Protecting the Integrity of the Legal System:
The Admissibility of Testimony from Mental Health Experts Under Daubert/Kumho Analyses

William M. Grove
University of Minnesota

R. C. Barden
National Association for Consumer
Protection in Mental Health Practices


Abstract

We discuss to what degree testimony from social science and mental health experts (psychologists, psychiatrists, social workers, therapists and others) meets admissibility requirements expressed by the U.S. Supreme Court in Daubert, Joiner and the recent Kumho decision. We review data on Daubert/Kumho indicia of reliability, emphasizing several areas commonly addressed in civil and criminal trials: repressed and “recovered” memories of childhood sexual abuse, personality assessment by means of the Rorschach and other “projective” assessment techniques, as well as the diagnoses of post-traumatic stress disorder and multiple personality disorder (dissociative identity disorder). We conclude that much (and perhaps most) of the testimony presently offered by mental health professionals in these areas should not survive scrutiny under the framework of Daubert, Joiner and Kumho, and argue that such controversial, unreliable testimony imperils the integrity of our legal system. We recommend increased attention to scientific methodology and analysis in the education of legal and mental health professionals. We note that mental health professionals and social scientists could become increasingly helpful to the legal system as expert testimony becomes more firmly grounded in reliable scientific evidence.

Protecting the Integrity of the Legal System:
The Admissibility of Testimony from Mental Health Experts Under Daubert/Kumho Analyses

Prior to the ruling of the U.S. Supreme Court in Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U. S. 579 (1993), testimony from mental health and social science experts—regardless of how unscientific and unreliable such testimony might be—was largely unregulated by the legal system. Although the Frye standard had been in place for decades, junk science testimony by mental health professionals regarding ink blot tests, post traumatic stress disorder, multiple personality disorder and other controversial issues was almost uniformly admitted without objection by counsel or the court (see Frye v. United States, 54 App. D.C. 46, 47, 293 F. 1013, 1014, for the rule that expert opinion based on a scientific technique is inadmissible unless
the technique has achieved "general acceptance in the particular field in which it belongs"; see also People v. Hughes, 59 N.Y.2d 523, 453 N.E.2d 484, 1983; People v. Zayas, 131 Ill. 2d 284, 546 N.E.2d 513, 1989 [using Frye to exclude hypnotically refreshed testimony]; Note 1).

The 1993 Daubert ruling of the U.S. Supreme Court changed this unfortunate situation and heightened interest in, and concern about, expert testimony based on "junk science" (Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 1993). In Daubert, the U.S. Supreme Court ruled that scientific expert testimony is admissible only if it is both relevant and reliable. The Court further held that the Federal Rules of Evidence "assign to the trial judge the task of ensuring that an expert's testimony both rests on a reliable foundation and is relevant to the task at hand" (id., at 597). In addition, the Court discussed specific factors including falsifiability (and actual testing), peer review, error rates, and "acceptance" in the relevant scientific community, some or all of which might prove helpful in determining the reliability of a particular scientific "theory or technique" (id., at 593–594). This analysis generally follows the philosophy of science model offered by Sir Karl Popper (1959). Below we discuss these factors.

Clarifying the review process enunciated in Daubert, the U.S. Supreme Court later ruled that the court of appeals must apply an abuse-of-discretion standard when it reviews the trial court's decision to admit or exclude expert testimony (General Electric Co. v. Joiner, 522 U. S. 136, 138-139). The Joiner ruling further emphasized the responsibility of the trial court to fulfill a mandatory gatekeeper role—the duty to exclude unreliable expert testimony—assigned by the Daubert ruling.

In a March 1999 decision with enormous importance for the regulation of the testimony of mental health professionals, the U.S. Supreme Court expanded the Daubert analysis to the testimony of essentially all expert witnesses (Kumho Tire, Inc. v. Carmichael, 119 S.Ct.1167 (1999). In Kumho, the defendant tire company moved to exclude the plaintiffs' expert testimony on the ground that his methodology failed to satisfy Federal Rule of Evidence 702, which says: "If scientific, technical, or other specialized knowledge will assist the trier of fact, a witness qualified as an expert may testify thereto in the form of an opinion" (emphasis added). Granting the motion to exclude the expert testimony in question and entering summary judgment for the defendants, the District Court acknowledged that it was acting as a reliability gatekeeper as required by Daubert.

Erroneously concluding that Daubert was limited to "scientific" testimony, the U.S. 11th Circuit Court of Appeals held that the Daubert factors did not apply to the expert's testimony, which it attempted to distinguish as "skill- or experience-based." The U.S. Supreme Court overturned the 11th Circuit's interpretation and reinstated the District Court opinion by holding that the Daubert factors may apply to the testimony of engineers and other experts who are not claiming a basis for their testimony in rigorous scientific research and peer reviewed publications (Kumho, pp. 7–13). In an opinion written by Justice Breyer—a justice highly trained in methodology and philosophy of science issues—the Court ruled that the Daubert "gate keeping" obligation applies not only to "scientific" testimony, but to all expert testimony. This landmark case should have broad and powerful effects on expert testimony by mental health, social science and psychotherapist experts.

With obvious and profound long range implications for all "clinical" expert witnesses, the Kumho court ruled that Federal Rules of Evidence 702 and 703 grant all expert witnesses, not just "scientific" ones, testimonial latitude unavailable to other witnesses—on the assumption that the expert's opinion will have a reliable basis in the knowledge and experience of his discipline. In addition, the Court ruled that it would prove difficult, if not impossible, for judges to administer evidentiary rules under which a gate keeping obligation depended upon a distinction between "scientific" knowledge and "technical" or "other specialized" knowledge, since there is no clear line dividing the one from the others and no convincing need to make such distinctions (Kumho, pp. 7–9).

As Justice Breyer, writing for the court, noted:
There is no convincing need to make such distinctions between "scientific" and "clinical" experts. Experts of all kinds tie observations to conclusions through the use of what Judge Learned Hand called "general truths derived from _ specialized experience." Hand, Historical and Practical Considerations Regarding Expert Testimony, 15 Harv. L. Rev. 40, 54 (1901). And whether the specific expert testimony focuses upon specialized observations, the specialized translation of those observations into theory, a specialized theory itself, or the application of such a theory in a particular case, the expert's testimony often will rest "upon an experience confessedly foreign in kind to [the jury's] own." Ibid. The trial judge's effort to assure that the specialized testimony is reliable and relevant can help the jury evaluate that foreign experience, whether the testimony reflects scientific, technical, or other specialized knowledge.

We conclude that Daubert's general principles apply to the expert matters described in Rule 702. The Rule, in respect to all such matters, "establishes a standard of evidentiary reliability." 509 U. S., at 590. It "requires a valid ... connection to the pertinent inquiry as a precondition to admissibility." Id., at 592.

Having applied the Daubert analysis to virtually all expert witnesses, the Kumho court continued to struggle with testimony offered by "experiential" experts: We agree with the Solicitor General that "[t]he factors identified in Daubert may or may not be pertinent in assessing reliability, depending on the nature of the issue, the expert's particular expertise, and the subject of his testimony." (Brief for United States as Amicus Curiae 19.)

At the same time, and contrary to the Court of Appeals' view, some of Daubert's questions can help to evaluate the reliability even of experience-based testimony. In certain cases, it will be appropriate for the trial judge to ask, for example, how often an engineering expert's experience-based methodology has produced erroneous results, or whether such a method is generally accepted in the relevant engineering community. Likewise, it will at times be useful to ask even of a witness whose expertise is based purely on experience, say, a perfume tester able to distinguish among 140 odors at a sniff, whether his preparation is of a kind that others in the field would recognize as acceptable.

To say this is not to deny the importance of Daubert's gatekeeping requirement. The objective of that requirement is to ensure the reliability and relevancy of expert testimony. It is to make certain that an expert, whether basing testimony upon professional studies or personal experience, employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field. (Kumho, March 1999; emphasis added.)

We note that the Kumho analysis regarding the importance of analyzing all expert testimony for reliability using Popperian factors is an historic, far reaching improvement in the legal system. We further note that permitting "expert" testimony based on "personal experience" without a methodological analysis of reliability would permit astrologers, ouija board analysts, astral projection evaluators and UFO "experts" to testify if they showed the "same level of intellectual rigor that characterizes the practice of an expert in the relevant (e.g. astrological) field."

This loophole has been duly noted. Writing for the Court, Justice Breyer closed the door on clinical "experts" relying upon inadequate methodologies: As the U.S. Supreme Court ruled in Joiner, "nothing in either Daubert or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the ipse dixit of the expert." 522 U. S., at 146. (Kumho, March 1999.)

In the Daubert, Joiner and Kumho cases the U.S. Supreme Court has begun the long-overdue process of educating legal professionals in the essential, minimal characteristics of science. The reliable exclusion of junkscience testimony will only be possible when attorneys and judges become familiar and comfortable with the process of "thinking like a scientist." Six factors of scientific analysis, or indicia of testimonial reliability, can be distinguished in Daubert and we have numbered these factors as follows:
1. Is the proposed theory or technique, on which the testimony is to be based, testable (i.e., can it be falsified, in Popper's (1959) sense)?
2. Has the proposed theory, method or procedure been tested using valid and reliable procedures?
3. Has the theory or technique been subjected to peer review?
4. What is the known or potential error rate of the scientific technique?
5. What are the standards controlling the technique's operation that maximize the reliability and validity of the technique?
6. Has the theory or technique been generally accepted as reliable and valid in the relevant scientific community?

Hereafter, we will call these considerations “testability,” “testing,” “peer review,” “error rate,” “standards,” and “general acceptance.”

As other contributors to this issue have pointed out (e.g., Krauss and Sales), these features were not enumerated as an exhaustive list. Further, the Daubert court did not require trial judges to combine these factors algorithmically in deciding on admissibility, nor did they assign weights to the factors. Hence, it has been left to case law to clarify the proper application of Daubert. The reader is referred to other articles in this issue where these features are discussed from a legal point of view (e.g., Lipton; Schopp, Scalora & Pearce) and pertinent cases are analyzed. Bersoff, Glass, Dodds, Eckl and Peters (1999) have compiled a list of federal appellate Daubert cases.

Daubert admissibility hinged on acceptability of a theory or methodology, and not on conclusions drawn from it. However, the Joiner court held that “[c]onclusions and methodology are not entirely distinct from one another; when the analytical gap between the data and the opinion proffered is simply too great [the evidence may be excluded]” (General Electric v. Joiner, 118 S. Ct. 512, 1997). Therefore, a seventh factor should be added to those above: do the expert’s conclusions reasonably follow from applying the theory to this case?

Kumho further extends Daubert in an historic and essential way by holding that a scientific analysis of reliability applies even when non-scientific expert testimony is offered. Henceforth, it should not avail the expert to claim that what they do is not “science-based” or to raise the pseudo-distinction between “hard” and “soft” sciences. We note that science is simply systematized, reliable knowledge and that unreliable, unsystematized information in the form of “expert” opinion should not be permitted to influence the legal rights of citizens.

In several controversial areas of forensic mental health testimony, admissibility under Daubert-Joiner analyses has finally become an important issue; we predict that Kumho analyses will also prove crucial. As a clinical psychologist-researcher and an attorney who is also a researcher and clinical psychologist, the present authors are particularly interested in expert mental health testimony relating to specific individuals, e.g., symptomatology, diagnosis, and personality traits; prognosis or prediction of future behavior or adjustment; and the reliability of fact witness testimony. Although the most egregious examples of what we regard as expert witness misconduct seem to appear in child custody cases, given present space limitations, we will cover three areas where there is significant scientific or professional literature, admissibility case law, or both: repressed or “recovered” memories of childhood abuse; the Rorschach test; and diagnoses of post-traumatic stress disorder and multiple personality disorder. We note that the present analysis certainly does not exhaust this important subject. For example, intentional or negligent expert witness misconduct and misinformation most commonly offered by “experts” in the family law system are important issues that will be addressed in subsequent publications.

Repressed and Recovered Memories of Abuse

This important controversy has attracted the attention of some of the most prominent social scientists in the world. Leading experts in a number of related fields have expressed strong concern about the harms caused by the pseudoscientific theory of “repression and recovered memories” (Loftus, 1997; Loftus & Ketcham, 1994; Ofshe & Singer, 1994; Ofshe & Watters, 1996; McHugh, 1992, 1995; Pope & Hudson,
Some of the most prominent figures in psychology and psychiatry have sounded the alarm in the scientific and legal communities with regard to the dangers posed by the junk science theories regarding the “repression” or “dissociation” of abuse “memories.”

The theory of repressed memories rarely enters the courtroom unless it is also claimed that the “repressed” memory was eventually “recovered”. Hence, we treat the concepts of “repressed” and “recovered” memories as one. This theory arises in several kinds of litigation. First, the theory may be used in an attempt to toll the statute of limitations, by claiming that a plaintiff repressed a trauma memory and could not have filed suit in a timely manner. Second, some malpractice litigants allege their therapist negligently encouraged or implanted false, ostensibly “repressed” abuse memories. In defending such suits, therapists often offer expert witnesses who testify that, far from being incompetent practice, the process of recovering repressed memories in therapy is buttressed by a generally accepted scientific theory of repression.

We note that the term “repression” is not always used; other common labels include “traumatic” amnesia, “psychogenic” amnesia or “dissociated memories”. The Diagnostic and Statistical Manual (DSM-IV) of the American Psychiatric Association (1994) includes a diagnostic category, dissociative amnesia, which is also sometimes used to label this purported phenomenon.

Before we can decide whether the theory of “repression,” “dissociation,” or “traumatic amnesia” for child abuse is testable, we need to state clearly what the theory means. Despite innumerable pages being spent debating the merits of this hypothesis, it seems that there is little general agreement on precisely what the hypothesis is. We would accept the following formulation, whether one believes the mechanism is best called “repression,” “dissociation,” “robust repression” (Ofshe & Singer, 1994) or something else: “repression” occurs when someone experiences a traumatic event that would be expected to be memorable. Subsequently, the victim experiences prolonged, complete inability to recall the traumatic event. For such forgetting to be counted as “repression,” the non-recollection must not be explainable by uncontroversial causes such as normal infantile amnesia, head injury, toxic states, or ordinary forgetting (Pope, Oliva & Hudson, 1999).

By “ordinary forgetting” we mean the type of recall and recognition failure studied for non-traumatic material in experimental psychology laboratories since Ebbinghaus, encompassing memory trace erosion as well as retrieval failure. We note that it is characteristic of ordinary forgetting that it is typically overcome by enhancing motivation to recall, greater effort, or providing reminders (Schachter, 1996, pp. 79–80).

1. Is this theory testable? In principle, yes. However, it can be made all but untestable by failing to define key terms, only vaguely specifying them, or by redefining key terms in a data-evading, post hoc way. It appears to us that advocates of the repression concept, in courtroom testimony, sometimes appear to trade on ambiguity in definitions. One can, for example, try to blur the distinction between ordinary forgetting and repression, more than warranted by facts, and then try to claim that evidence for forgetting is evidence for repression. In this category we would put some citations of studies that demonstrate forgetting the details of a traumatic event without loss of memory for the gist (e.g., citations contained in Brown, Scheflin & Hammond, 1998). For example, concentration camp survivors sometimes forget specific episodes of their wartime torments. However, they apparently do not forget that they were imprisoned and tormented in the camp (Wagenaar & Groeneweg, 1990).

2. Has this theory been tested? Yes—if by this one means that it has been tested in any way that could conceivably falsify it. (Pope, Hudson, Bodkin & Oliva, 1998) (summarizing studies related to all forms of trauma). However, most of the studies cited in support of this concept are not strong tests, i.e. they lack many of the principal features of a high-risk Popperian test. In fact, only a few studies possess even most of the features of such a test: publication via demanding peer-review; prospective study design; carefully and objectively determining the occurrence of
abuse; a representative sampling of subjects; designs which can convincingly exclude competing explanations for forgetting (e.g., reluctance to discuss the abuse). Only one study has essentially all these features; it offers no support for the repression hypothesis as all of the victims remembered the abuse (della Femina, Yeager & Lewis, 1990).

Defenders of the repression theory (e.g., Brown, Scheflin & Hammond, 1998) offer quite a different—although we believe non-mainstream—view about the strength of the evidence for repression, from substantially the same body of research. Brown et al. place great weight upon studies which few researchers and virtually no courts have accepted as credible, e.g., unpublished convention papers. One example of such idiosyncratic analysis is Brown, et al’s reliance upon a study of 3 childhood “amnesia” victims as support for the notion that children are likely to “repress” memories of abuse. (See Brown et al’s chart at page 164, citing Dollinger, 1985). Even a cursory review of the research study’s original text (an essential but rarely conducted part of preparing for cross examination) reveals that, in fact, two of the three children in the Dollinger study who suffered “amnesia” were hit and knocked unconscious by side flash lightning; the third child was killed. Most importantly, the entire large group of neurologically uninjured children who witnessed the lightning strike suffered no amnesia of any kind or degree. Like the other thousands of neurologically unimpaired victims of trauma studied over the past decades they also remembered the horrifying event all too well. Clearly, in the light of rational analysis, this study offers no evidence at all in support of “traumatic amnesia” for child abuse. As the case law records indicate, such pro-repression analyses contain misperceptions and errors of logic and methodology that provide large targets for science-informed cross-examination (see Barden, 1996, 1998).

In contrast to the small, seriously flawed body of research which purports to demonstrate complete “repression” of trauma memories (e.g., Herman & Schatzow, 1987), a vast array of peer reviewed, published research stretching over decades demonstrates that trauma is generally all too well remembered by its victims. These studies fall into several patterns: (a) studies in which the subjects report no amnesia at all; (b) studies in which the subjects report only partial amnesia for details but no global amnesia—this includes studies that fail to even distinguish global vs. partial amnesia (thus confusing cases where only details of a trauma were forgotten versus the total amnesia postulated by "repression" theory); (c) studies that fail to separate reports of amnesia in very young children—thus confounding normal infantile forgetting with repression; or (d) studies that do not specify whether all amnesia victims were subjected to head trauma, toxins, or other known physical causes of amnesia. This vast literature on trauma victims does not document or even support notions of global, persistent amnesia, for objectively verified trauma events, in persons older than age 4 who have not suffered cerebral insult. In sum, decades of research on repression and trauma sequelae, as well as extensive reviews of the relevant peer reviewed research, lead to the consensus of the relevant scientific community that “There is no empirical (scientific) evidence to support either repression or dissociation [of trauma memories]....” (Brandon et al., 1998, p. 304). (For extensive reviews of this literature see Pope, Hudson, Bodkin & Oliva, 1998 (concerning “repression” of all forms of trauma); Kendall-Tackett Williams & Finkelhor, 1993 (abused children); Beitchman, Zucker, Hood, DaCosta et al., 1992 (abused children); (Brandon et al., 1998 (general review); Pope, Olivia & Hudson, 1999 (extensive legal/scientific analysis).

3. Has the repression theory been supported by peer-reviewed studies? Although some would argue yes for a few studies, such studies are seriously flawed. For example, in the oft-cited Herman and Schatzow study, the goal of the therapy group under study was apparently to recover “repressed memories.” In addition, no reasonable corroboration was obtained for the patients’ “memories” of abuse. Finally, a number of the patients’ memories in this study were apparently contaminated by previous treatment with “recovered memory” therapists. In sum, the methodology of such studies cannot even distinguish between victims of actual
childhood abuse who later "recovered memories" and victims of suggestive, improper therapy practices.

The fact that such flawed studies were published in peer-reviewed journals is, in our view, cause for concern. Clearly, courts evaluating Daubert-Joiner-Kumho admissibility must pay attention not only to the fact of peer review, but also the quality of the journal involved and its peer-review processes. In the information age, there are many outlets for publication, including publication of arguably unsound research. For example, when a theory is not generally accepted, journals can be created as outlets for non-mainstream views; or existing journals can come to be seen as favored venues for such views. This arguably happened in the fields of multiple personality disorder and repression with the journal Dissociation. Many questions were raised about the clinical methods and practices of some editors and contributors to Dissociation, including federal criminal grand jury investigations, malpractice lawsuits, and state licensing board actions. Subsequently, in a recent newsletter of the International Society for the Study of Dissociation (the sponsoring organization for Dissociation), its President wrote: "[t]he [Society] is in crisis_ [regarding the journal Dissociation] we understand that its editor has decided to suspend publication" (Barach, 1999, p. 1). It is currently unknown what, if any, connection exists between the public controversies regarding the practices of some editors and contributors of Dissociation, on the one hand, and the suspension of publication, on the other.

Given that much of the "pro-repression" research was published by a rather small group of individuals and frequently appeared in non-mainstream, controversial journals like Dissociation, it is of grave concern that as-yet unanswered questions exist regarding possible scientific misconduct in the compilation and publication of some "pro-repression" research. For example, one noted researcher in this area has testified that one of his key research assistants pled guilty to fabricating data, in a plea agreement with federal investigators and the Harvard University Office of Scientific Integrity. (Barden, 1996, at pg 313). Another well-known theorist was criminally prosecuted on charges related to "recovered memory therapy" (United States of America v. Judith A. Peterson, Ph.D., et al., 1998; Smith, 1998). Other authors in this area have suffered adverse action against their professional licenses, or have such actions pending (Grumman, "Controversial Doctor Faces Loss of License: State Allege Psychiatrist Abused Position of Power," Chicago Tribune, August 13, 1998; In the Matter of Renee Fredrickson LP 2653, 1999). A federal court ordered production of the original data for one widely cited paper on repression. Rather than produce these data, the author refused to comply with the court’s order and withdrew from the case. (Smith v. O'Connell, et al., C.A. No. 93-0615-T, Federal District of Rhode Island, December 27, 1996). Researchers, clinicians, attorneys and judges may need to be particularly vigilant with regard to research findings published in some "specialty" peer-reviewed journals, especially when the underlying data are not available for review.

4. The most problematic Daubert criterion for the repression theory, as for many other controversial psychological theories and techniques, is the one concerning the known or potential error rate. Before examining this question, we note that the Daubert court did not specify an acceptable rate of error. Although rates of error that fall below coin-tossing should be clearly unacceptable, determining acceptable rates of error is a more difficult case by case task.

A rational analysis of error rates resulting from "repressed memory" evidence requires several steps. First, one needs to define the relevant kinds of error. False positive errors are problematic because claims of repression are often cited to toll the statute of limitations in personal injury and criminal cases. For example, a court may erroneously accept a "recovered" and heretofore "repressed" "memory" as substantially accurate, and hence toll the statute of limitations, enabling a junkscience lawsuit or criminal prosecution to go forward. False negative errors are equally troubling. They may result in an abuse victim being unfairly blocked from receiving justice, or recovering damages, in the justice system. To help courts assess the risk of such errors we need to know the false positive and false negative error
rates involved in accepting the repression theory as valid. We also need to know the base rates of sexual abuse and of repression. Knowing the relevant error rates will enable us to begin to calculate the “positive predictive value,” which tells the court how often it is likely to err if it relies on “repressed memory” testimony.

Almost no papers in the literature appear to have squarely addressed the critical question of error rate. Fish (1998) attempts to do so, but makes numerous fatal errors. First, he bases calculations on unverified assumptions about which form of Bayes Theorem applies to this problem. Second, he inserts numerous arbitrary assumed figures into his calculations: “I assume that 39% of clients who are not asked but have a CSA [child sexual abuse] history will spontaneously report it,” “I assume that clinicians actually routinely inquire of clients [about CSA history] 100% of the time,” etc. (Fish, 1998, p. 208). Third, he combines figures on the rate of reported child sexual abuse from one group of studies (e.g., Fish, 1998, last paragraph on p. 215), with ostensible rates of “repression” (our term, not his; see below) from another set (Fish, 1998, pp. 216–217). This is unjustified unless the two kinds of studies (on sexual abuse and repression, respectively) sample the same population, and define sexual abuse in the same way. These conditions are highly unlikely to be met generally, and certainly are not provably met here.

Fourth, Fish blurs the critical distinction between forgetting and repression. He calls the phenomenon we have labeled “repressed and recovered memories” and others have called “dissociated memories” by the nonce word “delayed memories.” As this term was coined de novo, there is no peer reviewed literature supporting its existence. In addition, it appears that no court has ever ruled on the Daubert admissibility of “delayed memories.”

Without research support, Fish goes on to speculate that subjects who don’t report abuse during interviews (e.g., subjects in the Williams, 1994 study) and subjects who report “memories” of abuse “recovered” in psychotherapy (e.g., subjects in the Williams, 1995 study) necessarily demonstrate the same phenomenon. He further speculates that these “memories” are what others have called “repressed” or “dissociated.” (Williams, 1994, does not state that her subjects “repressed” memories of abuse.)

Fish’s error rate appears quite unreliable, for at least two reasons. First, the false positive and negative error rate estimates are based chiefly on only two studies (Dalenberg, 1996; Williams, 1995). These studies contrast the ostensible accuracy of two kinds of reported memories. The first kind comprises reported memories of alleged abuse, retrospectively reported as having been temporarily forgotten. Alas, such reported memories may include (in unknown proportions) false or mistaken reports of forgetting, ordinarily forgotten memories, memories forgotten due to physical causes (e.g., blows to the head), and possibly repressed memories as well. Fish assumes all these memories were “delayed” (which by implication he substantially equates with “repressed”). This heterogeneous mixture is then contrasted with a group reporting never-forgotten abuse memories. The studies report no real difference in ostensible accuracy between these two types of reported memories. Both studies, however, suffer from numerous, serious methodological flaws when their results are used for Fish’s purpose. We mention only one problem here to save space: the distinction between “continuous” (never-forgotten) and “recovered” memories is entirely based on a subjective, uncorroborated one-time report about whether and when the subject remembers the event of forgetting. Answers to this odd question: When do you remember forgetting the abuse? seem unlikely to be highly reliable. If subjects made such reports inaccurately, then “noise” is introduced into the group membership determination. On this basis alone, one would predict precisely the finding obtained: no material difference in accuracy between unreliably assessed “continuous” and unreliably assessed “recovered” memories.

A second flaw with Fish’s speculative error rate calculation is that it assumes repression exists, in direct contradiction to a wide array of sound, published research. If the empirical research on trauma victims thus far is accurate, then the error rate of a report of a “repressed and recovered” memory will range upwards to 100%,
whether the ostensible memory arises in therapy or elsewhere. If repression does not exist, then all reports that a subject repressed (or dissociated), then recovered, an abuse memory, are necessarily 100% inaccurate. Until it is established to the scientific community’s general satisfaction that repression exists, error rate calculations cannot be reliable in the legal sense. All we really know is that the error rate of relying on repressed and recovered memories cannot be zero, and could well be as high as 100%. Courts are rightfully loath to restrict the civil liberties of citizens (parental rights, custody decisions, criminal convictions, etc) based upon such capricious information.

To obtain a reliable estimate of the error rate for repressed and recovered memories of traumatic events, one would need to do each of the following. First, one must establish that “repression” or “dissociation” of memories of objectively verified traumatic sexual abuse actually occurs (following design criteria like those suggested by Pope, Oliva & Hudson, 1999). These findings should be replicated in independent laboratories, in studies conducted by researchers of varied persuasions regarding this issue. Second, one must establish that an identifiable group of subjects really repressed memories in their entirety, for prolonged periods; demonstrations of partial or extremely transient forgetting will not do. Such studies would need to be prospective and longitudinal, so that the timing and extent of forgetting could be documented. Such studies would need to show that any amnesia was real, not feigned, and that it could not be accounted for by ordinary forgetting (e.g., the memories couldn’t be retrieved by standard cueing). Third, one would need to show that a group of these proven “repressors” later recovered their memories. Finally, one would need to show that a sizable proportion of such provably repressed, then recovered memories were reasonably accurate. For example, one would need to set up objective accuracy criteria: can the subject accurately identify perpetrators, identify approximate dates of offenses, and accurately identify the nature of the criminal acts?

The field is simply nowhere near meeting these requirements today. As discussed above, studies with reasonable methodological safeguards indicate that victims of trauma, including child abuse, generally remember the abuse all too well (see Pope, Hudson, Bodkin & Oliva, 1998, regarding forgetting all sorts of trauma; Kendall-Tackett, Williams & Finkelhor, 1993, Beitchman, Zucker, Hood, DaCosta et al., 1992, both on abused children; Brandon et al., 1998, a general review; and Pope, Olivia & Hudson, 1999, extensive legal/scientific analysis). Hence, any contemporary claim that we can accurately estimate the error rate of “recovering” accurate “memories” of trauma is clearly false and should be rejected at all levels of the legal system.

5. The next Daubert consideration is what we call “standards.” Here, there are no well-characterized methodological standards, which when followed ensure reliable application of the repression theory. It is sometimes claimed on the one hand that certain therapy or interview techniques lower the risk of inducing false or distorted memories, or on the other hand that suggestive therapy or interview techniques raise the risk. Fish (1998) and Brown et al. (1998) address these arguments from a pro-repression point of view, while Ofshe and Watters (1990) and Loftus and Ketcham (1994), among others, review them from the skeptics’ perspective. Clearly, the necessary and sufficient conditions for reliable vs. unreliable recollection of autobiographical events are not well characterized at the present time. This is because demonstrating a correlation between an experimental manipulation and subjects’ tendency to accept suggestive and misleading information, is logically quite different from demonstrating a minimum level of suggestiveness, below which false or distorted memory reports are highly unlikely to occur. Also, laboratory experiments on suggestion and memory do not and cannot (for ethical reasons) employ the kind of massive, persistent, and coercive persuasion which some individuals may experience before reporting questionable or false memories.

In fact, suggestive and coercive techniques reportedly used by some clinicians to elicit “recovered” memories have been roundly criticized by mainstream professional groups (e.g., Brandon, Boakes, Glaser & Green, 1998). A number of professional association pronouncements generally warn clinicians not to rely on
uncorroborated reports of “recovered” sexual abuse memories (Barden, 1998). This seems to be good advice for the courts as well.

6. As to “general acceptance” in the Daubert sense, it is important to distinguish this from the more simplistic, vague Frye analysis. Under Frye, the court is to determine “general acceptance in the particular field”—leaving the meaning of “field” quite unclear. Under Daubert, the court must determine acceptance within the “relevant scientific community.” Given the Supreme Court’s Popperian concept of science, all should agree that, the Daubert “relevant scientific community” does not include clinicians, or at least it only includes them to the extent that they also follow the methodological safeguards of the scientific method (Popper, 1959). Kumho makes the applicability of Daubert Popperian analyses to clinicians quite clear.

Some experts have confused the boundaries of the “relevant scientific community.” In the repression controversy, this kind of confusion can severely distort the Daubert analysis. Depending on how the community is demarcated, opinions about “general acceptance” may vary. For example, psychotherapists studying trauma victims’ “memories” (e.g., Terr, 1995) more often accept the repression hypothesis than do experimental scientists in the community of human memory researchers (e.g., Garry & Loftus, 1994). In addition, social scientists working in the area of coercive influence (attitude manipulation) often see “recovered memories” as classic examples of coercive manipulation—not involving “memories” at all (Ofshe & Singer, 1994). Trauma researchers, memory researchers and influence researchers are all most definitely appropriate parts of the relevant scientific research community on the repression controversy. It is clear that within this relevant scientific community, the theory of repressed or dissociated abuse memories is not generally accepted and is, in fact, generally repudiated (Pope, Olivia & Hudson, 1999). As one piece of evidence in favor of this conclusion, we note the American Psychological Association Working Group on Investigation of Memories of Childhood Abuse. This group was comprised of three general memory researchers and three trauma clinician/researchers. In the end, these two groups agreed on little that was specific and legally pertinent. In fact, they issued dissenting reports, each sharply critical of the other group’s position (American Psychological Association Working Group on Investigation of Memories of Childhood Abuse, in press).

In a more empirical approach to the question of general acceptance among therapists and scientists, a recent paper by Pope, Oliva, Hudson, Bodkin and Gruber (1999) provides survey data on the acceptance of “dissociative amnesia” as described in the DSM. “Dissociative amnesia” is similar to the repression concept as we have defined it (McDougall, 1920). In Pope et al.’s broad survey of board-certified psychiatrists, 35% of respondents thought dissociative amnesia belonged in the next DSM revision “without reservations”; 48% thought it warranted inclusion only with reservations; and 9% thought it should be excluded outright. This is not, be it noted, a study of a scientific community (those aware of valid and reliable Popperian methodologies), because most psychiatrists are not scientific researchers. Although respondents who had published scientific papers did not offer substantially different responses than psychiatrists who had not published, these data appear to indicate that researchers do not generally accept the current rendition of the concept. In any event, these data demonstrate that the psychiatric controversy regarding repressed memories is far from over. Courts would be wise to exclude such a controversial, unreliable concept as required by Daubert/Kumho.

In fact, virtually all courts looking carefully at this issue have found that theories regarding “repressed and recovered memories,” “dissociative amnesia,” “traumatic amnesia” and related concepts and procedures used to “recover memories” fail to meet minimal standards of reliability and should be excluded as “junk science.” The most lengthy and thoroughly litigated examination of this issue in U.S. history was recently decided in Rhode Island. Weeks of testimony, seven experts, several fact witnesses, and six attorneys including one psychologist/attorney were involved. The court ultimately excluded all expert testimony supporting the notion of “repressed memories,” noting: “The State has not met its burden of establishing that repressed recollection is reliable and admissible as

Exceptions to this tidal wave of rejection include two cases sometimes cited as support for inclusion of repressed memory testimony. The first, a Michigan case (Isley v. Capuchin, 877 F. Supp 1055 (E.D. Mich 1995), demonstrates the confusion that can result when critical distinctions are missed. The court confused the existence of Post Traumatic Stress Disorder with the reliability of “repressed memory,” an extraordinary error that, thankfully, no other court appears to have made. The second exception, Shazade v. Gregory (930 F. Supp. 673, D. Mass. 1996), involves a different problem. A pro-repression expert witness testified in a manner that was, in our opinion, misleading and incomplete, in describing “pro-repression” research. His testimony is discussed in a full chapter of Hagen’s Whores of the Court (Hagen, 1997; see also Barden, 1996, Faigman, et al, 1999). The court found that the relevant scientific community should be restricted to psychotherapists only— in direct contradiction to Daubert. No other court appears to have made this error. This case was never appealed; others that have been appealed have routinely led to appeals court rulings against the Frye/Daubert admissibility of “repressed memories.” In
sum, these cases do not argue for inclusion of repression testimony so much as for thoroughly prepared, highly skilled cross-examination of experts by attorneys well versed in science, and reliance on well-qualified and knowledgeable rebuttal experts.

Projective Testing for Diagnosis

There are many kinds of projective tests, including the Thematic Apperception Test, projective drawing tests, and others. In our opinion, there is scant empirical support for any of these tests. No projective technique has as much supporting research as the Rorschach. If it can be shown that testimony based on the Rorschach, when used for diagnosis of mental disorders, is inadmissible under Daubert, it will follow that none of the other projective techniques are likely to be admissible, either.

In dealing with the Rorschach, we would prefer to focus attention on Exner’s The Rorschach: A Comprehensive System (TRACS; Exner, 1993) method of administration, scoring and interpretation as this is the more thoroughly researched of the Rorschach scoring procedures. This would arguably put the Rorschach on its best footing. However, a restriction to Exner’s TRACS system is not feasible for two reasons. First, only a minority of the Rorschach literature has carefully adhered to TRACS scoring and interpretation. Second, many clinicians, especially senior practitioners, don’t use TRACS (Exner, 1980). Therefore, we are forced also to consider pre-TRACS literature on the Rorschach.

Daubert concerns scientific theories. However, the Rorschach is not a theory but a measurement technique. Although early advocates of the Rorschach proposed a specific theory to account for how the Rorschach worked, namely the projective hypothesis, it is not necessary to accept this hypothesis in order to use the Rorschach. Hence, we address Daubert issues surrounding a related theory: The Rorschach, when properly administered, scored, and interpreted, yields accurate diagnoses of psychological disorders. Where pertinent, we also touch on the literature concerning the Rorschach’s ability to measure personality traits, as some of these traits are related to psychological disorders.

1. It is this technique testable? Yes. It’s relatively straightforward to determine whether an instrument yields scores that are valid, in the sense of notably correlating with external criteria including psychiatric diagnoses.

2. Has this technique been tested? Yes, both for zero-order validity and for incremental validity. By the former, we mean tests of the Rorschach’s ability, considered in isolation, to diagnose psychological disorders and assess personality traits. By incremental validity, we mean the Rorschach’s ability to add to diagnostic accuracy or personality description, when added to other commonly obtained information (chart data, interviews, MMPI). The interpretation of zero-order validity studies is currently controversial, but we read the literature as suggesting that the Rorschach scores have low validity, typically in the .2–.3 range, somewhat lower than the MMPI (Garb, Florio & Grove, 1998). Incremental validity studies quite consistently show zero to negative validity for the Rorschach (Garb, 1985; Grove, 1986).

One could argue that because the Rorschach has appreciable validity, it passes this Daubert test. However, it is more important to remember that without incremental validity, the Rorschach—on average—adds nothing to diagnostic evaluations. Indeed, the Rorschach can be expected to detract from testimonial accuracy, if it has no incremental validity. This is because it will enhance confidence without enhancing validity. Psychologists (and other clinicians) often show increased confidence in their judgments when they have more information rather than less; but their judgments often don’t grow more accurate as more information accrues (Garb, 1980); in particular, the Rorschach apparently does not enhance diagnostic accuracy. Ignorant but confidently expressed opinions from a testifying expert are likely to mislead a court and thus potentially endanger the liberties of citizens. By rational analysis from its zero incremental validity, the Rorschach is a forensic failure in the area of psychological diagnosis.

3. Is the Rorschach generally accepted as reasonably valid? No; it is quite controversial and always has been. Courts are encouraged to analyze years of negative Rorschach reviews in authoritative sources including Buros’s Mental
4. What is the error rate associated with Rorschach interpretations? This question is not easily addressed in this context, because it would be expected to vary across applications. Unquestionably, the Rorschach can yield highly inaccurate diagnoses. For example, Little and Schneidman (1959) suggest the error rate can be 100%; in their report, the modal diagnosis given to psychiatrically normal individuals by Rorschach experts was one of psychosis. Albert, Fox and Kahn (1980) compared malingerers and psychotics on Rorschachs interpreted by nominated experts. They found 52% diagnostic errors among psychotics, 46–72% errors with malingerers, and 24% with normals. That is, the Rorschach missed over half of the psychotics, and mislabeled a quarter of quite normal individuals as not merely disturbed but psychotic, calling more fakers than real patients sick! On the other hand, if the Rorschach F+% (or X+%%) score is used, then one may do better than chance in diagnosing psychosis; but the error rate will still be substantial. Alas, experts seldom if ever testify solely on the basis of the F+% score. For non-psychotic diagnoses (e.g., depression), the Rorschach’s validity is also quite weak (Wood et al., 1996). [Xxx cite for Wood article on diagnosis here xxx] The adjudicated rights of citizens should not turn on such an error prone technique.

5. Has the validity of the Rorschach been subjected to peer review? Yes, no and maybe. The assessment community has been surprised to learn in recent years that many of Exner’s supporting studies were apparently never peer-reviewed; in fact, many do not even exist as actual written reports, and their data is not readily available to other investigators (Wood et al., 1996). Furthermore, pro-Rorschach studies often appear in a specialty journal, the Journal of Personality Assessment (originally called the Rorschach Research Exchange), which may potentially relate to quality of publication. Finally, the quality of peer review has been questioned with regard to reports in the Rorschach literature (e.g., Wood, Nezworski, Stejskal, Garven & West, 1999).

6. Standards: from the increasing popularity of Exner’s scoring and interpretation system, it might seem that using TRACS is a well-recognized way of enhancing the Rorschach’s accuracy. However, this has not been satisfactorily demonstrated. Hiller, Rosenthal, Bornstein, Berry and Brunell-Neuleib (in press) claim to find evidence for superior validity of TRACS when compared to other systems. However, they appear to be misinterpreting a non-significant difference (two-tailed p < .175) as favoring TRACS.

One example will show how there can be a large gap between the ostensible validity of the Rorschach and conclusions drawn from nonstandard uses of it. Recently, the first author consulted on a case where a psychologist used a Rorschach, along with other tests, to conclude that a young mother had “incipient Munchhausen by Proxy (MBP) syndrome.” That is, even though the mother had not even allegedly harmed her baby, it was recommended that her child be taken away lest she some day develop this syndrome, which is characterized by making the child ill. The psychologist claimed that the Rorschach was a valid instrument for making this “diagnosis.” However, “incipient MBP” doesn’t exist in the literature (nor is it in DSM-IV). No scientific studies support this idiosyncratic use of the Rorschach. Such abuses of the legal system would be prevented by appropriate Daubert/Kumho scrutiny.

McCann (1988) has published a more sanguine Frye-Daubert admissibility review. Guidelines for ostensibly appropriate use of the Rorschach in court were offered by Meloy (1991), and Weiner, Exner and Sciara (1996) noted that the Rorschach has seldom been excluded in court (they noted only 6 challenges, one of them successful, out of 7934 cases employing Rorschach testimony). This is consistent with our experience that very few attorneys have ever conducted Daubert (or even Frye) hearings to exclude unreliable, junkscience testimony.

Notwithstanding the success of the Rorschach in eluding challenge to date, and despite Meloy’s advice, we find the use of the Rorschach to be quite problematic under Daubert/Kumho. We believe the analysis in McCann (at pp. 133–140) is seriously flawed. McCann fails to explain how Rorschach-generated impressions of a
client are falsifiable, a fundamental Daubert criterion. He also misinterprets the
continuing controversy over this test as somehow constituting evidence for its
general acceptance. (A controversy can continue long after most disputants have
reached a negative opinion.) McCann also errs in relying on Exner’s (1993) self-cited,
unpublished reports, which as discussed above are often not peer-reviewed or even
physically extant manuscripts. McCann likewise relies on unreplicated and
unrepresentative accuracy figures for selected Rorschach variables in concluding that
the Rorschach has Daubert-acceptable error rates. In addition, McCann fails to
mention in his “error rate” section a study which was, at that time, the best review
of Rorschach accuracy: Parker, Hanson and Hunsley (1988). This meta-analysis
yielded two average Rorschach validities: .42 for “convergent” validity studies and
.07 for other studies. Skeptical of this report, the first author and colleagues
reanalyzed Parker et al.’s data (Garb et al., 1998), finding an average validity of .30
for a 35% Rorschach error rate. Such error rates are unacceptable for tests used in
litigation over parental rights and other “high stakes” matters. In sum, such
unreliable evidence presented to unsuspecting jurists as a “scientific psychological
test” poses an unacceptable risk to the integrity of the legal system.

Controversial Diagnoses (PTSD and MPD/DID)

We consider just two diagnoses here: post-traumatic stress disorder (PTSD)
and multiple personality disorder (MPD, now called Dissociative Identity Disorder or
DID in DSM-IV). First, we note that PTSD and MPD are labels, not theories. This is a
critical distinction as some experts have misled courts to falsely believe that the
existence of a diagnostic label in DSM-IV somehow proves general acceptance of the
existence of the described disorder as well as proposed causal mechanisms for the
etiology of such a disorder. This is a very serious error of logic and method. The DSM-
IV is simply an agreed upon set of terms and descriptions—a catalog. It does not
provide, and was not intended to provide, documentation of the general acceptance
of the existence of disorders. Further, the DSM-IV is not, in any way, documentation
general acceptance of the etiology (cause) of a disorder. To clarify with an
example, just because the word unicorn is in the dictionary and just because we all
agree on the concept and description of a unicorn as defined in the dictionary does
not, in any way, document the existence of unicorns.

Even the DSM labels may not be generally accepted, just because they are in
the manual. Consider an analogy: the Volstead Act was on the books for years after
public support for it had dramatically eroded. With regard to the DSM, no Gallup poll
of scientists is taken to decide whether to put a label into the manual. Instead, a
small (sometimes very small!) committee votes for a label, and this largely ensures its
inclusion.

Prof. Paul McHugh, Chairman of Psychiatry at Johns Hopkins Medical School,
clearly and accurately explained this critical issue before the court in Quattrocchi
(McHugh, 1998, pg 530-531):

15 Q. (Dr. Barden) As a psychiatrist, as one who trains psychiatrists
16 at Johns Hopkins Medical School, is the fact that the
17 term “dissociative amnesia” is found in the DSM any
18 evidence at all of the scientific validity of the
19 condition?
20 A. (Dr. McHugh) No, it’s not. DSM-IV is an attempt to, as
21 DSM-III was, is an attempt to develop reliability
22 among psychiatrists about what they are observing.
23 It is not intended to say that they are claims for the
24 existence of a particular condition as confirmed by

PAGE 531
1 its enclosure within DSM-IV. It’s a question of
2 reliability versus validity.

***

PAGE 636
1 (Dr. McHugh): It will take me only a
2 minute or two, Your Honor, to explain it. The
reason—psychiatry is the only discipline that runs itself on a catalog, and the reason it did that was that back in the 19—late 1960s and early 1970s, it was discovered that they couldn’t do much research in psychiatry because they couldn’t get agreement on what they were observing, let alone what they were calling things. So a patient with a set of symptoms called schizophrenia in Baltimore was called manic depressive in London and demoralized in San Diego. It was the decision of the American Psychiatric Association and of the American psychiatric community that we should build a catalog of reliability, not of validity, so that we could do research. So that when we said these patients are showing these symptoms that we are going to call Dissociative Disorder or schizophrenia, we could have a common language. After that, we could discover whether they had, in fact, anything like dementia, or whether they were behaving in a particular way for other reasons. So when I’m saying this (DSM) is a book of reliability, I’m saying this gives you a code that whereby you can describe people whom you say will be given—will satisfy the criteria in this book (DSM) for Dissociative Disorder, but whether Dissociative Disorder exists in itself is not proven or even claimed in this book (DSM).

To analyze these theoretical conditions with Daubert/Kumho reasoning one must consider that these diagnostic labels are generally linked to accompanying theories of etiology: namely, that these psychological disorders are caused by psychological trauma. In the case of PTSD, this includes a wide range of upsetting events; for MPD, the cause is purported to be early, severe (and almost always repressed) abuse, chiefly sexual abuse (Putnam, 1989). The concept of PTSD is testable in principle. However, to obtain a test placing the diagnosis at risk of falsification, one would need a benchmark for accuracy. Since many, even most, forensic PTSD evaluations now use structured interviews and claim to adhere to DSM criteria, validity studies will seldom have a benchmark better than the forensic diagnosis itself. Spitzer (1983) proposed a method for validating diagnoses that could be useful for PTSD: the “LEAD standard,” denoting Longitudinal, Expert, and All-Data based diagnoses. Unfortunately, this useful strategy has apparently never been used to assess the concept of PTSD (or MPD).

Testing the theory that PTSD is caused by exposure to trauma is also possible: observe people exposed vs. not exposed to an objectively verified, naturally occurring traumatic event, and find out whether the traumatized group gets PTSD more often then the unexposed. (Ethical principles preclude a direct experimental test.)

2. Has the causal theory of PTSD been tested? Yes, but the tests have been weak because of ethical and practical problems. Tests to date have yielded variable, mostly weak confirmations demonstrating the unreliable nature of the concept (Bowman, 1997). For example, McFarlane (1987; 1988a-d) found that just 9% of individuals’ symptoms could be accounted for by exposure to a disastrous fire, while Galante and Foa (1987) found no relationship between proximity to an earthquake and children’s behavioral disturbances. By contrast, pre-existing symptoms may predict as much as half of the variance in post-trauma symptoms (Nolen-Hoeksema & Morrow, 1991). Only for those exposed to the severest, most prolonged traumas (e.g., prisoners of war) does the rate of PTSD rise to one-half or more (Engdahl, Dikel, Eberly & Blank, 1997). In a review of 45 studies Kendall-Tackett, et al (1993) reported that abused children displayed more symptoms than non-abused children with abuse accounting for 15% to 45% of the variance (we note that none of the
thousands of abused children studied were reported to have “repressed” memories of
the abuse).

One might argue that if any symptoms are significantly associated with
trauma, this makes PTSD testimony probative. However, the relevant legal question
here will generally be whether or not a trauma was a “substantial factor” in creating
the disorder. With traumas often accounting for only a small percent of symptoms, a
psychologist may seldom be able to offer good faith, science-based “reasonably
certain” testimony that a trauma is the “cause” of a patient’s PTSD symptoms. For
example, if “reasonably certain” means “more probable than not,” then for a trauma
to be a “reasonably certain” cause of PTSD, ceteris paribus more than half of
sufferers of this particular type of trauma would need to develop PTSD. However,
causation is seldom so straightforward. Correlational studies relating traumas to
symptoms often show that third variables (such as other adversity, pre-existing
symptoms, or coping styles) can account for much of the apparent trauma-symptom
connection (Beitchman, et al, 1992; Kendall-Tackett, et al, 1993). If this is so, then
theories like “PTSD is caused by trauma” cannot be fairly characterized as having
survived strong risks of falsification. In sum, the causal connection between trauma
and PTSD appears too unreliable to survive a thorough Daubert/Kumho analysis.

3. Is the theory of PTSD generally accepted? The label is more controversial
than some (e.g., schizophrenia) but less so than others (e.g., MPD). We believe that
the syndrome is generally but by no means universally accepted.

However, as noted above, acceptance of the label need not imply acceptance of the
causal theory. PTSD positively invites misunderstanding on this score; its diagnostic
requirement of a trauma event invites experts to commit the post hoc ergo proper hoc
fallacy, assuming that the trauma caused the PTSD. The very name “post-traumatic
stress disorder” seems to imply what the criteria do not state, namely causality.

If we consider an analogy from the depressive disorders, the fallacy will be
clearer. If DSM-V erected a “reactive depression” category, with diagnostic criteria
requiring a negative life event plus certain depressive symptoms, would this justify
the conclusion that the negative event caused the symptoms? Of course not. In
fact, negative life events have little proven relation to depression and “neurotic”
ailments (Tennant, 1983; Tennant, Bebbington & Hurry, 1981).

4. What is the known or potential error rate of a PTSD diagnosis? A relevant
error rate would seem to be the frequency with which a court would err, if it equated
a PTSD diagnosis with proof of causation. If symptoms are only 9% predictable from
trauma exposure as McFarlane found, this would correspond to an error rate of 35% in
causal inferences (assuming a 50% base rate of actual trauma exposure). Since we
don’t know the true rate of trauma exposure suffered by plaintiffs, it is fair to say
that the error rate is unknown. However, it is important to note the error rate would
go up from 35% if the actual trauma exposure rate were substantially less than 50%;
near a rate of zero, the error rate would obviously approach 100%. Clearly, such
uncertain and high error rates should not survive thorough Daubert/Kumho scrutiny.

5. With regard to peer review, many studies of the relationship between
trauma and psychiatric symptoms have been published in mainstream journals (e.g.,
Archives of General Psychiatry) as well as in speciality journals (e.g., Journal of
Traumatic Stress, Child Maltreatment, Child Abuse and Neglect). However, as
mentioned above many of these studies report only weak support for the trauma-
illness connection.

6. A major problem with standards in diagnosing DSM-IV PTSD is the vagueness
of its “A” criterion: “The person has been exposed to a traumatic event to which
both of the following were present: (1) the person experienced, witnessed, or was
confronted with an event or events that involved actual or threatened death or
serious injury, or a threat to the physical integrity of self or others; and (2) the
person’s response involved intense fear, helplessness, or horror.” The inclusion of the
phrase “threat to the physical integrity of self or others” has the potential to allow a
biased or careless “expert” to over-diagnose PTSD. Also, criterion A(2) is obviously
quite subjective, as are most of the symptoms of PTSD. Malingering these symptoms
would not be difficult. Even though a PTSD diagnosis requires actual (not just
claimed) trauma, it has been the authors’ observation that many forensic clinicians do not even attempt to corroborate the trauma claim, nor make the PTSD diagnosis conditional on court’s determination of the accuracy of the trauma claim. This standards problem is especially striking given that, in many PTSD claims, it seems that the parties dispute all or part of the trauma claim.

If things are problematic for PTSD diagnoses under Daubert, they are much worse for MPD/DID. Like PTSD, the underlying theory (not stated in the criteria) is that severe abuse causes the disorder. For MPD the assumed abuse seems almost always to be sexual in nature, supposedly beginning in early to middle childhood (Putnam, 1989).

In assessing the credibility of the MPD hypothesis an assessment of the full range of theories regarding the etiology of MPD is important. Courts must be truthfully informed that variants of the theory of MPD -- crafted by the central founders of MPD theory and practice -- are unsupported by any credible evidence. Such theories have generated considerable controversy and litigation. For example, some of the leading proponents of MPD have blamed "international satanic cults" and "CIA mind control experiments" for the dramatic increase in the reported incidence of MPD.

One national leader of the early MPD movement, has offered one such theory of the etiology of MPD symptoms that is unsupported by any research or other evidence. This theory posits that MPD is often the product of "programming" by "intergenerational, international satanic cults". To an audience of psychotherapists, in a question and answer session, he offered the following hypotheses:

Q: Audience: What's the difference between this kind of program and cult-type abuse and Satanic abuse in the kind of cults with the candles and the...
A: Dr. Hammond: This type of programming will be done in the cults with the candles and all the rest. My impression is this is simply done in people where they have great access to them or they're bloodline and their parents are in it and they can be raised in it from an early age. If they are bloodline they are the chosen generation. If not, they're expendable and they are expected to die and not get well. There will be booby traps in your way if they aren't non-bloodline people that when they get well they will kill themselves. I'll tell you just a little about that. My belief is that some people that have ritual abuse and don't have this have been ritually abused but they may be may be part of a non-mainstream group. The Satanism comes in the overall philosophy overriding all of this. People say, "What's the purpose of it?" My best guess is that the purpose of it is that they want an army of Manchurian Candidates, ten of thousands of mental robots who will do prostitution, do child pornography, smuggle drugs, engage in international arms smuggling, do snuff films, all sorts of very lucrative things and do their bidding and eventually the megalomaniacs at the top believe they'll create a Satanic order that will rule the world."

(Hammond, C. 1992)

Similarly, a former president and co-founder of the International Society for the Study of Dissociation,” has offered his own theory of how MPD is often the result of abuse by "satanic cults":

[Dr. Bennett Braun] told the audience that children are often abused in day-care centers as a way of preparing them to join the cult. "You have to predispose the nervous system to this sort of behavior," he said. After the children are abused in day care, they are "then picked up in high school" and indoctrinated into the cult. The cult, he has come to understand, is networked with the Ku Klux Klan; neo-Nazi groups; the Mafia; big business; the intelligence community, including the CIA; and the military. He told the audience that he has developed twelve P’s for those involved in satanic abuse: "Pimps, Pushers, Prostitutes, Physicians, Psychiatrists, Psychotherapists, Principals and teachers, Pallbearers [meaning undertakers], Public workers, Police, Politicians and judges, and Priests and clergies from all religions.”

(Ofshe & Watters, 1996, p. 245)
Dr. Braun was recently a co-defendant in the largest psychotherapy negligence settlement in history, in a case alleging that Dr. Braun imposed his unique theories of abuse by "satanic cults" on a patient and her children using hypnosis and other means. Although Dr. Braun denied any negligence the plaintiff received a $10.6 Million settlement (Belluck, 1997). It is important to note that Dr. Braun served as an Editor of the journal Dissociation for years. Several therapists who have lost lawsuits and their medical license relied upon this "international satanic cult" theory of the etiology of MPD and applied it- with disastrous results- to patients. (Barden, 1994) It is essential that courts, in reviewing research in this field understand the social milieu in which MPD theories were constructed. Especially with regard to specialty journals, for courts to thoroughly assess the quality of a journal’s peer review process, it may well be necessary to explore the knowledge, training, experience and judgment of the peer reviewers.

Yet another former president and co-founder of the International Society for the Study of Dissociation has written a book proposal in which he explains his theory that MPD cases may be caused by CIA-military brainwashing experiments (Ofshe & Watters, 1996, p. 223). Dr. Colin Ross also testified about similar theories in Hamanne v. Humenansky (July 25, 1995, Ramsey County, MN, Case C4-94-203). Another well known therapist testified that she learned about links between the CIA and the "satanic cults" from Dr. Ross. (Barden, 1994 at pg. 131)

Even the most cursory analysis of these various MPD theories- put forth by national leaders in the MPD field-indicates that such theories will not survive Daubert (or Frye) scrutiny. In addition, these unusual theories alert courts to be skeptical of research published in specialty journals- edited by the creators of these unusual theories-that support particular diagnoses and points of view. Questions should be raised about whether articles published in such journals are as reliable as those published in mainstream journals, e.g., the APA clinical psychology journals, AMA journals and mainstream APA psychiatry journals. It would be irresponsible for experts or courts to conduct Daubert/Kumho analyses of MPD as a theory and diagnosis without incorporating and reviewing information regarding the complete social history of MPD theories.

Less lurid variants of the theory that MPD results from child abuse are also unlikely to survive Daubert/Kumho analysis.

1. As a causal theory, the hypothesis that MPD is caused by severe child abuse is testable in principle. A prospective longitudinal study of abused children could be conducted to yield evidence about whether severe child abuse significantly increases the risk of MPD symptoms. The use of proper controls, similar to the traumatized children in many relevant ways other than trauma exposure, would of course be crucial.

2. However, the theory has not been tested in this straightforward fashion. All so-called “tests” to date are retrospective and based on self-reported abuse histories. Since the diagnosis and the abuse histories both have unknown error rates, there is to date no good evidence that MPD is ever caused by severe child abuse, let alone that it is routinely so caused. To exhibit good evidence for the abuse theory, one would have to convincingly demonstrate the following: unmistakable signs of MPD (and not equivocal signs such as are shared with many other disorders), such as blatant alter switching, are prospectively observed in early to middle childhood; objectively verified, severe abuse preceded all these incontestable signs of MPD; and concomitant variables (e.g., non-abuse adversity, family history of psychopathology) can be convincingly ruled out as causes of the MPD symptoms. We know of no peer-reviewed, published (let alone independently replicated) study that even approaches meeting these simple methodological requirements. In a review of studies on abused children, the authors concluded there was insufficient evidence of any link between child abuse and “multiple personality disorder” (Beitchman et al, 1992).

3. Some studies of MPD/DID are peer reviewed, but many of these appear in specialty journals like Dissociation and the Journal of Traumatic Stress. Much of the material is in the form of books and conference presentations, which are either not peer-reviewed at all, or are typically reviewed much less stringently than are most
journal submissions. Material skeptical of MPD is more likely to be published in mainstream journals (e.g., British Journal of Psychiatry).

4. It appears to us, from our reading of numerous books and papers by MPD advocates as well as an extensive review of numerous MPD clinical cases, that almost all forensic MPD patients’ childhood sexual abuse “memories” (and all “memories” of Satanic ritual abuse; Ross, 1995) are “recovered” (i.e., ostensibly formerly “repressed”) in regressive psychotherapy—almost all following therapy using hypnosis or related techniques. Hence, the MPD-abuse link is apparently no stronger than the case for repression, addressed above. In any case, there appear to be no well-conducted longitudinal prospective studies showing that severe abuse precedes MPD, let alone that such abuse causes MPD.

5. We have the gravest concern about the circumstances and frequency with which MPD diagnoses may be misapplied. There are only two main criteria for MPD/DID in the current DSM: two or more personalities that alternately control the patient’s behavior; and amnesia. We have observed the diagnosis of MPD being made by clinicians who interpret a patient’s varying moods, inconsistent attitudes, and varied likes and dislikes as “alter personalities” where another, more skeptical clinician would not see these behavioral changes as anything out of the ordinary. We have seen reports of clinicians telling patients that any significant forgetting of childhood events (including for events before age 2!) constituted “amnesia,” and then using that “symptom” to qualify a patient for an MPD diagnosis. Indeed, supporters of the validity of MPD report that patients often do not manifest any MPD-type behavior when first seen (Kluft, 1984). Most have been treated by various clinicians and for several years under other diagnoses (Coons, Bowman & Milstein, 1988; Putnam, Guroff, Silberman, Barban, et al, 1986), with the MPD diagnosis often only made after months of contact with the patient, and sometimes made without the symptoms being present (Kluft, 1984).

Clearly, under the current state of the art “experts” cannot offer responsible testimony to courts about what conditions favor accurate MPD diagnoses. To begin with, we will need much clearer, objective behavioral or physiological criteria. MPD advocates and skeptics alike must generally accept these criteria. Advocates and skeptics will also need to be able to routinely agree on their presence or absence, in individual cases. Until that time, expert testimony supporting MPD diagnoses should not survive an appropriately conducted Daubert/Kumho attacks on “standards” grounds.

6. As far as we can tell, MPD/DID is one of the least generally accepted of all DSM diagnoses. There are extremely heated debates about the “reality” of MPD/DID, with some defenders (Bliss, 1986; Cohen, Berzof & Elin, 1995; Kluft & Fine, 1993; Putnam, 1989; Ross, 1997) and many skeptics (Aldridge-Morris, 1989; Merskey, 1992; North, Ryall, Ricci & Wetzel, 1993; Ofshe & Watters, 1994; Piper, 1997; Spanos, 1996). We prefer to think of the controversy not in terms of the “reality” of MPD, but rather in terms of its prevalence (near-zero, vs. much higher) and its etiology. As a diagnostic label, MPD may conceivably serve as a reasonable label for a certain behavioral syndrome. This syndrome may be rare or more common. It remains to be determined whether it is primarily caused by abuse (Putham, 1989), by misguided assessment and therapy techniques (Merskey, 1992; Ofshe & Watters, 1994; Piper, 1997), by media influences (North et al., 1993; Ofshe & Watters, 1994), or is typically feigned or represents a role enactment (Spanos, 1996).

The recent Pope et al. (1999) paper, cited above in connection with repression or dissociative amnesia, also asked U.S. psychiatrists about MPD. About a third (35%) of respondents favored including MPD in the next DSM “without reservations”; 15% thought it should not be included at all; and 43% thought it should be included “only with reservations.” An earlier report by Dell (1988) stated that “a total of 52% of the psychiatrists, 80% of the psychologists... stated that other professionals had overtly told them that 'there is no such thing as multiple personality' [disorder]” (pg. 529) Clearly there is and has been no general acceptance of the current or previous versions of the MPD diagnosis even among general mental health professionals.
It has been claimed by a few that the relevant scientific community for MPD testimony is limited to those who publish research on numerous MPD patients. This superficially plausible requirement reflects several serious errors. First, MPD skeptics—many highly respectable scientists and clinicians—are highly unlikely to treat many MPD patients as they view the disorder as created by improper therapy that focuses attention on MPD symptoms. Hence, like most clinicians in general, MPD skeptics are not in a position to see, let alone publish, research on MPD patients. Prior to the recent explosion of the MPD diagnosis, most clinicians went a lifetime without ever making this diagnosis. Thus, with MPD, as with other potentially improper diagnostic labels, skeptics who doubt the existence of the disorder will not spend years studying a mistaken diagnostic label. Courts should rely upon experts in diagnosis, assessment, measurement, uses of the DSM and the history of psychiatry to provide guidance as to the reliability of new and controversial diagnostic labels.

A second problem with such a restriction is that pro-MPD researchers often publish in “specialty” journals such as the now apparently defunct Dissociation—a fact that arguably generates skepticism about the quality of these papers. Third, of the few clinicians who have reported treating dozens to hundreds of MPD patients, many have been successfully sued for negligently creating, diagnosing and/or treating MPD symptoms, and encouraging or even implanting “recovered memories” of sexual abuse in these very patients. (Bellack, 1997; Guthrey & Kaplan, 1996) Further creating doubts about the reliability of peer review in this specialty field, a number of nationally prominent MPD, repressed memory experts have been sanctioned, prosecuted or de licensed by governmental agencies. (Barden, 1997 at pg 33,34; USA v. Peterson, 1998; In the Matter of Renee Fredrickson LP 2653, 1999; Grumman, 1998; Warmbir, 1999; Langner, 1995).

We believe it is clear that the relevant scientific community for an appropriate Daubert/Kumho analysis of MPD would include: researchers on disorders frequently present before, after, alongside, or arguably instead of MPD in ostensible MPD patients (e.g., affective disorders, anxiety disorders, personality disorders); researchers who study MPD patients; developmental psychologists studying children’s responses to trauma, experts in the practice and methodology of diagnosis and assessment, experts in the history of other diagnostic misadventures in psychiatry, experts in human memory (because MPD supposedly involves amnesia), experts in social influence (such research documents the process by which MPD could be the product of improper suggestion), and experts in sociology (whose research suggests that MPD, and theories about “international satanic ritual abuse,” are simply modern examples of social hysteria).

Finally, it is up to those proposing a controversial theory to win over the scientific and legal communities with convincing data. Those proposing a high rate of MPD necessarily suggest that there are up to millions of “hidden cases” of MPD coming to light. Such extraordinary claims require extraordinary evidence. (Piper, 1997) Those advocating the more extreme abuse-etiology theories (CIA "mind control programming", "satanic cults") must explain how these conspiracies so successfully and uniformly escape forensic detection when rapists, “ordinary” child molesters, serial killers, the Mafia, and bungled White House-directed burglaries have not. Pro-MPD theorists need to satisfactorily explain why thousands of trauma victims have been longitudinally studied over the past several decades, and such individuals do not develop MPD. (Beitchman et al, 1992; Kendall-Tackett, et al, 1993)

Investigators in the fields of child psychology, sociology, history of psychiatry, general clinical psychiatry and psychology and memory research commonly share a simpler alternative explanation for MPD—improper suggestion, chiefly by zealous and poorly trained therapists. It is apparent that state licensing boards increasingly agree with this perspective. The theory that MPD is the result of repressed and dissociated memories of horrific abuse is no more reliable than the theory of repression that so many courts have now excluded as junkscience. We conclude that the MPD concept, and especially the abuse theory of MPD etiology, does not currently come close to general acceptance under Daubert/Kumho analyses. Summary and Recommendations
The foregoing has implications for courts, expert witnesses and attorneys. Following Daubert/Kumho, federal judges are now on notice by the U.S. Supreme Court that they bear an affirmative duty to actively exclude junk science testimony and thus protect the integrity of the legal process. We hope that state courts will quickly follow this overdue analysis, as the admission of irrelevant, unreliable, junkscience “expert” testimony is contrary to public policy and endangers the integrity of the legal system. Too many citizens have been harmed by inappropriate, junkscience “expert” testimony; proper implementation of Daubert/Kumho analyses will go far towards correcting this serious social problem.

Experts should be aware, based on the literature we have cited, that a significant portion of mental health-related social science testimony may be unreliable and unlikely to withstand a well-conducted Daubert/Kumho hearing. Moreover, in our view experts have an affirmative ethical duty to refuse to give testimony that would not reasonably be expected to pass Daubert/Kumho scrutiny. This is true even if opposing counsel fail to challenge its admissibility. In addition, experts have a similar duty to accurately report to judges and juries the kind of research we have summarized. This is true even if opposing counsel fails to introduce such research. We say this because professionals are bound by their oath to tell the whole truth; trying to “slip one by” opposing counsel is hardly that. Moreover, unethical testimony will inevitably bring the profession into disrepute (American Psychological Association Code of Ethics, 1992; APA Division 41 Specialty Guidelines for Forensic Psychologists, 1991; Hagen, 1997).

Experts wishing to practice competently in a well-conducted Daubert/Kumho hearing will find the new environment a spur to improving their testimony about complex science issues. By contrast, nonchalant experts in Daubert/Kumho cross-examinations may reveal culpable technical and ethical errors, documentable on the record, which may later haunt them. Examples include misreporting of credentials; improper use of “projective testing” or other, unvalidated tests; misinterpretation of test results; erroneous (implicit) application of Bayes Theorem in the form of “profile testimony” and misapplying symptom, test, and diagnostic probability data; failure to disclose limitations on the validity of assessment techniques; failure to disclose data discrepant with the expert’s favored conclusions; misrepresentation of statements by professional associations; and misrepresentation of research results. It is up to experts to uphold the highest standards of their respective professions, disclose fully and fairly the bases for their opinions, rely to the greatest extent possible on solid scientific findings, understand and explain the uncertainties in their inferences, and be frank about the degree to which their theories and methods pass, or fail to pass, Daubert/Kumho scrutiny. The integrity of the legal system demands the most careful scrutiny of experts whose opinions may impact the lives and rights of American citizens.

Knowledgeable experts can also contribute in another way. Helping educate legal and mental health professionals about scientific methods, and what is good science vs. junkscience, will help guard our legal system from many forms of bogus expert testimony.

Attorneys may also need to modify traditional practices, where Daubert/Kumho hearings are involved. The common practice of “getting up to speed” by a hasty reading of general tomes is likely to fail miserably in highly complex, science-intensive hearings dealing with multivariate analysis, sophisticated research methodologies, philosophy of science, and detailed facts from decades of research findings. In the world of Daubert/Kumho analysis, a science/law team should be the minimal standard of legal practice, to help ensure that these complexities are properly addressed. We believe attorneys have an affirmative duty to consult with or defer to expert attorneys or scientists in the relevant fields. Improper loss of a Daubert/Kumho hearing may yield dire consequences for clients (e.g. false imprisonment of an innocent client, or an innocent child’s continued exposure to an abusive environment) and could even lead to a new area of legal malpractice claims. The demand for specialized education and knowledge created by Daubert, Kumho and related decisions is likely to hasten the end of the “Lone Ranger” approach to
complex cases and hasten the advent of the multidisciplinary team approach to science intensive litigation.

References

Case Law


Blackowiak v. Kemp, 546 N.W. 2d 1, Minn., April 19, 1996.


Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).


In the Matter of Renee Fredrickson LP 2653, Order of the Minnesota Board of Psychology, May 7, 1999


People v. Shirley, 31 Cal. 3d 18, 1982.


Travis v. Ziter, 681 So. 2d 1348, Ala., July 12, 1996.


Publications


Warmbir, Steve Doctor will testify against her colleague, Chicago Daily Herald, July 10, 1999)


Author Note
William M. Grove, Ph.D., Department of Psychology, University of Minnesota, N218 Elliott Hall, 75 East River Road, Minneapolis, Minnesota 55455–0344. R. Christopher Barden, Ph.D., J.D., National Association for Consumer Protection in Mental Health Practices, 1093 Duffer Lane, North Salt Lake, Utah 84054.
The authors thank many colleagues for reviewing a draft of this article. Correspondence concerning this article should be addressed to William M. Grove at the address above, or at William.M.Grove-1@tc.umn.edu.

Reference Notes