provided by the client. Although we are sometimes weighed down by methodology, nevertheless, it is our greatest desire to understand how our clients change, in order to become more effective in helping them do so.

References

- Barkham, M., Rees, A., Stiles, W.B., Shapiro, D.A., Hardy, G.E., & Reynolds, S. (1996). Dose-effect relations in time-limited psychotherapy for depression. Journal of Consulting and Clinical Psychology, 64, 927-935.
- Bohart, A.C., & Boyd, G. (Dec., 1997). Clients' construction of the therapy process: A qualitative analysis. Poster presented at meeting of North American Chapter of the Society for Psychotherapy Research.
- Bohart, A.C. (June, 2000). A qualitative "adjudicational" model for assessing psychotherapy outcome. Paper presented at meeting of Society for Psychotherapy Research, Chicago, IL.
- Cook, T.D., & Campbell, D.T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally.
- Derogatis, L. R. (1983). SCL-90-R administration, scoring and procedures manual - II. Towson, MD: Clinical Psychometric Research.
- Elliott, R. & James, E. (1989). Varieties of client experience in psychotherapy: An analysis of the literature. Clinical Psychology Review, 9, 443-467.
- Elliott, R., Shapiro, D.A, & Mack, C. (1999). Simplified Personal Questionnaire Procedure. Manual, Department of Psychology, University of Toledo.
- Elliott, R., Slatick, E., & Urman, M. (2001). Qualitative Change Process Research on Psychotherapy: Alternative Strategies. Psychologische Beiträge, 43(3), 69-111. [Also published as: J. Frommer and D.L. Rennie (Eds.), Qualitative psychotherapy research: Methods and methodology (pp. 69-111). Lengerich, Germany: Pabst Science Publishers.]
- Elliott, R., Smith, S., Magaña, C.G., Germann, J., Jersak, H., Partyka, R., Urman, M., Wagner, J., & Shapiro, D.A. (June, 2000). Hermeneutic single case efficacy design: A pilot project evaluating Process-Experiential Therapy in a naturalistic treatment series. Panel presented at meeting of Society for Psychotherapy Research, Chicago.
- Fishman, D.B. (1999). The case for pragmatic psychology. New York: New York University Press.
- Fitts, W.H., & Warren, W. L. (1996). Tennessee Self-Concept Scale (2nd ed). Los Angeles, CA: Western Psychological Services.
- Gosciny, R., & Uderzo, A. (1961/1969). Asterix the Gaul [Trans. A. Bell & D. Hockridge]. London, UK: Hodder Dargaud.
- Greenberg, L.S. (1986). Change process research. Journal of Consulting and Clinical Psychology, 54, 4-9.
- Haaga, D.A.F., & Stiles, W.B. (2000). Randomized clinical trials in psychotherapy research: Methodology, design, and evaluation. In C.R. Snyder & R.E. Ingram, Handbook of psychological change (pp. 14-39). New York: Wiley.
- Hayes, S. C., Barlow, D. H., & Nelson-Gray, R.O. (1999). The scientist practitioner: Research and accountability in the age of managed care (2nd ed.). Needham Heights, MA: Allyn & Bacon.
- Haynes, S.N. & O'Brien, W.O. (2000). Principles of behavioral assessment: A functional approach to psychological assessment. New York: Plenum.
- Horowitz, L.M., Rosenberg, S.E., Baer, B.A., Ureño, G., Villaseñor, V.S. (1988). Inventory of interpersonal problems: psychometric properties and clinical applications. Journal of Consulting and Clinical Psychology, 56, 885-892.

Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology, 59, 12-19.

Julius Caesar, G. (1960). War Commentaries of Caesar (R. Warner, Trans.). New York: New American Library (Original work published 51 B.C.E.).

Kazdin, A.E. (1981). Drawing valid inferences from case studies. Journal of Consulting and Clinical Psychology, 49, 183-192.

Kazdin, A.E. (1998). Research design in clinical psychology (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Kazdin, A.E. (1999). The meaning and measurement of clinical significance. Journal of Consulting and Clinical Psychology, 67, 332-339.

Llewelyn, S. (1988). Psychological therapy as viewed by clients and therapists. British Journal of Clinical Psychology, 27, 223-238.

Mohr, L.B. (Oct., 1993). Causation and the case study. Paper presented at meeting of the National Public Management Research Conference, University of Wisconsin, Madison, WI.

Morgan, D.L., & Morgan, R.K. (2001). Single-participant research design. American Psychologist, 56, 119-127.

Ogles, B.M., Lambert, M.J., & Sawyer, J.D. (1995). Clinical significance of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. Journal of Consulting and Clinical Psychology, 63, 321-326.

Partyka, R., Elliott, R., Alperin, R., Dobrenski, R., Wagner, J., Castonguay, L., Watson, J., & Messer, S. (Nov., 2001). An Adjudicated Hermeneutic Single Case Efficacy Study of Brief Experiential Therapy for Panic Disorder. Panel submitted for presentation at meeting of North American Chapter of the Society for Psychotherapy Research, Puerto Vallarta, Mexico.

Peterson, D.R. (1968). The clinical study of social behavior. New York: Appleton-Century-Crofts.

Piper, W.E. (2001). Collaboration in the new millennium. Psychotherapy Research, 11, 1-11.

Rennie, D. L., (1999). Qualitative research: A matter of hermeneutics and the sociology of knowledge. In M. Kopala & L.A. Suzuki (Eds.), Using qualitative methods in psychology (pp. 3-13). Thousand Oaks, CA: Sage Publications.

Schneider, K.J. (1999). Multiple-case depth research. Journal of Clinical Psychology, 55, 1531-1540.

Schön, D.A. (1983). The reflective practitioner: How professionals think in action. New York: Basic Books.

Sidman, M. (1960). Tactics of scientific research. New York: Basic.

Strupp, H.H., Horowitz, L.M., & Lambert, M.J. (eds.) (1997). Measuring patient changes in mood, anxiety, and personality disorders: Toward a core battery. Washington, DC: APA.

Notes

¹In this, I had hoped to be aided by the scientific incarnation of Asterix and Obelix, those two intrepid (but fictional) adventurers who, according to the graphic novels of Goscinny and Uderzo (1961), so plagued Julius Caesar in his attempts to conquer Gaul. Unfortunately, copyright restrictions severely limit the use of these characters, and especially modified forms of their names (e.g., "Hermeneutix" and "Critico-Reflectix").

²This is not to say that "enriched" RCTs (Piper, 2000) cannot be used to identify specific causal process, only that carrying out the essential elements of an RCT does not provide this understanding.

Table 1
Outcome Data for Client PE-04 (Paul)

Measure:	Cas e- nes s	RC Min (p < .2)	Pre	Post	Pre- post diff.	6-month follow- up	Pre-6 month diff.
SCL-90-R GSI	.93	.51	1.17	.80	.37	.47	.70+
Inventory of Interpersonal Problems-26	1.5 0	.57	1.69	1.62	.07	.96	.73+
Personal Questionnaire	3.0	.53	4.44 (4.1 1) ^a	2.44 (2.7 8)	2.00+ (1.33+)	2.22	2.22+ (1.89+)

Note. Caseness: cut-off for determining whether client is clinically distressed; RC Min: minimum value required for reliable change at $p < .2$; "+" = Reliable improvement from pretherapy. Sources for values given: Barkham et al. (1996; Inventory of Interpersonal Problems, Personal Questionnaire); Ogles, Lambert & Sawyer (1995; SCL-90-R GSI).

^aUsing median of three successive weekly scores to smooth instability of scores.

Table 2
Indirect Evidence: Methods for Evaluating the Presence of Nontherapy Explanations

Non-Change/Non-Therapy Possibility	Efficacy hypothesis	Evaluation Methods
1. <u>Nonimprovement</u> . 1a. Apparent changes are <u>trivial</u> .	Changes are important.	<ul style="list-style-type: none"> •Calculate reliability and clinical significance on outcome measures. •Look at client manner and detail for indications of importance. •Ask about importance of changes.
1b. Apparent changes are <u>negative</u> .	Changes are in positive direction.	<ul style="list-style-type: none"> •Ask client, therapist about negative changes. •Test for reliable negative change on outcome measures.
2. <u>Statistical artifacts</u> . 2a. Apparent changes reflect <u>measurement error</u> .	Changes are statistically reliable.	Assess for reliable change (RCI calculations).
2b. Apparent change reflects <u>outlier or regression to mean</u> .	Change is discontinuous with stable baseline.	<ul style="list-style-type: none"> •Use multiple pretests. •Assess duration/stability of client problems.
2c. Apparent change is due to <u>experimentwise error (fishing expedition)</u> .	Change is general .	Replicate change across multiple measures (global reliable change).
3. <u>Relational artifacts</u> . Apparent changes are superficial attempts to please therapist /researcher.	Change exists independent of relational issues.	<ul style="list-style-type: none"> •Look for specific or idiosyncratic detail. •Ask client about negative as well as positive descriptions of therapy. •Assess client tendency to respond in socially desirable manner.
4. Apparent changes are result of <u>client expectations</u> (therapy "scripts") or <u>wishful thinking</u> .	Change exists independent of client expectations or hopes.	<ul style="list-style-type: none"> •Evaluate stereotyped vs. idiosyncratic nature of the language used to describe changes. •Ask client whether changes were expected vs. surprising.
5. <u>Self-correction</u> . Apparent changes reflect self-help and self-limiting easing of short-term or temporary problems.	Changes would not have occurred without therapy.	<ul style="list-style-type: none"> •Assess duration/stability of client problems (by multiple pretests or retrospectively). •Assess client-perceived likelihood of changes without therapy. •Look for evidence of self-help efforts begun before therapy.
6. Apparent changes can be attributed <u>extra-therapy life events</u> , such as changes in relationships or work.	Influence of therapy is separate from or interacts with life events.	<ul style="list-style-type: none"> •Look for extra-therapy events which might have influenced changes. •Assess client-perceived likelihood of changes without therapy. •Consider mutual influence of therapy

		and life events on one another.
7. <u>Psychobiological factors</u> . Apparent changes can be attributed to medication or herbal remedies, or recovery from medical illness.	Influence of therapy is separate from or interacts with medical factors.	<ul style="list-style-type: none"> •Collect information on changes in medication or herbal remedies. •Consider role of recovery from illness as possible cause.
8. Apparent changes can be attributed to <u>reactive effects of research</u> , including relation with research staff, altruism.	Research enhances but does not separately influence outcome.	<ul style="list-style-type: none"> •Ask client about effects of research. •Use nonrecruited clients, unobtrusive data collection.

Figure 1
PQ Means Across Sessions: PE-04

