

August, 1998

# 1. Where researchers fear to tread: Interpretive differences among testifying experts in child sexual abuse cases.

Thomas D. Lyon, *University of Southern California*  
Jonathan J. Koehler

# Expert Witnesses in Child Abuse Cases

What Can and Should Be Said in Court

Edited by  
Stephen J. Ceci and Helene Hembrooke

AMERICAN PSYCHOLOGICAL ASSOCIATION

# 12

## WHERE RESEARCHERS FEAR TO TREAD: INTERPRETIVE DIFFERENCES AMONG TESTIFYING EXPERTS IN CHILD SEXUAL ABUSE CASES

THOMAS D. LYON AND JONATHAN J. KOEHLER

Debates regarding the admissibility of expert testimony in child sexual abuse cases are often characterized as between clinicians and researchers. Clinicians base their judgment on personal experience and anecdotes, whereas researchers base their judgment on scientific findings. Clinicians are willing to testify that a particular child has been sexually abused, whereas researchers cautiously avoid rendering a judgment about any particular case. Clinicians believe that they can interpret children's statements and behaviors to validate abuse, whereas researchers warn that children's statements and behaviors may be shaped by adults, including clinicians. Clinicians are happy to testify (typically for the prosecution), comfortably adopting the role of advocate for their position. Researchers reluctantly agree to testify (typically for the defense) and are unhappy having done so, finding their neutrality challenged by the inherently adversarial nature of the trial. In short, clinicians rush in where researchers fear to tread.

One might think that debates would cease if clinicians were banished

from the courtroom. However, disagreements among researchers are suppressed when they face the common enemy of intuitive, unscientific judgment. Once that enemy is removed, differences within the ranks are highlighted, and the debate is repeated, albeit on a higher plane.

In this chapter, we address two areas of disagreement among researchers, involving the two most widely discussed areas of expert testimony in child sexual abuse cases: behavioral symptoms and the suggestibility of children. In each case, subjective differences among researchers in their interpretation of the research lead to differences of opinion regarding what an expert may appropriately testify to in court. There are several questions about which reasonable researchers disagree. How does one weigh false-positive results against false-negative results? How should methodological limitations be weighed? How closely must the characteristics of a study match those of a given situation for the results to apply? How many studies must support a proposition before one should accept the proposition as true? The questions are not new, but their persistence bears repeating in the light of strong claims regarding the ethics and admissibility of various types of expert testimony.

In the case of behavioral symptoms, observational research documenting symptoms in some nonabused children leads different researchers to different conclusions. Some believe that this renders symptoms irrelevant as evidence of abuse; others believe that symptoms may serve as evidence (but not conclusive evidence) that abuse occurred. We argue that a few symptoms are relevant evidence of abuse and the proper subject of expert testimony but that experts should acknowledge the methodological shortcomings of the observational research. In the case of suggestibility, laboratory research demonstrating that some children can be led to make false claims also leads researchers to disagree. Some believe this justifies concerns that a nontrivial number of sexual abuse allegations are the product of coercive interviewing techniques, whereas others question the applicability of the research to most sexual abuse investigations. We believe that laboratory research is often relevant but that experts must acknowledge the potential differences between the research and the nature of interviewing and the dynamics of sexual abuse.

## BEHAVIORAL SYMPTOMS OF ABUSE

Several contributors to this volume discuss the admissibility of expert testimony that behavioral symptoms are diagnostic of sexual abuse (Berliner, chapter 1; Fisher & Whiting, chapter 8; Kovera & Borgida, chapter 9; Lawlor, chapter 5; Mason, chapter 10). The consensus is that such testimony should not be admitted, although the chapters by Berliner and Kovera and Borgida are more tolerant of such testimony than are the others.

In our view, experts can reasonably testify that some symptoms are relevant for proving that abuse occurred. Disagreements are inevitable, however, because the research fails to provide conclusive answers.

In assessing the arguments, it is important to emphasize a distinction noted by Kovera and Borgida (this volume, chapter 9) between symptoms that increase the likelihood that abuse occurred and symptoms that are dispositive of abuse. If a symptom occurs more frequently in abused children than in nonabused children, then the presence of that symptom in a child increases the likelihood that the child has been abused (Lyon & Koehler, 1996). In legal terms, the symptom is relevant or probative evidence that the child was abused (*Federal Rules of Evidence* Rule 403). However, a symptom may increase the likelihood of abuse without justifying the ultimate judgment that the child was abused. One can conclude that a symptom is dispositive of abuse only if the symptom occurs in some abused children and is nonexistent in nonabused children. If the symptom appears in some abused children and some smaller number of nonabused children, the symptom is relevant but not definitive evidence of abuse. Whether one can conclude that abuse occurred depends on the strength of the other evidence in the case and on one's standard regarding how sure one must be to diagnose abuse.

If there are symptoms that are relevant but not conclusive evidence of abuse, then such symptoms could properly be the subject of expert testimony that a child exhibited symptoms probative of abuse. The expert could inform the jury that particular symptoms increase the likelihood that abuse occurred but should refrain from expressing an opinion whether the child in question had been abused. An examination of the expert might inquire into the extent to which the symptoms appear in abused and nonabused children and whether the causes of symptomatology in nonabused children can be discerned. As Mason (this volume, chapter 10) points out, such testimony would be consistent with legal proscriptions against testimony in which an expert testifies as to the ultimate judgment whether abuse occurred.

Reviews of the research on the sequelae of sexual abuse tend to find that many symptoms are in fact more common in abused children than in nonabused children (Kendall-Tackett, Williams, & Finkelhor, 1993). Because of their existence in nonabused children, however, reviewers have concluded that such symptoms are insufficient standing alone to prove conclusively that abuse occurred. Kendall-Tackett et al. (1993) concluded that "symptoms cannot be easily used, without other evidence, to confirm the presence of sexual abuse" (p. 175). Berliner and Conte (1993) found that "the qualities or characteristics of children will not be, in themselves, determinative of sexual abuse" (p. 116).

In this volume, Fisher and Whiting (chapter 8) refer to Kendall-Tackett et al.'s (1993) review for the proposition that "practitioners cannot

rely on the presence or absence of symptom patterns to validate sexual abuse" (italics added, Fisher & Whiting, this volume, p. 168). If one reads "rely" as referring to sole reliance, then the statement captures the position taken by Kendall-Tackett et al. However, Fisher and Whiting's position is different, as revealed in their next paragraph, which concludes that "there is no empirical evidence that such [sexual and other] behaviors can serve as evidence that a child has been sexually abused" (italics added, p. 168).

Why is there this difference of opinion? In part, we believe it arises from disagreements among researchers regarding the emphasis that one should place on the probability of false-positive errors—in this case, false allegations of abuse. Symptoms are relevant but not conclusive when they occur in some nonabused children. Their presence in nonabused children raises the potential for false allegations because nonabused children with symptoms may be misdiagnosed as abused. If one has no tolerance for false allegations, then one would require symptoms to be conclusive to be admissible. Fisher and Whiting might have argued that many symptoms do not merely occur in some nonabused children but in fact are quite common, thus increasing the risk of false positives. Yet, the point at which the risk of false positives is so high that testimony regarding symptoms should be inadmissible is a subjective judgment, about which ethical experts may disagree.

A second reason researchers might disagree over the relevance of symptoms of abuse stems from a misunderstanding about the significance of the fact that most abused children do not exhibit any particular symptom. It is neither necessary nor sufficient for a symptom to appear in a majority of sexually abused children to increase the likelihood that abuse occurred. The symptom may appear in a minority of abused children but also be relevant because it appears in a smaller minority of nonabused children. The symptom may appear in a majority of abused children but be irrelevant because it is as common in nonabused children. It is therefore of limited significance that "no one symptom characterize[s] a majority of sexually abused children" (Fisher & Whiting, this volume, chapter 8, p. 168); that "children who have been sexually abused can exhibit widely varying emotional reactions and behaviors, including showing no overt emotional disturbance" (Lawlor, this volume, chapter 5, p. 110); or that "the many studies that have been performed on the indicators of sexual abuse fail to reveal any consistent pattern of behavior that can be identified in the majority of sexually abused children" (Mason, this volume, chapter 10, p. 232).

Consider physical evidence of sexual abuse. No consistent pattern of physical signs can be identified in a majority of sexually abused children because approximately half of sexually abused children appear normal on examination (Bays & Chadwick, 1993). Erythema is the most common physical sign, but it is irrelevant because it appears in the same proportion

of nonabused children as abused children (Emans, Woods, Flagg, & Freeman, 1987). Gonorrhea is a rare sign, appearing in fewer than 5% of abused children, but it is highly suggestive of abuse because it is virtually nonexistent in nonabused children (Bays & Chadwick, 1993). In short, the frequency with which a sign appears is a poor proxy for its relevance. Some signs that occur frequently are irrelevant; some signs that occur infrequently are quite relevant.

Lawlor (this volume, chapter 5) notes that whereas 10% of sexually abused children depict genitalia in their drawings only 2% of nonabused children do so. He concludes that because "the numbers are extremely low in each case[,] . . . there is not much predictive value from genitalia in drawings of young children" (p. 116). Yet if the numbers are reliable—that is, if research were to consistently find a 5:1 ratio of abused to nonabused children whose drawings show genitalia—then drawings of genitalia would be probative evidence that abuse occurred.

The fact that many abused children do not exhibit symptoms of abuse is important information but not because it demonstrates the irrelevance of such symptoms. Rather, it raises the potential that clinicians might erroneously conclude that a child has not been abused because that child failed to exhibit the appropriate symptoms. Fisher and Whiting (this volume, chapter 8) cite Finkelhor (1987) to support their position that symptoms of posttraumatic stress disorder (PTSD) should not be interpreted as evidence of sexual abuse. Finkelhor (1987) was indeed critical of the PTSD model:

First, [the PTSD model] does not adequately account for all the symptoms. Second, it accurately applies only to some of the victims. Finally, and most seriously, it does not truly present a theory that explains how the dynamics of sexual abuse lead to the symptoms noted. (p. 350)

A moment's reflection shows that although these objections undermine PTSD's utility as an explanatory model for understanding the effects of sexual abuse, they are consistent with the possibility that PTSD symptoms are much more common in abused children than nonabused children. Indeed, Finkelhor (1987) never argued that PTSD is not evidence of abuse. Rather, he warned that because many abused children do not exhibit PTSD, it would be an error for experts to "testify in court cases that alleged victims probably were not abused because they do not manifest PTSD" (p. 352).

In summary, symptoms are relevant for proving that abuse occurred if they are more common in abused children than nonabused children. Having said that, we hasten to add that methodological concerns lead us to adopt a conservative criterion for determining when a symptom is indeed more common in abused children. We believe that unless symptoms are more common in abused children than nonabused children in treatment,

prudent experts ought to avoid labeling such symptoms as relevant for proving that abuse occurred. Because most of the research selects sexually abused children from sexual abuse evaluations or treatment programs (Kendall-Tackett et al., 1993), the extent to which abused children are symptomatic may be exaggerated (Friedrich, Urquiza, & Beilke, 1986). Unless nonabused children are drawn from clinical populations, differences between abused and nonabused children may be attributable to selection bias rather than to real differences in symptom prevalence.

Our position reflects the possibility that selection biases artificially inflate the apparent relevance of abuse symptoms, but the extent to which they do is unknown. Moreover, we recognize that requiring that symptoms occur in a higher percentage of abused children than clinical nonabused children may lead one to understate true differences in symptoms between abused and nonabused children, in part because asymptomatic abused children are more likely than asymptomatic nonabused children to be referred to treatment simply "because of something done to them (i.e., abuse)" (Kendall-Tackett et al., 1993, p. 165). We prefer to risk understating rather than overstating the relevance of symptoms and, therefore, prefer to rely on comparisons between abused children and nonabused children in treatment.

If our criterion is accepted, then which symptoms are relevant? According to Kendall-Tackett et al. (1993), "sexually abused children showed only two symptoms consistently more often than non-abused clinical children: PTSD (just one study) and sexualized behavior (six of eight studies)" (p. 165). Given Fisher and Whiting's (this volume, chapter 8) extensive critique of PTSD and sexualized behavior, this conclusion may be surprising. The differences of opinion are attributable to disagreements about how the research ought to be interpreted.

Fisher and Whiting (this volume, chapter 8) note that "only" 48–55% of abused children meet the criteria for PTSD. But this figure, standing alone, says little about PTSD's potential as a relevant indicator of abuse. Deblinger, McLeer, Atkins, Ralphe, and Foa (1989) appear to be the only researchers who have examined differences between abused and nonabused children. Fisher and Whiting report that Deblinger et al. (1989) found "no significant group differences . . . when a PTSD symptom checklist was used to compare sexually abused psychiatric patients with physically abused and nonabused patients" (Fisher & Whiting, this volume, chapter 8, p. 166). Strikingly, Deblinger et al.'s is the one study Kendall-Tackett et al. (1993) believe supports the proposition that PTSD is more common in abused children than nonabused clinical children.

Who is right? According to Deblinger et al. (1989), the differences among the percentages of sexually abused children (20.9%), physically abused children (6.9%), and nonabused children (10.3%) who met the diagnostic criteria for PTSD were not statistically significant. However,

they also reported that "significant differences were found across groups on specific PTSD categories" (p. 405), namely reexperiencing phenomena and autonomic hyperarousal. Hence, depending on the level of analysis used to examine the Deblinger et al. data, one may conclude that the researchers either did or did not find significant differences among the groups.

One might conclude that because differences in PTSD symptoms between abused and nonabused children have been examined in only one study, it is premature to conclude that PTSD is relevant for proving abuse. Still, one finding reported by Deblinger et al. (1989) is consistent with a substantial amount of other research: Sexually abused children exhibited more sexually inappropriate behaviors than either physically abused or nonabused children (Gale, Thompson, Moran, & Sack, 1988; Goldston, Turnquist, & Knutson, 1989; Hibbard & Hartman, 1992; Kolko, Moser, & Weldy, 1988; White, Halpin, Strom, & Santilli, 1988). Fisher and Whiting (this volume, chapter 8) acknowledge that the rates of sexualized behavior are higher among abused children. In addition, they note that in several well-controlled studies, "Friedrich and his colleagues have demonstrated that sexually abused children exhibit more sexual behavior problems than either nonabused clinical samples or nonreferred children" (p. 169; also see Einbender & Friedrich, 1989; Friedrich, 1995; Friedrich et al., 1986, 1992). Friedrich (1993) himself has reviewed the literature and concluded that "a growing body of increasingly sophisticated empirical research has demonstrated that sexual abuse is related to increased sexual behavior following the abuse" (p. 64).

Why, then, do Fisher and Whiting (this volume, chapter 8) conclude that there is "no empirical evidence that such behaviors can serve as evidence that a child has been sexually abused" (p. 168)? They note that some nonabused children exhibit sexualized behavior but that no more than 42% of abused children do so. As we have discussed, such data indicate that sexualized behavior is not conclusive evidence of abuse but do not undercut its utility as relevant evidence. Indeed, Friedrich et al.'s (1992) position, as cited by Fisher and Whiting (this volume, chapter 8), is that sexualized behavior ought "not be relied on in isolation as the primary indicator of sexual abuse" (p. 169). Kendall-Tackett et al. (1993) similarly concluded that sexualized behavior may "indicate sexual abuse but is not completely diagnostic because children can apparently appear to be sexualized for other reasons" (p. 173).

Fisher and Whiting (this volume, chapter 8) also argue that sexualized behavior may discriminate less well for girls than boys, that patterns of specific behaviors vary across different studies, and that variables other than sexual abuse have been linked to sexual behavior. Whether these are damning criticisms is a question about which ethical experts might disagree.

What appears to be the most serious criticism—the existence of al-

ternative causes for sexualized behavior—is troubling only so long as a testifying expert ignores those causes. The very process by which abuse is diagnosed involves consideration of alternative explanations for a child's statements and behavior. Consider physical abuse. Subdural hematomas are much more common in abused children than in nonabused children. However, subdural hematomas are also associated with nonabusive causes, such as blood disorders, meningitis, blood vessel abnormalities, neurodegenerative disease, birth trauma, and accidental trauma (Lyon, Gilles, & Cory, 1996). An awareness of alternative explanations, far from rendering subdural hematomas irrelevant for proving that abuse occurred, allows for a differential diagnosis that increases the physician's confidence that a particular child's injury is attributable to abuse. Therefore, if a previously healthy infant without a history of major trauma presents with a subdural hematoma, physicians consider it highly suggestive of physical abuse (Case, 1994). Similarly, if alternative explanations for a child's sexualized behavior can be considered and eliminated (e.g., Is the child regularly exposed to nudity in the home?), then the probative value of such behavior increases.

Both Lawlor (this volume, chapter 5) and Fisher and Whiting (this volume, chapter 8) assert that behaviors attributed to sexual abuse may actually be the effects of abuse investigation and therapy, what one might call *iatrogenic effects*. Certainly, the investigation and prosecution of child sexual abuse are stressful for any child and will, in many cases, lead to stress reactions (Goodman et al., 1992). With respect to sexualized behavior, however, there is little evidence for iatrogenesis. Friedrich and Reams (1987) observed that sexual problems emerged during therapy for some of their patients, but Hewitt and Friedrich (1991) and Friedrich, Luecke, Beilke, and Place (1992, as cited in Friedrich, 1995) found decreases in sexual behavior over time. Gomes-Schwartz, Horowitz, and Cardarelli (1990) found little change in inappropriate sex play (a slight decrease) or sexual preoccupation (a slight increase) in their 18-month follow-up of sexually abused children 12 years old and younger. Even if sexual behaviors were to increase over time in therapy, it would be unclear whether therapy played a causal role, just as it would be difficult to credit therapy with reducing such behaviors over time.

The use of anatomically correct dolls has evoked some concern that they might elicit sexualized play. The evidence on this point is mixed. Bruck, Ceci, Francouer, and Renick (1995) interviewed parents of 3-year-olds who had been interviewed with anatomically correct dolls about a pediatric examination. The authors concluded that "parents did not feel that the dolls or the session with the dolls provoked any unusual sexualized behavior in their children" (p. 103). However, Boat, Everson, and Holland (1990) found that following exposure to anatomically correct dolls, many 3- to 4-year-olds were more interested in sexuality and their behaviors became more sexually focused.

We are less skeptical than Fisher and Whiting (this volume, chapter 8) regarding the potential relevance of sexualized behavior in diagnosing sexual abuse, but we do have several concerns. First, we believe that for diagnosis, the multi-item Child Sexual Behavior Inventory may be less useful than a focus on rare but highly probative behaviors. Although the inventory produces a true-positive rate of 70–92%, it produces a false-positive rate of 45–65% (Friedrich et al., 1992). Consequently, a high score on the inventory is only moderately probative of abuse. In contrast, some types of sexualized behavior are infrequent in abused children but so much more infrequent in nonabused children that they are highly probative of sexual abuse. For example, masturbation with an object occurs in more than 10% of abused children but less than 1% of nonabused children (Friedrich et al., 1992).

Several methodological concerns with the research on sexualized behavior in abused children must also be addressed. Sackett (1979) has identified a number of potential biases in case-control research—two of which are particularly important to consider in reviewing research on the behavioral consequences of sexual abuse. First, *exposure suspicion bias* occurs when the awareness of an individual's symptoms (sexualized behavior) increases the intensity of the search for and recognition of a putative cause (sexual abuse). Children who behave sexually are more likely to be questioned by their parents and professionals regarding possible sexual abuse (Friedrich & Reams, 1987; Sorensen & Snow, 1991). If sexualized behavior increases the likelihood that abuse will be discovered, children known to have been abused would exhibit more sexualized behaviors than children not believed to have been abused, even if abused children in general were no more likely than nonabused children to behave sexually. Exposure suspicion bias is minimized to the extent that samples of sexually abused children were originally detected and ultimately diagnosed as abused without reference to suspicious behaviors.

Second, *diagnostic suspicion bias* occurs when the awareness of an individual's exposure to a putative cause (sexual abuse) increases the intensity of the search for and recognition of symptoms (sexualized behavior). Sexual abuse, once discovered, may lead parents and professionals to watch more carefully for sexualized behaviors. Behaviors that might be interpreted as innocent play or natural curiosity in a child not suspected of being abused may be reinterpreted as sexual (Friedrich et al., 1992). Moreover, parents may be more likely to recall sexualized behavior once they believe their child has been abused. Diagnostic suspicion bias is minimized to the extent that sexualized behaviors can be objectively observed and recorded. Both exposure suspicion bias and diagnostic suspicion bias may exaggerate the probative value of some sexualized symptoms, but it would be difficult to measure these effects.

In summary, we believe that expert witnesses can point to research

supporting the claim that sexualized behavior and some components of PTSD are probative of sexual abuse. Symptoms need not appear in a majority of sexually abused children or be nonexistent in nonabused children to be relevant. We regard this as a matter of logic, not opinion. Other issues are more subjective. How much more common must symptoms be in abused children than nonabused children before the symptom has practical significance as a symptom of child abuse? How many studies must indicate that a symptom is relevant before one should conclude that it is? How should one weigh methodological objections—should one presume that their existence explains the observed result or should one require evidence that they are operating? Disagreements over these issues do not separate the ethical from the unethical. Instead, they reflect justifiable differences of opinion regarding how certain one must be before giving testimony that a symptom is relevant for proving that abuse occurred.

## THE SUGGESTIBILITY OF CHILDREN

Experts in sexual abuse cases frequently spend a lot of time discussing the suggestibility of children. However, we were surprised to find that suggestibility received little attention in most of this volume's chapters. Perhaps this is because the topic fails to evoke skepticism. Testimony on suggestibility is usually based on laboratory research, much of which would seem to satisfy the U.S. Supreme Court's requirement that such testimony "must be derived by the scientific method" (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 1993, p. 590; see also Kovera & Borgida, this volume, chapter 9). Furthermore, most suggestibility experts avoid direct comments on the credibility of any particular child witness. For example, in describing her own experience as an expert witness, Bruck (this volume, chapter 4) emphasized that she avoided making statements "explicitly tied to the facts at bar" (p. 89). Thus, suggestibility testimony avoids the infirmities of clinical opinion, which is often based on unsystematic observation and is often case specific.

However, the strengths of such testimony are also its vulnerabilities. There is substantial disagreement among researchers about the applicability of laboratory research on suggestibility to children who actually appear in court (Ceci, 1991). The problem concerns ecological validity. Are the techniques used to interview children in the laboratory sufficiently similar to those used to interview an allegedly abused child in a particular case? Are the events about which children are interviewed in the laboratory comparable with a sexually abusive event? If not, how significant are the differences? In answering such questions, testifying experts are often forced to rely on their subjective judgments and personal experience.

Bruck (this volume, chapter 4) provides illustrative examples of cases in which suggestibility research was discussed, describing her expert testimony in two cases: the Martensville trials (*The Queen v. Ronald Sterling, Linda Sterling, and Travis Sterling*, 1994) and the Little Rascals case (*State v. Robert Fulton Kelly, Jr.*, 1992). In both cases, issues of ecological validity were raised. In *The Queen v. Linda Sterling*, the prosecutor questioned how well the research matched the interviewing actually conducted in the case at bar. In *State v. Kelly*, the prosecutor challenged assumptions that the research applied to allegations of sexual abuse that the child witnesses appeared to find embarrassing, aversive, and painful.

It is often impossible for an expert to know how a child has been interviewed in a particular case, even when the available reports relate what appear to be convincing narratives of abuse: As Bruck notes (this volume, chapter 4),

I learned that just because a professional concluded in a report that the child was abused, it did not necessarily mean that the child had reported abuse in the manner contained in the report or that the child had spontaneously reported the abuse. I learned that these reports could reflect situations in which the child was repeatedly questioned and finally assented to abuse in response to an interviewer's leading questions. (p. 88)

Verbatim records of interviews are frequently unavailable, particularly those documenting the initial allegation, which is typically made before a formal investigation begins.

Without a verbatim record, a testifying expert can only speculate about whether suggestive methods were used. The expert might assume that interviews were suggestive on the basis of a review of transcripts from other cases, but doing so risks a generalization from a potentially unrepresentative sample (Ceci & Bruck, 1995). Systematic research on suggestive practices in actual interviews is in its infancy (Lamb et al., 1996; McGough & Warren, 1994). Consequently, we know little about the ecological validity of findings from the suggestibility literature. These include conclusions about the impact of such coercive techniques as stereotype induction (Lepore & SESCO, 1994; Leichtman & Ceci, 1995) and questioning in which the child is told (rather than asked) whether the events occurred (Ceci, Loftus, Leichtman, & Bruck, 1994; Leichtman & Ceci, 1995).

For ethical reasons, research on children's suggestibility cannot directly test the proposition that children can be manipulated to recall abusive experiences that never occurred. To the extent that abusive experiences are different than experiences that can be tested, the applicability of suggestibility research to sexual abuse allegations is open to question (Ceci, Bruck, & Rosenthal, 1995; Lyon, 1995). Moreover, even assuming

agreement that differences exist, the practical significance of such differences is a matter of subjective opinion.

In legal terms, these issues address what the U.S. Supreme Court in *Daubert* (1993) called fit. *Fit* refers to the extent to which the expert's testimony is "tied to the facts of the case" (p. 591). A lack of fit renders an expert's testimony inadmissible under the *Federal Rules of Evidence* (Rule 702). A less than perfect fit raises issues of prejudicial impact (Rule 403) and can lead to the exclusion of expert testimony because juries are often deferential to expert opinion, even when the expert's testimony is of questionable relevance (Kovera & Borgida, this volume, chapter 9). Given the tendency of juries to defer decision making to experts (Miller & Allen, this volume, chapter 7), courts can use issues of fit and prejudice to weed out expert testimony when its persuasiveness exceeds its probative value.

One might argue that because scientific data on interview practices are limited, researchers should be barred from testifying about the suggestibility of children unless they have seen videotapes of the interviews or have at least reviewed transcripts of the interviews in a particular case. However, because documentation is frequently unavailable, this would dramatically limit the ability of experts to warn jurors about potentially suggestive interviewing techniques. Similarly, one might argue that the differences between the events examined by suggestibility researchers and actual sexual abuse justify exclusion of all expert testimony on suggestibility research. However, because it would be unethical for researchers to use coercive techniques to convince nonabused children that they were, in fact, abused, this argument would permanently foreclose all expert testimony on suggestibility. We believe that although expert testimony on suggestibility necessarily goes beyond what has been (or can be) scientifically proven, it should not be treated as inadmissible per se. If jurors are relatively uninformed about the suggestibility of children (or if they harbor misconceptions), expert testimony may be helpful.

At the same time, we agree with several of the contributors to this volume (Lavin & Sales, chapter 3; Pruett & Solnit, chapter 6) that experts are obligated to be candid about the limits of their knowledge: "When research is not available to address the question, they [experts] should present their testimony hemmed with disclaimers alerting the jury that it is not scientifically based and as yet unproven or unprovable" (Lavin & Sales, p. 79). Given Bruck's (this volume, chapter 4) warning that attorneys are often untrained in scientific methods and Kovera and Borgida's (this volume, chapter 9) discovery that cross-examination may fail to alter the jury's initial impressions of an expert's testimony, we would add that the expert should initiate a discussion of such limits in his or her direct testimony rather than wait for cross-examination.

## CONCLUSION

Interpretive differences among researchers in child sexual abuse cases reveal the limitations of scientific knowledge. Those who study the behavior of sexually abused children in the real world are unable to subject randomly selected participants to abuse and observe the effects. Consequently, their conclusions may be influenced by illusory correlations born of selection and diagnostic biases. Those who study the suggestibility of children in a laboratory setting are typically unable to recreate the multitude of factors that influence a child's report in the field. Both types of researchers make inferential leaps and judgment calls in court when they attempt to translate their research into case-relevant knowledge. Disagreements are inevitable, not because some researchers are ethical and others are not, but because all researchers are human.

## REFERENCES

- Bays, J., & Chadwick, D. (1993). Medical diagnosis of the sexually abused child. *Child Abuse & Neglect*, 17, 91–110.
- Berliner, L., & Conte, J. R. (1993). Sexual abuse evaluations: Conceptual and empirical obstacles. *Child Abuse & Neglect*, 17, 111–125.
- Boat, B., Everson, M., & Holland, J. (1990). Maternal perceptions of non-abused young children's behaviors after the children's exposure to anatomical dolls. *Child Welfare*, 64, 389–399.
- Bruck, M., Ceci, S. J., Francouer, E., & Renick, A. (1995). Anatomically detailed dolls do not facilitate preschoolers' reports of a pediatric examination involving genital touching. *Journal of Experimental Psychology: Applied*, 1, 95–109.
- Case, M. E. S. (1994). Head injury in child abuse. In J. A. Monteleone & A. E. Brodeur (Eds.), *Child maltreatment: A clinical guide and reference* (pp. 75–87). St. Louis, MO: G. W. Medical.
- Ceci, S. J. (1991). Some overarching issues in the children's suggestibility debate. In J. Doris (Ed.), *The suggestibility of children's recollections: Implications for eyewitness testimony* (pp. 1–9). Washington, DC: American Psychological Association.
- Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony*. Washington, DC: American Psychological Association.
- Ceci, S. J., Bruck, M., & Rosenthal, R. (1995). Children's allegations of sexual abuse: Forensic and scientific issues: A reply to commentators. *Psychology, Public Policy, and Law*, 1, 494–520.
- Ceci, S. J., Loftus, E. F., Leichtman, M. D., & Bruck, M. (1994). The possible role of source misattribution in the creation of false beliefs among preschoolers. *International Journal of Clinical and Experimental Hypnosis*, 42, 304–320.

- Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S., 113 S. Ct. 2786 (1993).
- Deblinger, E., McLeer, S. V., Atkins, M. S., Ralphe, D., & Foa, E. (1989). Post-traumatic stress in sexually abused, physically abused, and non-abused children. *Child Abuse & Neglect*, 13, 403–408.
- Einbender, A. J., & Friedrich, W. N. (1989). Psychological functioning and behavior of sexually abused girls. *Journal of Consulting and Clinical Psychology*, 57, 155–157.
- Emans, S. J., Woods, E. R., Flagg, N. T., & Freeman, A. (1987). Genital findings in sexually abused, symptomatic and asymptomatic girls. *Pediatrics*, 79, 778–785.
- Federal rules of evidence*, 28 U.S.C. (West 1994).
- Finkelhor, D. (1987). The trauma of child sexual abuse: Two models. *Journal of Interpersonal Violence*, 2, 348–366.
- Friedrich, W. N. (1993). Sexual victimization and sexual behavior in children: A review of recent literature. *Child Abuse & Neglect*, 17, 59–66.
- Friedrich, W. N. (1995, Spring). The clinical use of the Child Sexual Behavior Inventory: Commonly asked questions. *American Professional Society on the Abuse of Children Advisor*, 1, 17–20.
- Friedrich, W. N., Grambsch, P., Damon, L., Hewitt, S., Koverola, C., Lang, R., Wolfe, V., & Broughton, D. (1992). The Child Sexual Behavior Inventory: Normative and clinical findings. *Psychological Assessment*, 4, 303–311.
- Friedrich, W. N., & Reams, R. A. (1987). Course of psychological symptoms in sexually abused young children. *Psychotherapy*, 24, 160–170.
- Friedrich, W. N., Urquiza, A. J., & Beilke, R. L. (1986). Behavior problems in sexually abused young children. *Journal of Pediatric Psychology*, 11, 47–57.
- Gale, J., Thompson, R. J., Moran, T., & Sack, W. H. (1988). Sexual abuse in young children: Its clinical presentation and characteristic patterns. *Child Abuse & Neglect*, 12, 163–170.
- Goldston, D. B., Turnquist, D. C., & Knutson, J. F. (1989). Presenting problems of sexually abused girls receiving psychiatric services. *Journal of Abnormal Psychology*, 98, 314–317.
- Gomes-Schwartz, B., Horowitz, J. M., & Cardarelli, A. P. (1990). *Child sexual abuse: The initial effects*. Newbury Park, CA: Sage.
- Goodman, G. S., Taub, E. P., Jones, D. P. H., England, P., Port, L. K., Rudy, L., & Prado, L. (1992). Testifying in criminal court. *Monographs of the Society for Research in Child Development*, 57(Serial No. 229).
- Hewitt, S. K., & Friedrich, W. N. (1991). Effects of probable sexual abuse on preschool children. In M. Q. Patton (Ed.), *Family sexual abuse* (pp. 57–74). Newbury Park, CA: Sage.
- Hibbard, R. A., & Hartman, G. L. (1992). Behavioral problems in alleged sexual abuse victims. *Child Abuse & Neglect*, 16, 755–762.
- Kendall-Tackett, K. A., Williams, L. M., & Finkelhor, D. (1993). Impact of sexual abuse on children: A review and synthesis of recent empirical studies. *Psychological Bulletin*, 113, 164–180.

- Kolko, D. J., Moser, J. T., & Weldy, S. R. (1988). Behavioral/emotional indicators of sexual abuse in child psychiatric inpatients: A controlled comparison with physical abuse. *Child Abuse & Neglect*, 12, 529–541.
- Lamb, M. E., Hershkowitz, I., Sternberg, K. J., Esplin, P. W., Hovav, M., Manor, T., & Yudilevitch, L. (1996). Effects of investigative utterance types on Israeli children's responses. *International Journal of Behavioral Development*, 19, 627–637.
- Leichtman, M. D., & Ceci, S. J. (1995). The effects of stereotypes and suggestions on preschoolers' reports. *Developmental Psychology*, 31, 568–578.
- Lepore, S. J., & SESCO, B. (1994). Distorting children's reports and interpretations of events through suggestion. *Journal of Applied Psychology*, 79, 108–120.
- Lyon, T. D. (1995). False allegations and false denial in child sexual abuse. *Psychology, Public Policy, and Law*, 1, 429–437.
- Lyon, T. D., Gilles, E. E., & Cory, L. (1996). Medical evidence of physical abuse in infants and young children. *Pacific Law Journal*, 28, 93–167.
- Lyon, T. D., & Koehler, J. J. (1996). The relevance ratio: Evaluating the probative value of experts' testimony in child sexual abuse cases. *Cornell Law Review*, 82, 43–78.
- McGough, L. S., & Warren, A. R. (1994). The all-important investigative interview. *Juvenile & Family Court Journal*, 45, 13–29.
- The Queen v. Ronald Sterling, Linda Sterling, and Travis Sterling, Jud. Centre of Saskatoon, Sask., Q.B.J. No. 74 (1994).
- Sackett, D. L. (1979). Bias in analytic research. *Journal of Chronic Disease*, 32, 51–63.
- Sorensen, T., & Snow, B. (1991). How children tell: The process of disclosure in child sexual abuse. *Child Welfare*, 70, 3–15.
- State v. Robert Fulton Kelly, Jr., Superior Criminal Court, Pitt County, NC, #91-CRS-4250–4363 (1991–1992).
- White, S., Halpin, B. M., Strom, G. A., & Santilli, G. (1988). Behavioral comparisons of young sexually abused, neglected, and nonreferred children. *Journal of Clinical Child Psychology*, 17, 53–61.