THE SUGGESTIBILITY OF CHILDREN: SCIENTIFIC RESEARCH AND LEGAL IMPLICATIONS

Stephen J. Ceci[1] & Richard D. Friedman[2]

1 Helen L. Carr Professor of Developmental Psychology, Department of Human Development, Cornell University.
2 Ralph W. Aigler Professor of Law, University of Michigan Law School.

In this Article, Professors Ceci and Friedman analyze psychological studies on children's suggestibility and find a broad consensus that young children are suggestible to a significant degree. Studies confirm that interviewers commonly use suggestive interviewing techniques that exacerbate this suggestibility, creating a significant risk in some forensic contexts—notably but not exclusively those of suspected child abuse—that children will make false assertions of fact. Professors Ceci and Friedman address the implications of this difficulty for the legal system and respond to Professor Lyon's criticism of this view recently articulated in the Cornell Law Review. *Using Bayesian probability theory, Professors Ceci and Friedman assess the implications of children's suggestibility for fact-finding in adjudication. Based on the constitutionally compelled principle that an inaccurate criminal conviction is a far worse result than a failure to gain an accurate conviction, even a slight risk of false allegations is significant. Professors Ceci and Friedman present several policy implications that follow from their analysis. First, interviewers should use leading questions only as a last resort, and they should completely avoid some strongly suggestive techniques that create particularly significant risks of false allegation. Second, except in very limited circumstances the fact that a child has been subjected to suggestive questioning should not preclude her from testifying. Instead, in appropriate cases, courts should be receptive to expert evidence on the suggestibility of children. Furthermore, in some extreme cases in which the child's allegation is essential to the prosecution and the child was subjected to very strongly suggestive influences, a criminal conviction should be precluded. To the extent that reliability is a factor in determining the admissibility of hearsay statements, in some circumstances children's statements should be considered unreliable. Finally, absent exigent circumstances, all interviews conducted as part of a criminal abuse investigation should be videotaped, to reduce the uncertainty as to whether interviewers have used suggestive questioning techniques.*

Introduction

Young children have historically been viewed as particularly vulnerable to suggestion. Within the mainstream scientific community, scholars agree that young children are more susceptible than older individuals to leading questions and pressures to conform to the expectations and desires of others. At the same time, children may hesitate to disclose matters such as sexual abuse without significant prompting. In some circumstances, these frailties aggravate the already difficult task of determining whether a child's statement is truthful. This matter is of immense concern because of the large number of young children who are interviewed each year during the course of abuse and neglect investigations.[3] The vulnerabilities of young children have far-reaching

[3] In each of the past three years there have been approximately two million reports alleging the maltreatment of nearly three million children. Among the forty-nine states reporting to the National Center on Child Abuse and Neglect 1997 data system, there were 126,095 cases of *substantiated* sexual abuse of minors–or nearly 13% of all reported cases

implications for the juvenile and criminal justice systems. Arguably, these vulnerabilities may affect how an investigator should interview the child; whether she should be allowed to testify in court; whether her hearsay statements should be admitted; whether expert evidence concerning her vulnerability should be admitted; and whether a criminal conviction based principally on her testimony should be allowed.

Recently, however, a number of scholars—most notably John E.B. Myers and Thomas D. Lyon on this side of the Atlantic—have vigorously criticized this mainstream view.[4] These scholars have chastised scientific researchers for fueling what they deem to be a backlash against believing children's claims of abuse. They believe that for at least two reasons the results of the scientific research have little bearing on the real world. First, they argue that there is scant empirical evidence to support the assumption that child-abuse interviewers often employ highly suggestive interviewing techniques that are potentially damaging to the accuracy of children's statements.[5] Second, they argue that these techniques, even if commonly used in interviews, would not result in suggestibility errors of the

---

and 37% of all sexual abuse allegations. Although only a fraction of these children actually testified in criminal court, virtually all of them were interviewed by law enforcement officials and/or child protective service workers, and many gave depositions or unsworn testimony. The most recent national incidence data make clear that young children are at least as likely to be sexually abused as older children:

Age range	*Incidence (Cases per 1,000)*
3-5	5.1
6-8	3.2
9-11	3.7
12-14	2.6
15-17	2.7

ANDREA J. SEDLAK & DIANE D. BROADHURST, U.S. DEP'T OF HEALTH & HUMAN SERVS., THE THIRD NATIONAL INCIDENCE STUDY OF CHILD ABUSE AND NEGLECT: FINAL REPORT 4-13 fig.4-6 (1996). Across all reports of sexual abuse, children aged seven and younger comprise 40.65% of cases.

To this evidence, we can add the data from a recent government report: "When victims are grouped in 4-year age categories, those 4-7 years old were the highest proportion of victims (26.2 percent)." U.S. DEP'T OF HEALTH AND HUMAN SERVS., CHILD MALTREATMENT 1997: REPORTS FROM THE STATES TO THE NATIONAL CHILD ABUSE AND NEGLECT DATA SYSTEM 4-2 (1999). If one adds to the above data children embroiled in acrimonious custody disputes, PINS (Persons in Need of Supervision) actions, neglect proceedings, and termination of parental rights suits, then the absolute numbers of young children enmeshed in the justice system is very large.

4 *See* Thomas D. Lyon, *False Allegations and False Denials in Child Sexual Abuse*, 1 PSYCHOL. PUB. POL'Y & L. 429 (1995); Thomas D. Lyon, *The New Wave in Children's Suggestibility Research: A Critique*, 84 CORNELL L. REV. 1004 (1999) [hereinafter Lyon, *New Wave*]; John E. B. Myers, *New Era Of Skepticism Regarding Children's Credibility,* 1 PSYCHOL. PUB. POL'Y & L. 387 (1995) [hereinafter Myers, *New Era*]; John E.B. Myers et al., *Psychological Research on Children as Witnesses: Practical Implications for Forensic Interviews and Courtroom Testimony*, 28 PAC. L.J. 3 (1996) [hereinafter Myers et al., *Psychological Research*]; Helen Westcott, Book Review, 89 BRIT. J. PSYCHOL. 525 (1998).

5 *See* Myers, *New Era*, *supra* note 2, at 396. Myers explains that:

> The reader comes away from Ceci and Bruck's articles with the impression that many, if not most, interviews are conducted improperly . . . .
>
> If Ceci and Bruck supported their assertions with sufficient evidence, they would have to be taken seriously. However, they fail to support their indictment of investigative interviewing and children's credibility. . . . [T]heir articles convey an unnecessarily pessimistic picture of the child protection system and children's credibility.
>
> . . . In the final analysis, the articles [by Ceci and Bruck] fuel unwarranted skepticism of children and the system designed to protect them.

Id.

magnitude that scientific studies suggest. Those studies "neglect[] the characteristics of child sexual abuse that both make false allegations less likely and increase the need to guard against a failure to detect abuse when it has actually occurred."[6]

In this Article, we summarize and analyze the principal findings of psychological research concerning children's suggestibility as well as other factors that may affect the credibility of a child's allegation of abuse. We demonstrate that what Lyon characterizes as a "new wave" of research[7] is actually a broad and long-standing scientific mainstream. We argue that the results of this research do, indeed, raise significant concerns for the real world of abuse and abuse investigation and thus engender significant legal implications.

Part I of this Article briefly describes the history and current state of research into children's suggestibility. In this Part, we argue that, although psychological researchers disagree considerably over the degree to which the suggestibility of young children may lead to false allegations of sexual abuse, there is an overwhelming consensus that children are suggestible to a degree that, we believe, must be regarded as significant. In presenting this argument, we respond to the contentions of revisionist scholars, particularly those recently expressed by Professor Lyon in the *Cornell Law Review*.[8] We show that there is good reason to believe the use of highly suggestive questions remains very common, and that these questions present a significant possibility that children will make false allegations even on matters such as sexual abuse.

Part II develops a framework, using Bayesian probability theory, for considering the findings described in Part I. We argue that there is merit to the traditional—and constitutionally compelled—view that an inaccurate criminal conviction is a far worse result than a failure to reach an accurate conviction, and that this perspective should inform the design of legal systems. With this in mind, we explain that even relatively slight probabilities of false allegations are potentially significant. Moreover, we show that the very substantial probability that a child who has been abused will fail to reveal the abuse tends, perhaps counter-intuitively, to diminish the probative value of an allegation of abuse when it is actually made.

Part III discusses the legal implications of our analysis. We conclude, in line with a rather broad consensus, that leading questions should be used in interviewing children only as a last resort, and stronger suggestive techniques should be avoided altogether. However, we express doubt about emerging legal doctrine that would preclude a child from testifying if she had been exposed to unduly suggestive questioning. We contend that, in some circumstances, the admissibility of children's hearsay statements cannot be justified on the basis that such statements are

---

6 Lyon, *New Wave*, *supra* note 2, at 1004.

7 Lyon describes the "new wave" as a:

> prestigious group of researchers in developmental psychology who argue that children are highly vulnerable to suggestive interviewing techniques. Because of its scientific credentials, its moderate tone, and its impressive body of research, the new wave presents a serious challenge to those who have claimed that children are unlikely to allege sexual abuse falsely.

Id.

8 *Id.*

particularly reliable. We believe that courts should be receptive in appropriate cases to expert evidence on the suggestibility of children, and that in some cases doubts about the credibility of the child should preclude a criminal conviction. We also show that the array of responses to the suggestibility problem available to the legal system substantially diminishes the significance of the issue of how common suggestive questioning is in abuse cases. The court can and should tailor its decisions to the case at hand without worrying about whether the case is typical or not. As an aid to that end, we recommend the videotaping of all interviews of children that are conducted as part of investigations of suspected abuse.

We agree with Lyon that "no science is value-free."[9] He himself acknowledges that "an acute awareness of true cases of abuse and the difficulty abused children have in revealing abuse" affects his analysis.[10] We share this awareness, and we believe we appreciate the grievous harm that can be wrought by child sexual abuse. Moreover, because some scholars have suggested that Ceci has been grudging in his recognition of children's testimonial strengths,[11] we emphasize our belief that when children allege that they have been subjected to sexual abuse they are frequently, even usually, accurate. As Ceci and Bruck have written, "Children have enormous strengths in recollecting their past. Even very young preschool children are capable of providing highly detailed and accurate accounts of prior interactions, provided that the adults who have access to them do not do anything to usurp their memories."[12] We believe, and Ceci has written, that there are probably many more unreported cases of child sexual abuse than false allegations of abuse, and this is a most serious problem.[13] We are confident that fair-minded readers of this Article will not conclude that we are biased against child witness es. Indeed, some of our recommendations have been met with a cold blast of fury from defense lawyers.

At the same time, we share with our legal system an abhorrence of convicting a person for a crime he did not commit, whether the crime is

---

9 *Id.* at 1084. Ceci has expressed the same thought elsewhere. Stephen J. Ceci et al., *Human Subjects Review, Personal Values, and the Regulation of Social Science Research*, 40 AM. PSYCHOL. 994, 1001 (1985).

10 Lyon, *New Wave*, *supra* note 2, at 1014.

11 Myers, *New Era*, *supra* note 2, at 396; Westcott, *supra* note 2, at 527 (speaking of Ceci and Bruck's supposed "almost exclusive focus on children's weaknesses" and relative inattention to "children's strength as witnesses"); *supra* note 3.

12 Stephen J. Ceci & Maggie Bruck, *Children's Testimony: Applied and Basic Issues*, *in* 4 HANDBOOK OF CHILD PSYCHOLOGY: CHILD PSYCHOLOGY IN PRACTICE 713 (William Damon ed., Irving E. Sigel & K. Ann Renninger vol. eds., 5th ed. 1998) (emphasis removed).

13 *See* STEPHEN J. CECI & MAGGIE BRUCK, JEOPARDY IN THE COURTROOM: A SCIENTIFIC ANALYSIS OF CHILDREN'S TESTIMONY 24 (1995) ("Although there is some debate over the exact prevalence of child sexual abuse in North America, the existing data lead us to believe that the incidence data most certainly underestimate the number of annual cases of sexual abuse."). Ceci and Bruck elaborate:

> We recoil at an earlier generation's denial of the magnitude of the problem that exists at all levels of society and at rates that are so high as to call into question some basic assumptions that we hold about ourselves as a people and about the inherently healthy status of the American family. . . . [A]lthough false claims exist, and perhaps in nontrivial numbers, we also believe that, in absolute numbers, the incidence of sexual abuse is very large and must never be minimized.

Id. at 4.

child sexual abuse or, say, assault, kidnaping, rape, or murder. False convictions are an ever-present possibility, partly as a result of the suggestive techniques that are sometimes used to investigate abuse. In earlier times, many adults were executed because of children's statements.[14] Some of the suggestive factors used then to coax children's disclosures remain in operation today. Society must be ever vigilant to prevent the occurrence of a modern counterpart of such tragedies.[15]

I

Suggestibility Research in Historical Perspective

We begin with a very brief history of research on children's suggestibility.[16] We will show that although there are obviously disagreements among researchers, there is also a substantial degree of consensus. To a large extent, the disagreements focus on which aspects of the data warrant emphasis. The so-called "new wave" is actually part of a very old and dominant view.

A. Early Studies

---

14 In Sweden, a great witch panic occurred between 1668 and 1676. *See* Rickard L. Sjöberg, *Child Testimonies During an Outbreak of Witch Hysteria: Sweden 1670-1671*, 36 J. CHILD PSYCHOL. & PSYCHIATRY 1040 (1995). Village priests interviewed approximately 600 children about the presumed sorcery of neighbors. *Id.* As a result of the children's statements, fourteen adults in one community were burned at the stake, and in another community twenty-seven adults were either beheaded or burned at the stake. *Id.*

In America, the most notorious witch trials occurred in Salem, Massachusetts during the final decade of the seventeenth century. CECI & BRUCK, *supra* note 11, at 8. A group of children known as the "circle girls" testified that they witnessed the following: the defendants flying on broom sticks, celestial apparitions in the form of speaking animals, and the defendants instructing insects to fly into their mouths and deposit bent nails and pins in their stomachs. *Id.* at 8-9. Ceci and Bruck relate that:

> On the basis of [this] testimony, 19 defendants were convicted and put to death, and a dozen more were spared execution when they threw themselves on the mercy of the court and admitted their participation in witchcraft. In the aftermath of these executions, some of the child witnesses publicly recanted their testimonies.

15 The authorship of this Article is rather unusual. One of us, Ceci, is a psychological researcher, who for nearly two decades has examined the question of children's suggestibility. He is one of three scholars whom Lyon identifies as being at the forefront of the "new wave" of suggestibility research. Lyon, *New Wave*, *supra* note 2, at 1010. The other author of this Article, Friedman, is a legal academic who writes on evidence law, among other areas. Neither of us claims any expertise in the field of the other. We have joined in this collaboration, however, because we believe it may be productive in leading to a better understanding of how the current state of knowledge of children's capacities and vulnerabilities should affect the operation of the legal system. We have both participated in writing the entire Article (the order of authorship is alphabetical only), and we are both comfortable with its entire contents. But neither of us presumes to offer evaluations or make recommendations outside the bounds of our own expertise; when we step beyond those bounds, we are each relying on our coauthor.

16 For fuller historical reviews, see CECI & BRUCK, *supra* note 11, at 47-61; Stephen J. Ceci & Maggie Bruck, *The Suggestibility of the Child Witness: A Historical Review and Synthesis*, 113 PSYCHOL. BULL. 403 (1993) [hereinafter Ceci & Bruck, *Historical Review*]; Gail S. Goodman, *Children's Testimony in Historical Perspective*, 40 J. SOC. ISSUES 9 (1984).

Scientific researchers have examined the question of children's testimonial competence for more than a century.[17] The work of the renowned French psychologist Alfred Binet offers a useful perspective for an overview of the early research.[18] Later studies have echoed several of his conclusions.

First, Binet concluded that, although older children and even adults are suggestible to a significant degree,[19] young children are more highly suggestible.[20] He argued that this suggestibility reflects the operation of two different factors, one cognitive and the other social.[21] The first factor, which he called "auto-sugges tion," develops as a response to a child's expectation of what is supposed to happen.[22] In contrast, the social factor is a desire to conform to the expectations or pressures of an interviewer, and thus reflects a form of mental obedience to another.[23]

---

17 The first scientific study appeared in Maurice H. Small, *The Suggestibility of Children*, 4 PEDAGOGICAL SEMINARY 176 (1896). Small asked school-aged children to raise their hand when they could smell the fragrance emitted from a bottle of clear liquid he uncorked in the front of the room. In actuality, the bottle contained water, yet many children claimed to smell its fragrance. *Id.* at 178. Furthermore, Small found that many of the children were influenced by their classmates' responses. *Id.* at 178, 180.

18 *See* ALFRED BINET, LA SUGGESTIBILITÉ (1900). Binet developed the Stanford-Binet test, still one of the most widely used intelligence tests for school-aged children.

19 BINET, *supra* note 16, at 2.

20 *Id.* at 390. *See* HAROLD ERNEST BURTT, APPLIED PSYCHOLOGY 252-54 (2d ed. 1957) (asserting that "children are more suggestible than adults"); Irene Case Sherman, *The Suggestibility of Normal and Mentally Defective Children*, COMP. PSYCHOL. MONOGRAPHS, Aug. 1924, ser. 9, at 1, 12 ("Previous experiments have indicated that suggestibility tends to decrease with age in normal children."); Ramona Messerschmidt, *The Suggestibility of Boys and Girls Between the Ages of Six and Sixteen Years*, 43 J. OF GENETIC PSYCHOL. 422-23 (1933) (summarizing Binet's conclusions); Margaret Otis, *A Study of Suggestibility of Children*, 11 ARCHIVES PSYCHOL. 5, 94-99 (1924) (concluding that resistance to suggestibility increases with mental age).

21 *See* Ceci & Bruck, *Historical Review*, *supra* note 14, at 406 (identifying two aspects of memory–recall and suggestibility); Philip S. Dale et al., *The Influence of the Form of the Question on Eyewitness Testimony of Preschool Children*, 7 J. PSYCHOLINGUISTIC RES. 269 (1978) (analyzing children's memory); Otto Lipmann, *Pedagogical Psychology of Report*, 2 J. EDUC. PSYCHOL. 253 (1911) (same). In the early years of the century, some researchers suggested that very young children often have difficulty distinguishing fantasy from reality. *See* JEAN PIAGET, JUDGMENT AND REASONING IN THE CHILD 202 (C. K. Ogden ed., Marjorie Warden trans., 1959) (1928).

22 BINET, *supra* note 16, at 86. In one of Binet's experiments, the children were presented with five lines of increasing length, followed by a series of "target" lines that were of the same length as the longest line of the series. Researchers then asked the children to draw each line as accurately as possible. The children tended to be influenced by the expectation of ever-increasing lines; that is, their reproductions of the target line were systematically too long because they expected that it would be longer than the line that had preceded it. Binet questioned the children after the study to determine why they had drawn the target lines so long and discovered that many were aware that the lines they had drawn were incorrect. In fact, they were able to redraw the lines more accurately on demand. Binet argued that this experiment demonstrated the children's ability to escape the influence of auto-suggestion. *See id.* at 110-60.

23 *See id.* at 209. For example, Binet explained that after being led to expect certain events, some children sometimes asserted that they had witnessed those events whether or not the events actually occurred. *See id.* at 241-43.

Other researchers also emphasize the significance of questioning by influential adults. According to the German psychologist William Stern, children view suggestive questions by influential adults as imperatives. William Stern, *Abstracts of Lectures on the Psychology of Testimony and on the Study of Individuality*, 21 AM. J. PSYCHOL. 270 (1910). The Belgian psychologist J. Varendonck conducted notable experiments in preparation for his testimony in a murder case in which child witnesses had been subjected to repeat,

Second, Binet reported that the examiner's language and method of questioning are among the external forces that can affect children's responses.[24] When children were asked for free recall—to write down everything they observed, without being aided by specific questions—they tended to provide little information, but the information they did provide was highly accurate.[25] Children responding to questions that focused their attention on a particular detail were somewhat less accurate than those giving free recall, but significantly more accurate than those asked either leading questions that suggested an inaccurate answer or questions that were misleading in that they assumed false information.[26] Modern commentators generally agree that there is a strong relationship between the nature of questioning and the accuracy of the response;[27] moreover, the arguments of Myers and Lyon depend quite heavily on the proposition that leading questions elicit more information.[28]

Third, Binet reported that children's answers to repeat questions are often characterized by exactness and confidence, regardless of their accuracy level.[29] Once a child gives an erroneous response, Binet surmised, it becomes incorporated into her memory.[30]

Fourth, Binet concluded that children are more suggestible in groups than when alone.[31]

---

suggestive questioning. J. Varendonck, *Les Témoignages d'Enfants dans un Procès Retentissant [Child Witnesses in a Famous Trial*], 11 ARCHIVES DE PSYCHOLOGIE 129 (1911). In one study, seven-year-old children were asked about the color of a teacher's beard. *Id.* at 138. Sixteen of eighteen children provided a response even though the teacher in question did not have a beard. *Id.* In another study, a teacher from an adjoining classroom came into Varendonck's classroom and without removing his hat talked in an agitated fashion for approximately five minutes. *Id.* at 141. After this teacher had left the classroom, the children were asked in which hand the teacher had held his hat. *Id.* Only three of the twenty-seven students claimed that the hat was not in his hand. *Id.*

24 BINET, *supra* note 16, at 245-49.

25 *See* BINET, *supra* note 16, at 255-56.

26 BINET, *supra* note 16, at 294.

27 *See, e.g.*, Ceci & Bruck, *Historical Review*, *supra* note 14, at 406 (noting Binet's finding that free recall questions produce the most accurate answers); Jacqueline L. Cunningham, *Contribution to the History of Psychology: XLVI. The Pioneer-Work of Alfred Binet on Children as Eyewitnesses*, 62 PSYCHOL. REP. 271, 273 (1988) (noting that "[r]esults showed the now familiar phenomenon that fewer but more accurate details are recalled by subjects who report information spontaneously in comparison with those who respond to specific questioning").

28 *See, e.g.*, Lyon, *New Wave*, *supra* note 2, at 1046; Myers, *Psychological Research*, *supra* note 2, at 11-26.

29 BINET, *supra* note 16, at 324-25. Even among adults, there is often a low correlation between an eyewitness's confidence and accuracy. *See* Robert K. Bothwell et al., *Correlation of Eyewitness Accuracy and Confidence: Optimality Hypothesis Revisited*, 72 J. APPLIED PSYCHOL. 691, 694 (1987).

30 *See* BINET, *supra* note 16, at 325. Binet based this belief on the fact that, in contrast to the auto-suggestion study in which children could later redraw the line correctly, children in the study testing the effect of the examiner's language were unable to correct their wrong answers later. *Id.* at 324.

31 *Id.* at 359. When researchers asked a group of three children a series of misleading questions and told them to call out the answer to each question as quickly as possible, the children who responded second and third were most likely to give the same answer as the first respondent—even if that answer was inaccurate. *Id.* at 351-56.

The late-17th-century Swedish witch trials present an interesting analog to Binet's findings. Sjöberg analyzed statements made by 805 children to parish priests or a "Royal Commission of inquiry" and reported that the children were more likely to claim they had witnessed extraordinary events if they gave their testimony to the parish priest after waiting in line with other witnesses to attend prayer meetings. Sjöberg, *supra* note 12, at 1042-43.

Early in the century, other researchers reached results consonant with Binet's and drew conclusions that continue to find support today—most notably, they concluded that repeat questioning can have a particularly powerful effect. For example, the German psychologist William Stern concluded that a child is more likely to remember her answers to earlier questions than the underlying events themselves.[32]

This early research is of limited usefulness in analyzing issues of forensic significance. First, although some of the early researchers had forensic uses in mind, the subject matter of the questions they posed bore little resemblance to the subject matter of statements that children give in actual cases. In the early experiments, researchers often asked children leading questions about details that the children likely regarded as peripheral and of little significance. For instance, a researcher might ask questions about the color of a stranger's beard,[33] which of several lines was longer,[34] or whether the child smelled an odor when the questioner opened a mysterious-looking bottle containing only water.[35] In contrast, in actual forensic investigations—most of which involve abuse of the child—the interviewer usually questions the child about bodily actions that, if they occurred as alleged, were experienced rather than merely witnessed by the child herself; that were central to the event in question; and that are frequently associated with embarrassment, fear, and pain.[36]

Thus, whatever the early experiments might show about the reliability of children under the conditions of the experiments, they lack external or ecological validity for the context of principal contemporary significance.

Sjöberg concluded that these children were influenced by the other witnesses waiting in line. *Id.* He wrote that "only 59% of the children testifying at other places than prayer meetings were sure about the real life quality of their experiences of the witches' sabbath whereas as many as 91% were sure about it after standing in line at prayer meetings." *Id.* In addition, Sjöberg reported that the youngest witnesses in the Swedish witch trials, who ranged from one to six years old, were significantly more suggestible than older children. *Id.*

32 *See* Stern, *supra* note 21, at 274.

33 *See* Varendonck, *supra* note 21, at 138.

34 *See* BINET, *supra* note 16, at 284.

35 *See* Small, *supra* note 15, at 178. Gail S. Goodman & Alison Clarke-Stewart, *Suggestibility in Children's Testimony: Implications for Sexual Abuse Investigations*, *in* THE SUGGESTIBILITY OF CHILDREN'S RECOLLECTIONS 92 (John Doris ed., 1991), extends this criticism to some of the more modern studies as well:

> [M]ost research on children as eyewitnesses has relied on situations that are very different from the personal involvement and potential trauma of sexual abuse. Researchers have used brief stories, films, videotapes, or slides to simulate a witnessed event. A few have used actual staged events, but these events—for example . . . [a] man tending plants—are also qualitatively different from incidents of child abuse. The children are typically bystanders to the events, there is no bodily contact between the child and adult, and it is seldom even known whether the events hold much interest for the children. Of even more importance . . . , the questions the children are asked often focus on peripheral details of the incident, like what the confederate was wearing, rather than on the main actions that occurred or, more to the point, whether sexual acts were committed.

Id.

36 In other cases, these conditions do not necessarily hold. Thus, Varendonck claimed that the types of questions that he used in his studies were similar to those used with one of the child witnesses in a murder case in which he participated. Varendonck, *supra* note 21, at 137-38. In a case like that one, where the child witness is not the victim of the crime and may not even have observed it, she may have information to offer that is important to the case even though it did not have much salience for her.

That is, they cannot be relied on with confidence to show how suggestible children are in the real-world context of the greatest interest—when a child makes an allegation about personal abuse.

Nevertheless, this brief historical review indicates that recent research on the suggestibility of children—their susceptibility to leading questions, peer pressure, and repeated questioning, the tendency to perceive conditions in conformity with the expectancies created by adults, and the need to comply with adults' wishes—is not a modern departure from earlier understandings. On the contrary, it fits squarely within what has been the dominant view for the last century.

B. The Modern Era

1. *An Overview*

Research into children's suggestibility was virtually non-existent mid-century. However, the late 1970s witnessed a resurgence of interest in this area, and this interest has continued. This virtual explosion of research was fueled by various factors, most significantly a dramatic increase of reports of child abuse and growing recognition of the commonness of abuse.[37] Thus, researchers began to focus on the reasons so many children failed to report abuse and on the fact that when children did allege abuse their reports were often met with skepticism.

As discussed below in Part I.B.2, researchers have responded to increased abuse by emphasizing the potential value of suggestive or other directed questioning in securing disclosure of abuse.[38] Of course, directed questioning also raises the problem of false positives.[39] Some researchers, notably Gail Goodman, have contended that in settings of particular importance—when the questioning concerns a matter involving events in which the child was a participant or a victim, rather than a mere bystander, and the questioning addresses central aspects rather than peripheral details of the event—children are much less suggestible than the early experiments show.[40] Part I.B.3 discusses Professor Goodman's research and demonstrates that, notwithstanding the optimistic interpretation scholars have sometimes superimposed on it, the data actually indicate that even in such circumstances young children can be substantially suggestible.

Part I.B.4 illustrates that suggestibility is an even greater problem—even in matters involving intimate contact between an adult and a child—when interviewers use more strongly suggestive procedures.

---

37 *See* Ceci & Bruck, *Historical Review*, *supra* note 14, at 408. Ceci and Bruck note that other factors included greater receptivity by courts to expert psychological testimony, the increased focus of social scientists on socially relevant issues (including children's rights and protection of minors), and increased interest in the study of eyewitness testimony of adults. *See id.*

38 *See, e.g.*, Goodman & Clarke-Stewart, *supra* note 33, at 103; Lyon, *New Wave*, *supra* note 2, at 1046-49; Myers, *Psychological Research*, *supra* note 2, at 9.

39 *See e.g.*, CECI & BRUCK, *supra* note 11, at 18.

40 *See* Goodman & Clarke-Stewart, *supra* note 33, at 95, 103; Goodman et al., *Children's Concerns and Memory: Issues of Ecological Validity in the Study of Children's Eyewitness Testimony*, *in* KNOWING AND REMEMBERING IN YOUNG CHILDREN, 249, 256 (Robyn Fivush & Judith A. Hudson eds., 1990).

Whether a child has been subjected to such strong suggestion is a matter for the judicial system to determine on the facts of the particular case; those facts, rather than general practice, must control the response of the system. In Part I.B.5, however, we demonstrate that the use of suggestive questioning has been a pervasive problem, not merely an occasional one.

In considering the modern research, we think it is a mistake to frame the debate as a conflict between two irreconcilable groups of scientists. Scholars working in this area are far more interactive[41] and closer in agreement with one another than a reader might infer from portions of Lyon's article. As Lyon recognizes, scholarship by Ceci and Bruck has received broad professional and academic endorsement.[42] Furthermore, the amicus brief on suggestibility written by the Committee of Concerned Social Scientists and submitted in support of the defense in the famous case of Kelly Michaels[43] was signed by forty-three of the forty-six memory researchers who were asked to do so.[44] Goodman was one of the three who declined to sign the brief, but Lyon recognizes that "several of the nation's most well-respected researchers in psychology" were among the forty-three signers.[45] Goodman's works, as we will show, present evidence that strongly supports Ceci and Bruck's arguments. The reverse is also true; Ceci and Bruck's writings are also replete with provisos about children's strengths as witnesses. Virtually all research in the scientific mainstream, including that of Goodman, pays at least some attention to the dangers of both false positives and of false negatives.[46] Most scholars in the field produce work that is helpful in some contexts to the defendant and in some contexts to the prosecution or plaintiff; indeed, several of the scholars who are part of what Lyon calls the "new wave" have consulted for both sides in litigation.[47] As one would expect, the diverse scholars in this field vigorously disagree about some matters of

---

41 For example, Goodman has published some of her work in Ceci's volumes and has participated in several symposia he has organized. The same is true in reverse: Ceci has appeared in symposia organized by Goodman, and Poole, Bruck, and others, either to collaborate on writing projects with Goodman or to write chapters for volumes she edits.

42 *See* Lyon, *New Wave*, *supra* note 2, at 1010 (citing Ceci and Bruck's *Historical Review*, *supra* note 14, which was named the best article of the year on child abuse by the Society for the Psychological Study of Social Issues of the American Psychological Association (APA)). In addition, Ceci and Bruck's *Jeopardy in the Courtroom* won the APA's annual William James Book Award for 2000 for the book that best integrates research across diverse areas of psychology. CECI & BRUCK, *supra* note 11.

43 State v. Michaels, 642 A.2d 1372 (N.J. 1994).

44 *See* Brief *Amicus Curiae* of the Developmental, Social, and Psychological Researchers, Social Scientists and Scholars, State v. Michaels, 642 A.2d 1372 (N.J. 1994), *reprinted in* Maggie Bruck & Stephen J. Ceci, *Amicus Brief for the Case of* State of New Jersey v. Michaels *Presented by Committee of Concerned Social Scientists*, 1 PSYCHOL. PUB. POL'Y & L. 272, 312-14 (1995) (listing the signatories).

45 Lyon, *New Wave*, *supra* note 2, at 1012 n.39. One of the signatories was Alison Clarke-Stewart, who coauthored one of the papers on which Lyon relies. *See id.* (citing Goodman & Clarke-Stewart, *supra* note 33).

46 There is also no stylistic divide between groups. Lyon asserts that "[t]he new-wave researchers are not only scientists, but also storytellers who disseminate their most impressive subjects as aggressively as their cumulative data." Lyon, *New Wave, supra* note 2, at 1084 (footnote omitted). Attendance at international conferences during the past decade, however, would suffice to convince any observer that the use of vivid horror stories is not the exclusive province of one type of researcher.

47 This list includes Ceci, Poole, Lamb, Warren, and Bruck. Some, like Ceci, have not accepted remuneration for consulting on child sexual abuse cases.

interpretation. These disagreements are not merely between one group and another—assuming the groups can be well defined—but are within the groups as well.

2. *The Benefits of Directed Questioning*

Since the time of Binet, psychologists have understood that a child's free recall tends to be more accurate than her responses to suggestive questioning.[48] However, free recall also tends to be extremely sparse. When asked for free recall, children usually give correct but very brief answers, and they often omit important details.[49] This is especially true for very young children, particu larly in the abuse context; as Lyon emphasizes, fear, embarrassment or loyalty may inhibit a child from disclosing abuse.[50]

Abuse investigators therefore often use more directed and focused approaches, such as leading and repeated questions, thereby attempting to secure useful information from the child. Some modern research highlights the potential value of these techniques. For example, Karen Saywitz and Gail Goodman conducted a study discussed approvingly by Myers and Lyon.[51] The study showed that when girls whose pediatric examinations had included an exterior vaginal and anal examination were asked for their free recall, only eight of thirty-six (22%) correctly mentioned the vaginal touch, and only four of thirty-six (11%) mentioned the anal touch.[52] Directed questioning with the aid of anatomically correct dolls raised the numbers to thirty-one (86%) and twenty-five (69%),

---

48 *See supra* notes 23-25 and accompanying text.

49 *See* BINET, *supra* note 16, at 255-56, 294.

50 Lyon, *New Wave*, *supra* note 2, at 1048. According to Lyon, Ceci and Bruck argue that "threats do not suppress disclosure." *Id.* at 1060. Though Ceci and Bruck do make statements to that effect, they present mixed results from several studies of different types. *See* CECI & BRUCK, *supra* note 11, at 35, 141 n.1, 263-65, 301. We concede Lyon's point that studies of disclosure that examine only those children who eventually are identified as having been abused are imperfect indices of the efficacy of threats, because they may miss many children who were threatened into never disclosing. However, in addition to the real-world studies, laboratory studies by Peters and others seem to indicate that threats, fears, and other forces often do not affect ultimate disclosure after some passage of time and when the child feels safe. *See id.* In laboratory studies, we know with certainty which children were threatened.

We are unclear what Lyon means when he writes that "Ceci and Bruck omit the discussion of Peters's work as documenting a reluctance to disclose transgressions from their 1995 book," Lyon, *New Wave, supra* note 2, at 1050. In fact, Ceci and Bruck write:

> [P]erpetrators may use threats . . . to silence their victims. There have been a few studies that have examined the effectiveness of such threats in a laboratory setting. Peters (1990) staged an event in which children between 5 and 9 years of age saw a "thief" come into the testing room and steal a book. The "thief" told the child not to tell. When later questioned about the event by a parent, only 5% of the children disclosed the theft if the thief was also present. If the thief left the room, however, 67% of the children told about the theft.

See CECI & BRUCK, *supra* note 11, at 145 n.1 (citation omitted); *see also* LUCY S. MCGOUGH, CHILD WITNESSES: FRAGILE VOICES IN THE AMERICAN LEGAL SYSTEM (1994) (reviewing other studies concerning threats).

51 *See* Karen J. Saywitz et al., *Children's Memories of a Physical Examination Involving Genital Touch: Implications for Reports of Child Sexual Abuse*, 59 J. CONSULTING & CLINICAL PSYCHOL. 682 (1991) (cited in Lyon, *New Wave*, *supra* note 2, at 1017-18 & n.59, and in 1 JOHN E.B. MYERS, EVIDENCE IN CHILD ABUSE AND NEGLECT CASES § 1.27, at 65 (3d ed. 1997)).

52 Saywitz et al., *supra* note 49, at 686-87.

respectively.[53] There is no serious doubt that directed questioning will often be far more effective than requests for free recall in securing disclosure of abuse.[54]

This is one side of the coin. The risk of the false positives potentially created by suggestive questioning is the other.

3. *Studies of Suggestibility by Goodman and Her Colleagues*

The modern research on suggestibility is profuse and varied. Rather than attempt a full summary, we discuss four studies conducted by Gail Goodman and her colleagues. We have chosen Goodman because she is the scholar most favored by child advocates. Ceci and Bruck have written that "[p]erhaps no researcher has done more to redress the historical imbalance in favor of child witnesses than Gail Goodman."[55] Lyon, who relies heavily on Goodman's work, suggests that she "became the researcher-heroine of the child protection movement in the 1980s because her research supported claims that false allegations of abuse rarely, if ever, occur."[56] And yet her studies provide strong evidence that children, especially young children, are suggestible to a significant degree—even on abuse-related questions.[57]

a. *The Pediatric Exam Study*

We have already described one part of a study by Saywitz and Goodman, in which girls whose genitalia and anus were touched during a pediatric examination were much more likely to report that touching occurred in response to doll-aided directed questioning than in response to open-ended questions.[58] The other part of the study posed the same questions to girls whose genitalia had *not* been touched during the exam.[59] The vast majority of these girls correctly denied a genital touch.[60] But one out of thirty-five (2.86%) did answer affirmatively when

---

53 *Id.* Against such results must be placed studies indicating that children tend to respond more informatively to open-ended invitations when they have been conditioned to answer such questions, rather than more focused questions, in the early part of the interview. Kathleen J. Sternberg et al., *Effects of Introductory Style on Children's Abilities to Describe Experiences of Sexual Abuse*, 21 CHILD ABUSE & NEGLECT 1133 (1997).

54 *See* Lyon, *New Wave*, *supra* note 2, at 1064; *cf. id.* at 1067 (interpreting one study as providing "a compelling anecdote that children may be reluctant to affirm events that they believe are naughty").

55 Ceci & Bruck, *Historical Review*, *supra* note 14, at 410.

56 Lyon, *New Wave*, *supra* note 2, at 1015 (footnote omitted).

57 As Lyon points out, Goodman was one of only three researchers, out of forty-six asked to cosign the amicus brief on suggestibility in support of the defense in the famous Kelly Michaels case, who declined to do so. *Id.* at 1011-12.

58 *See supra* notes 48-50 and accompanying text. In this study, the mean rates of correct answers in response to misleading abuse questions ranged from 96% to 99%. *See* Saywitz et al., *supra* note 49, at 688. But when the children were asked direct, non-misleading questions about potentially abusive events (for example, "Did the man kiss you?"), their accuracy rates were not as high, ranging from 77% to 87% accuracy. *Id.* The basis for this perplexing finding is unclear. Ceci and Bruck have suggested that perhaps the reason is that the misleading questions were so unusual and "unmotivated" (in other words, the studies did not give the children any reason to assent to misleading questions by the use of threats, bribes, peer pressure, or cajoling) that the children perceived them as ridiculous or silly, and easily rejected them. Ceci & Bruck, *supra* note 14, at 426. In contrast, the nonmisleading questions were at least plausible to the children. *See id.*

59 Saywitz et al., *supra* note 49, at 684.

60 *See id.* at 687.

asked about a genital touch, and two out of thirty-six girls (5.56%) answered affirmatively when asked about an anal touch.[61] Saywitz and her colleagues concluded that "although there is a risk of increased error with doll-aided direct questions, there is an *even greater risk* that not asking about vaginal and anal touch leaves the majority of such touch unreported."[62]

b. *The Delayed Inquiry Study*

In another study, Goodman and her colleagues asked three- to six-year-olds to play a game with a strange man for approximately five minutes.[63] During this time, the man did not engage in any sexually provocative behaviors.[64] Four years later, the researchers reinterviewed fifteen of these same children, now between seven and ten years old, and asked them what they could recall of their prior experience with the strange man.[65] Not surprisingly, the children could not remember much.[66] Then, to create an "atmosphere of accusation," the interviewers asked questions such as: "Are you afraid to tell?" and "You'll feel better once you've told."[67] Goodman wrote that "[t]he children were more accurate on the abuse than the nonabuse questions."[68] Nevertheless, these children were quite susceptible to abuse-related questioning. Four of the fifteen children agreed with the interviewer's false suggestion that the stranger had kissed or hugged them; four out of the fifteen agreed that the stranger had taken pictures of them; and one child even agreed she had been given a bath by the stranger.[69] Goodman and her colleagues acknowledged that some of these errors "might well lead to suspicion [of abuse]."[70]

c. *The Trailer Study*

Rudy and Goodman conducted a third study in which pairs of four- and seven-year-olds, eighteen children of each age, were left in a trailer with a strange adult.[71] One child watched while the adult played

---

61 *Id.*

62 *Id.* at 690 (emphasis added).

63 Gail S. Goodman et al., Children's Testimony Nearly Four Years After an Event 5 (1989) (unpublished manuscript, on file with author Ceci).

64 *Id.*

65 *Id.* at 5-6.

66 *Id.* at 7-8.

67 *Id.* at 7.

68 *Id.* at 15.

69 *Id.* at 12.

70 *Id.* at 14.

71 Leslie Rudy & Gail S. Goodman, *Effects of Participation on Children's Reports: Implications for Children's Testimony,* 27 DEVELOPMENTAL PSYCHOL. 527, 529 (1991). In this study, as in the Pediatric Exam study, the children's answers to misleading abuse questions were correct more frequently than were their answers to directed but nonmisleading questions about potentially abusive events, such as "Did the man kiss you?" *See id.* at 532-33 In the Trailer Study, the means for the first type ranged from 88% to 94%, *see id.* at 533, and from 82% to 90% for the second type. *See id.* at 532. Again, the explanation may be that the children were sufficiently confident to reject the misleading questions out of hand. Indeed, Goodman and her colleagues, in discussing the Trailer Study, noted that the children often giggled in response to the misleading questions or otherwise acted in a way reflecting their perception of the question as silly. Goodman et al., *supra* note 38, at 266 ("We also noticed that children's demeanor changed once we

games with the other, helped her dress in a clown's costume, lifted her onto a desk, and took two photographs of her.[72] Ten to twelve days later, Rudy and Goodman asked each child various questions about the incident.[73] Some of the questions involved actions that might be of special concern in child abuse investigations, such as, "How many times did he spank you?" and "Did he put anything into your mouth?"[74] Rudy and Goodman reported that the "[s]even-year-olds did not make a single commission error to the specific abuse questions."[75] The four-year-old participants "made very few commission errors," while the four-year-old bystanders "evidenced a slightly higher, but still low, error rate"; these error rates were 3% and 7%, respectively.[76]

d. *The Mt. Sinai Study*

Goodman and her colleagues also conducted an experiment involving 108 children between the ages of three and fifteen who were examined at Chicago's Mt. Sinai Hospital as part of a five- day assessment of children with suspected histories of abuse.[77] The sound idea behind this experiment was that researchers could gain ecological validity by studying children who were actually involved in abuse investigations.[78] On day one of their stay, the children received a medical check-up.[79] On day two, the abused children were given an anogenital examination and swabbed for culture.[80] On day five, the children were interviewed, and the interview included misleading or other suggestive questions.[81]

Despite fairly substantial levels of incorrect answers to misleading, abuse-related questions by the pre-schoolers, Goodman and her colleagues developed an optimistic reading of the data:

> [D]espite performing more poorly than their older counterparts, the 3-to-5-year-olds still demonstrated *relatively good resistance* to misleading information in answering the abuse-related questions. When presented with misleading questions related to abusive or inappropriate behavior by the doctor and/or nurse (e.g., "How many

---

began to ask the abuse questions. Many showed signs of embarrassment by giggling or smiling. Others looked surprised.").

72 Rudy & Goodman, *supra* note 69, at 529-30.

73 *Id.* at 530.

74 *Id.* at 529.

75 *Id.* at 532.

76 *Id.*

77 Mitchell L. Eisen et al., *Memory and Suggestibility in Maltreated Children: New Research Relevant to Evaluating Allegations of Abuse*, *in* TRUTH IN MEMORY 163, 175-77 (Steven Jay Lynn & Kevin M. McConkey eds., 1998). Approximately 31% of these children were classified as having been sexually abused, 10% as physically abused, 10% as neglected, 16% as having addicted parents, and 22% as a control group (that is, neither abused nor neglected). *Id.* at 178 tbl.7.1. In preliminary analyses the authors found no significant differences in errors among these various groups of children, so they lumped them together in reporting their results and instead focused on age differences. *Id.* at 178-80.

78 The authors expressed the view that, because of various problems that they suffered, this group might be more suggestible than the children typically involved in suggestibility studies. *See id.* at 172-73. "It is also possible," they wrote, "that abused children are hypervigilant regarding abusive actions or abuse suggestions and, as a result, would be more resistant to such questioning than nonabused children." *Id.* at 175.

79 *Id.* at 176.

80 *Id.*

81 *Id.*

> times did the doctor kiss you?"), 3- to-5-year-olds answered 79% of the questions without making commission errors.[82]

Of course, the unstated implication is that this group did make commission errors in answering 21% of the misleading abuse-related questions. Moreover, the authors reported that the preschool-aged children gave incorrect answers to 40% of all misleading, abuse-related questions.[83] They also pointed out that "approximately 40% of the errors made by 3- to 5-year-olds in response to the misleading abuse-related questions were produced by only 6 of the 29 children in this group."[84] Thus, although the group's "proportion of commission errors to misleading abuse-related questions was relatively low on average," some children were "more error-prone than others."[85] "If such children were interviewed in an abuse investigation," the authors acknowledge, "a false accusation could potentially result."[86] Children in the older groups performed substantially better, but still answered a nontrivial percentage of the misleading abuse-related questions incorrectly—16% for the six- to ten-year-olds, and 9% for the eleven- to fifteen-year-olds.[87]

These studies are each important for understanding children's intellectual development and for revealing the underlying mechanisms of suggestibility and memory. But any use of them for forensic purposes must take into account six points.

First, as Lyon acknowledges, "calling nondisclosure the `greater risk' implicitly assumes that both kinds of risk are weighed equally."[88] That is a dubious assumption. Consider two polar cases. If false disclosures could be disproved without cost, then, for reasons we discuss in Part II, our system might tolerate a large number of them to generate some true disclosures; the initial disclosure would thus be treated as a screening device. If, at the other extreme, it is certain that a potential defendant would be convicted given a disclosure (whether true or not) and acquitted without a disclosure, then, for reasons we also discuss in Part II, our system treats the cost of a false disclosure as far greater than the cost of failure to make an accurate disclosure. Thus, it is far from clear that nondisclosure is the greater risk simply because it is numerically more prevalent.

Second, the probability that a child would answer any given question falsely should not be confused with the probability that a child would answer some questions falsely. For example, in the Trailer Study, the 3% and 7% error rates for four-year-old participants and bystanders in response to misleading abuse questions must be understood in conjunction with the fact that there were only nine children in each group; this means that *at least one of the nine children* gave at least one false, affirmative answer to an abuse-related question.

---

82 *Id.* at 179 (emphasis added).

83 *Id.* at 179 tbl.7.2. We are unsure of the basis for the difference between this 40% value and the 21% implied by the quote above. Perhaps the basis is the distinction between abuse-related questions that concerned the doctor or nurse, versus abuse-related questions concerning others.

84 *Id.* at 179.

85 *Id.* at 180.

86 *Id.*

87 *Id.* at 179.

88 Lyon, *New Wave*, *supra* note 2, at 1019 n.74.

Third, in the above studies, the children responded less accurately to the so-called "direct" questions concerning abuse than to affirmatively misleading questions. It may be that the content of the misleading questions made them more difficult to accept, or that some subtle verbal or nonverbal cues in the phrasing of these questions alerted children to the fact that they were in fact misleading.[89] Whatever the explanation, the most important measure of children's vulnerability to suggestion is their reliability in response to the type of question *most* likely to lead them into misstatement.

Fourth, one must avoid the temptation, into which several scholars have fallen, to draw the following type of *non sequitur*:

> The children in the studies by Goodman and associates were generally accurate in reporting specific and personal things *that had happened to them*. If these results can be generalized to investigations of abuse, they suggest that normal children are unlikely to make up details of sexual acts *when nothing abusive happened*.[90]

The argument is illogical. In determining the probability of a false positive, the premise is that "nothing abusive happened"; if abuse *did* occur, then there cannot be a false positive. Thus, it seems mistaken in this context to speak of abuse-related questions as being central or salient to an event—by hypothesis there *was* no such event. We would expect older children and adults to be better equipped to reject false suggestions about events that *would* be perceived as salient and central had they happened, because these older individuals would recognize that had such events occurred they would likely have left vivid memory traces. This is called an "intuitive theory of memory," meaning that adults try to determine whether a recollection is real (that is, due to actual experience) or fictitious (for example, due to suggestion, fantasy, or imagination) by asking themselves whether it is likely that they could have had such a salient experience and forgotten it.[91] Younger children are much less likely to have this capacity to interrogate the workings of their own memory.[92]

Fifth, even if we were to assume that the probability is only 3% that a child would make a statement wrongfully alleging abuse, this does *not* mean that, if a child does make a statement alleging abuse, the probability is only 3% that abuse did not occur. Equating these two probabilities—the probability of the event given the hypothesis and the probability of the hypothesis given the event—is a common error, often referred to as "transposing the conditional" or, in some settings, as the "prosecutor's fallacy."[93] If, apart from the child's statement, the probability is very low

---

89 *See supra* note 56 and accompanying text.

90 John E.B. Myers, *Taint Hearings for Child Witnesses? A Step in the Wrong Direction*, 46 BAYLOR L. REV. 873, 922 n.232 (1994) (emphasis added) (quoting Goodman & Clarke-Stewart, *supra* note 33, at 102-03).

91 *See generally* Michael Ross, *Relation of Implicit Theories to the Construction of Personal Histories*, 96 PSYCHOL. REV. 341, 341-42 (1989) (analyzing the process of personal recall).

92 *See* John H. Flavell, *The Development of Children's Knowledge About the Mind: From Cognitive Connections to Mental Representations*, *in* DEVELOPING THEORIES OF MIND 244, 246-47 (Janet W. Astington et al. eds., 1988).

93 *See, e.g.*, BERNARD ROBERTSON & G.A. VIGNAUX, INTERPRETING EVIDENCE: EVALUATING FORENSIC SCIENCE IN THE COURTROOM 19-20, 91-93 (1995) (defining the concept of transposing the conditional and the prosecutor's fallacy); David J. Balding &

that the child was abused, then even a 3% probability that a child who was not abused would say she was could result in a very high probability that the child was not abused. Suppose that a random sample of the child population, only 2% of whom have been abused, is questioned in such a way that 90% of those who have been abused but only 3% of those who have not been abused will confirm that they have been abused. Then more than 60% of those who answer that they have been abused will be speaking falsely.[94] We return to this problem, which Lyon recognizes, in Part II.

Finally, although the four studies we have discussed used suggestive questions—"Did the doctor touch you there?"; "How many times did he spank you?"; "Did he put anything into your mouth?"; "How many times did the doctor kiss you?"—they did not use highly suggestive *techniques* such as repeated questioning over time,[95] coercion, or peer pressure.[96] Moreover, in three of the four studies, the suggestive questions were embedded in neutral or supportive interviews. These studies therefore pose weak tests of young children's vulnerability to suggestion. They do not indicate the limits on how a combination of motives, strong suggestions, threats, and inducements might lead a child to make an inaccurate report. As we now show, these highly suggestive techniques can produce much higher error rates.

To recap, even if one looks no further than the body of research favored by child advocates, that of Gail Goodman and her associates, the proportion of false claims—that is, the proportion of children who were not exposed to a given type of behavior who nevertheless asserted that they were—ranged between 3 and 40%. Even the lowest end of this range can lead to unacceptably high decision-making errors in some circumstances, as we demonstrate with the aid of probability theory in Part II. But, as we have already suggested, this research does not indicate the rate of false claims that occur when the child is subjected to stronger forms of suggestion. We now develop this latter point.

4. *The Impact of Highly Suggestive Techniques*

As we have indicated, the studies described above may underestimate the susceptibility of young children to stronger suggestions.

Peter Donnelly, *The Prosecutor's Fallacy and DNA Evidence*, 1994 CRIM. L. REV. 711, 716-17.

94 Suppose there are 10,000 children, 2% of whom (200) were abused, and 90% of those (180) confirm that they have been abused. There are 9,800 nonabused children, and if only 3% of them say they have been abused, that yields 294 false positives. Thus, of the 10,000 children, 474 say they were abused—but this is false in 294 cases, or about 62% of the time.

95 *See generally* Debra Ann Poole & Lawrence T. White, *Tell Me Again and Again: Stability and Change in the Repeated Testimonies of Children and Adults*, *in* MEMORY AND TESTIMONY IN THE CHILD WITNESS 24 (Maria S. Zaragoza et al. eds., 1995) [hereinafter Poole & White, *Tell Me Again and Again*] (reporting on the effects of question repetition in test subjects); Debra A. Poole & Lawrence T. White, *Two Years Later: Effects of Question Repetition and Retention Interval on the Eyewitness Testimony of Children and Adults*, 29 DEVELOPMENTAL PSYCHOL. 844 (1993) [hereinafter Poole & White, *Two Years Later*] (same).

96 This is a point that the authors recognized. *See* Saywitz et al., *supra* note 49, at 690-91. It was principally this factor that led Maggie Bruck, in an unfortunate choice of words (for which Lyon calls her to task) in giving testimony, to refer to Saywitz's study as "meaningless." *See* Lyon, *New Wave*, *supra* note 2, at 1018, 1020.

To test this hypothesis, researchers have conducted a number of studies incorporating stronger forms of suggestion—techniques that have been used by investigators in some well-publicized child abuse cases.[97] Among these stronger forms of suggestion that researchers have shown to lead to false assertions are the following: repetition of questions within the same interview,[98] stereotype inducement,[99] guided imagery,[100] peer pressure, and selective reinforcement.[101] As we demonstrate below, numerous studies show that when children are exposed to these forms of suggestion the error rates can be very high, sometimes exceeding 50%. Moreover, this phenomenon holds true even when the questions concern events that supposedly affect the child herself, as opposed to events to which she was a mere bystander, even when the questions are central, rather than peripheral, to the alleged event, and even when the questions concern abuse-related matters. As we did with respect to the research favored by child advocates, we will discuss here only a small selection of the relevant studies.

Sena Garven and her colleagues used strong suggestions—such as reinforcing answers that were consistent with interviewers' hunches and invoking pressure to conform—based on the use of those same tactics in the McMartin daycare sexual abuse case.[102] Garven and her colleagues found a 58% false claim rate as to various behaviors in which an adult had supposedly engaged, versus only a 17% error rate when weaker suggestions were used.[103] In a follow-up publication, these researchers found between 35 and 52% false claims, including statements that the adult had tickled the child's tummy or had kissed the child on the nose, in response to strong suggestions, versus 13 to 15% when strong suggestions were not used.[104] Other researchers who have used these stronger

---

97 *See* CECI & BRUCK, *supra* note 11, at 9-18.

98 *See* Poole & White, *Tell Me Again and Again, supra* note 93, at 36; Poole & White, *Two Years Later, supra* note 93, at 851.

99 *See* Michelle D. Leichtman & Stephen J. Ceci, *The Effects of Stereotypes and Suggestions on Preschoolers' Reports*, 31 DEVELOPMENTAL PSYCHOL. 568, 572 (1995). Stereotype inducement is the creation of false expectations that lead subsequently to false reports consistent with those expectations.

100 Ceci and Bruck describe a forensic example of guided imagery in which a therapist engaged the child-client in fantasy play about being trapped in a crash with the defendant. CECI & BRUCK, *supra* note 11, at 214-15. A research example is an experiment in which the researcher encourages a child to imagine an event that allegedly transpired (for example, getting her hand caught in a mousetrap and having to go to the hospital to get the hand released) by providing perceptual cues, such as what the child was wearing when she supposedly by went to the hospital or how her finger was bandaged. Stephen J. Ceci et al., *The Possible Role of Source Misattributions in the Creation of False Beliefs Among Preschoolers*, 42 INT'L J. CLINICAL & EXPERIMENTAL HYPNOSIS 304, 306-07 (1994).

101 Sena Garven et al., *More than Suggestion: The Effect of Interviewing Techniques from the McMartin Preschool Case*, 83 J. APPLIED PSYCHOL. 347, 356 (1998). Selective reinforcement refers to interviewer conduct giving more positive signals to statements by the child consistent with the interviewer's hypothesis than to statements inconsistent with that hypothesis.

102 *Id.* at 348-50.

103 *Id.* at 354.

104 Sena Garven et al., *Allegations of Wrongdoing: The Effects of Reinforcement on Children's Mundane and Fantastic Claims*, 85 J. APPLIED PSYCHOL. 38 (2000). The following exchange, which occurred in the interviewers' study, is an example of a combination of conformity pressure with positive reinforcement:

> I: The other kids say that Paco took them to a farm. Did Paco take you to a farm?

suggestive techniques also have reported high error rates for children's claims that a strange man "put[] something yucky into their mouths" during a visit to a science exhibit,[105] took off their clothes and kissed them,[106] or touched them inappropriately.[107]

Studies focusing on repeated questions include a study by Bruck and her colleagues in which three-year-olds were repeatedly asked strongly suggestive questions about a doctor touching their anogenital regions, such as, "Show me on the doll how Dr. F. touched your genitals."[108] Among girls whom the doctor did not touch, fully 50% falsely claimed the doctor had inserted objects into their anogenital cavities.[109] After a third exposure in a period of a week to an anatomically correct doll, one three-year-old child reported that her pediatrician had strangled her with a rope, inserted a stick into her vagina, and hammered an earscope into her anus.[110]

Similarly, Steward and Steward and their associates interviewed children aged three to six four times after a pediatric clinic visit.[111] With each interview these researchers conducted, children's false reports of anal touching increased; by the final interview, which took place six months after the initial visit, more than one-third of the children in this study falsely reported anal touching.[112]

Poole & White interviewed some four-, six-, and eight-year-olds, and adults immediately after a staged encounter with a man, and again one week later.[113] They interviewed another group of subjects only once, after

---

C: Yes.

I: Great. You're doing excellent now.

Id. at 41. The specific false claims that various children in this study made concerning Paco Perez's conduct, in addition to those mentioned in the text, were that Paco Perez (a) tore a book, (b) stole a pen from the teacher's desk, (c) broke a toy, (d) told the child a secret, (e) threw a crayon at a child who was talking, (f) said a bad word, (g) took the child on a helicopter ride, (h) took the child to a farm, and (i) took the child on a horse ride. *Id.* at 41-43.

105 Debra A. Poole & D. Stephen Lindsay, *Interviewing Preschoolers: Effects of Nonsuggestive Techniques, Parental Coaching, and Leading Questions on Reports of Nonexperienced Events*, 60 J. EXPERIMENTAL CHILD PSYCHOL. 129, 143 (1995).

106 *See* Stephen J. Lepore & Barbara Sesco, *Distorting Children's Reports and Interpretations of Events Through Suggestion*, 79 J. APPLIED PSYCHOL. 108, 112 (1994).

107 *See* Jane Mary Rawls, *How Question Form and Body-parts Diagrams Can Affect the Content of Young Children's Disclosures* (paper presented at the NATO Advanced Study Institute, Recollections of Trauma: Scientific Research and Clinical Practice, 1996), summarized in E-mail from Winston Wealleans to Stephen J. Ceci (Apr. 11, 1996, 20:55:41) (on file with author Ceci).

108 Maggie Bruck et al., *Anatomically Detailed Dolls Do Not Facilitate Preschoolers' Reports of a Pediatric Examination Involving Genital Touching,* 1 J. EXPERIMENTAL PSYCHOL.: APPLIED 95, 100 (1995).

109 *Id.* at 102. Among boys, the error rates were lower, in the neighborhood of 20%. *Id.* This differential is explained by the fact that the girl doll had one more aperture than the boy doll, giving the girls additional opportunity for insertion of props or fingers. *Id.*

110 *Id.* at 106.

111 *See* Margaret Steward et al., *Interviewing Young Children About Body Touching and Handling*, *in* 61 MONOGRAPHS OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT v (4-5 Ser. 248) (1996).

112 *See id.* at 116 tbl.31, 123-24.

113 Debra A. Poole & Lawrence T. White, *Effects of Question Repetition on the Eyewitness Testimony of Children and Adults*, 27 DEVELOPMENTAL PSYCHOL. 975, 976- 77 (1991).

a delay of one week.[114] Some of the repeated questions were open-ended, such as "What did [the man] look like?", whereas others were closed or yes-or-no, such as "Did the man hurt Melanie?"[115] Poole and White reported that the "repeat-interview" four-year-olds were significantly more likely than the "single-delayed" ones to give a false affirmative answer to the question, "Did the man hurt Melanie?"; the repetition of the open-ended questions, in contrast, did not result in more errors.[116] But there was no difference between the two groups with respect to the question, "Did the man ask nicely for the pen?"[117]

Finally, in a study by Rawls, thirty five-year-olds and seven four-year-olds participated in a series of benign play events with a male adult.[118] Over the course of four interviews, the children were asked both open-ended questions, such as, "Where were you with X?" and "What sort of things did you do with him?", and closed questions, such as, "Do you know why he touched that part of your body?"[119] Rawls reported that nearly a quarter of her sample falsely claimed the man inappropriately touched them, with three of the children (10%) falsely reporting genital touching, two (7%) falsely reporting anal touching, and two additional children reporting mutual adult-child touching (for example, claiming the adult pretended to rub cream into their bodies).[120] And, finally, in the so-called "Monkey Thief" Study, which Lyon discusses,[121] Bruck and Ceci found that over half of the youngest children made false claims of witnessing theft of food in their daycare facility after exposure to repeated suggestions and pressures.[122]

These studies and others like them indicate quite clearly that if children, especially very young children, are subjected to highly suggestive questioning techniques, their rates of false claims, even on abuse-related questions, may be very high—far higher than reported in studies not using such highly suggestive techniques. Of course these studies have limited utility for forensic purposes unless children are in fact exposed to these techniques in the real world of abuse investigation. We now turn to that question.

---

114 *Id.* at 977.

115 *Id.*

116 *Id.* at 978-81.

117 *Id.* at 981. Regardless of the subjects' age and gender, those in the repeat condition reported that the man hurt Melanie (60%) much more frequently than those in the single interview condition (33%). *Id.*

118 Rawls, *supra* note 105.

119 *Id.*

120 According to Rawls, "[R]eports of mutual undressing without touching were also common, although this often reflected a confusion between dress-up items and ordinary clothes." *Id.*

121 *See* Lyon, *New Wave*, *supra* note 2, at 1024-26.

122 *See* Bruck et al., *Children's Reports of Pleasant and Unpleasant Events*, *in* RECOLLECTIONS OF TRAUMA 199, 203, 205 fig.1, 211 (J. Don Read & D. Stephen Lindsay eds., 1997). For additional examples of false claims—involving bodily touching or witnessing a bicycle theft—see, for example, William S. Cassel & David F. Bjorklund, *Developmental Patterns of Eyewitness Memory and Suggestibility: An Ecologically Based Short-Term Longitudinal Study*, 19 LAW & HUM. BEHAV. 507, 522- 23 (1995) (bicycle theft experiment); Peter A. Ornstein et al., *The Influence of Prior Knowledge on Children's Memory for Salient Medical Experiences, in* MEMORY FOR EVERYDAY AND EMOTIONAL EVENTS 83, 88-90 (Nancy L. Stein et al. eds., 1997) (doctor experiment).

5. *Suggestiveness in the Real World*

The research on the effects of suggestive questioning has real-world applicability *at least* to the extent of revealing the risks of such questioning. The research should therefore be a factor in deciding whether, in given circumstances, suggestive questioning is appropriate. But child advocates contend that research indicating the suggestibility of children has little real-world applicability in judging the credibility of children because it is predicated on a false assumption about the nature and incidence of suggestive questioning by investigative interviewers. Thus, Lyon says that "[t]he new wave . . . . present[s] a distorted picture of the suggestibility problems in the typical case."[123] Myers similarly accuses Ceci and Bruck of implying falsely that "the norm in child abuse investigations is to interview children multiple times and to overuse and abuse leading questions."[124] This issue is the linchpin of the child advocates' argument. Indeed, Lyon is probably accurate in saying, "[d]ifferences between the new wave and Goodman and her colleagues may derive more from differing assumptions about what interviews are like than from differing beliefs about children's vulnerability."[125]

At the outset, it is important to bear in mind that, so far as any decisions the legal system must make in a given case are concerned, the question of suggestiveness fundamentally depends on the facts of the particular case and not on the commonness or uncommonness of a given style of interrogation. If a given style of questioning was used in the particular case before the court, then research related to it would be highly relevant—no matter how unusual that type of questioning may be. This point is a double-edged sword.

On one hand, we fully agree that studies involving a given type of suggestion will have little or no bearing on a case in which that type of suggestion was absent.[126] For example, as Lyon points out, the Monkey Thief Study involved asking children several suggestive questions during each of several interviews, which also included peer pressure and encouragement to visualize an incident.[127] Error rates as high as those observed in this study are unlikely in cases that contain far fewer suggestions. Ceci and Bruck themselves have made this point repeatedly.[128]

---

123 Lyon, *New Wave*, *supra* note 2, at 1013.

124 Myers, *New Era, supra* note 2, at 394.

125 Lyon, *New Wave*, *supra* note 2, at 1027-28.

126 *Cf.* Ceci & Bruck, *Historical Review*, *supra* note 14, at 433 (listing various conditions that affect interviewees). Ceci and Bruck note:

> If the child's disclosure was made in a nonthreatening, nonsuggestible atmosphere, if the disclosure was not made after repeated interviews, if the adults who had access to the child prior to his or her testimony are not motivated to distort the child's recollections through relentless and potent suggestions and outright coaching, and if the child's original report remains highly consistent over a period of time, then the young child would be judged to be capable of providing much that is forensically relevant. The absence of any of these conditions would not in and of itself invalidate a child's testimony, but it ought to raise cautions in the mind of the court.

127 Lyon, *New Wave*, *supra* note 2, at 1039.

128 For example, see Stephen J. Ceci & Maggie Bruck, *Child Witnesses: Translating Research into Policy*, SOC. POL'Y REP., Fall 1993, at 18-19:

On the other hand, if it is clear—as it has been in some notorious cases[129]—that highly suggestive techniques *were* used in the particular case, then research on that type of suggestibility obviously becomes a significant matter, without regard to whether those techniques are typical or rare.

Lyon acknowledges this latter point.[130] But, he argues,

> how interviewers conducted interviews is largely unknown in many, if not most, cases. Although many jurisdictions require videotaping or taping of investigatory interviews, most do not. Furthermore, it would be impractical to impose a requirement that individuals record the first contact with the child giving rise to a suspicion of abuse because such contact arises between children and parents or teachers, rather than during a formal abuse investigation. . . . In sum, to make judgments in individual cases, courts often must make assumptions about how interviewers typically interview children.[131]

Two aspects of this argument strike us as most curious. First, for investigatory interviews, a videotaping requirement is certainly feasible; Lyon concedes that many jurisdictions follow this practice.[132] Given this flexibility, if failure to record an investigatory interview leaves substantial uncertainty about how the interview was conducted, courts should not generally resolve that uncertainty in favor of the prosecution.[133]

Second, the fact that the first contact giving rise to a suspicion of abuse is most often with parents or teachers, not with official investigators, is cause for concern rather than for comfort. Official investigators may be trained to avoid suggestiveness; most parents and teachers are not. Remarkably, Lyon asserts that "[p]arents are unlikely to pursue the hypothesis that a spouse or a brother has abused their child."[134] Obviously, this is often true—and just as obviously it is sometimes not

---

> Measures can be taken to lessen the risk of suggestibility effects. To date, the factors that we know most about concern the nature of the interview itself—its frequency, degree of suggestiveness, and demand characteristics.
>
> > A child's report is less likely to be distorted, for example, after one interview than after several interviews (the term "interviews" here includes informal conversations between parents and child about the target events).
>
> . . . .
>
> Thus, at one extreme we can have more confidence in a child's spontaneous statements made prior to any attempt by an adult to elicit what they suspect may be the truth. At the other extreme, we are more likely to be concerned when a child has made a statement after prolonged, repeated, suggestive interviews. Unfortunately, most cases lie between these extremes and require a case-by-case analysis.

129 *See, e.g.*, CECI & BRUCK, *supra* note 11, at 9-11 (the Little Rascals Day Care Case), 11-13 (the Kelly Michaels Case), 13-14 (the Old Cutler Presbyterian Case), 14-16 (the Country Walk Babysitting Service Case).

130 Lyon, *New Wave*, *supra* note 2, at 1026 ("Obviously, if one knows whether a particular child was interviewed with suggestive techniques, then one need not ask what most interviews are like.").

131 *Id.* at 1026-27.

132 *Id.* at 1026. We argue in favor of the practice *infra*, Part III.E.

133 *Cf.* Commonwealth v. Allen, 665 N.E.2d 105, 108-09 (Mass. App. Ct. 1996) (noting that "neither the initial interview, nor [most of the] subsequent interviews . . . were videotaped or recorded, and that, without such recordings, it is difficult for the defendant to substantiate an offer of proof in support of a request for a pretrial competency hearing" (footnote omitted)).

134 Lyon, *New Wave*, *supra* note 2, at 1032.

true.[135] The data on parental support show great diversity in parental reactions to the suspicion of child abuse by their current or former partners.[136] Many experts can cite cases—in acrimonious custody disputes, for example—in which one parent is accused of coaching the child against the other parent.[137] We do not know what percentage of custody cases involve such acrimony, but clearly many do. Even if this percentage is very low, it would nevertheless amount to many cases, given the large number of custody disputes. Moreover, suggestive questioning by a parent or teacher—an important and influential person to the child—may yield more false positive statements than questioning by an investigator whom the child does not know.

Having stated these threshold points, we will now concentrate on the question of the frequency of suggestive techniques in formal investigations. Of course, child advocates do not deny that investigators frequently question children suggestively.[138] Indeed, much of their argument is that some forms of suggestive questioning are necessary and involve little risk.[139] Any question that articulates a specific proposition and asks the child to confirm or deny it must be considered suggestive to a considerable degree. Thus, Lyon concedes that "`leading' questions are certainly common in investigative interviews."[140] But child advocates do contend that there is no basis for concluding that the stronger forms of suggestiveness are common.[141]

First, consider the mere fact of repetitive questioning, which, as we have shown, can in itself (and in conjunction with other suggestive techniques) be a powerful producer of false positives.[142] Myers chides Ceci and Bruck for contending that the norm is multiple interviews—though deep in a footnote he "fully concede[s] that too many children are interviewed too many times by too many professionals."[143] Such evidence as exists suggests strongly that multiple interviews by government agents, for both social services and law enforcement, are very common.[144] Of course, those interviews almost inevitably follow at least

---

135 *See, e.g.*, D. Corwin & E. Olafson, *Videotaped Discovery of a Reportedly Unrecallable Memory of Child Sexual Abuse: Comparison with a Childhood Interview Videotaped 11 Years Before*, 2 CHILD MALTREATMENT 91, 92 (1997) (reporting admission on tape by a child witness that her mother and grandmother pressured her to make a false claim against her father).

136 See the studies cited by Lyon, *New Wave*, *supra* note 2, at 1056 nn.288-89. Although many mothers do not support their children's disclosures of abuse, many *are* supportive, especially if the defendant is an estranged husband or partner rather than a current one. *See id.* at 1056-57. In many studies, the support rate is between 50% and 85%. *See id.* at 1056 nn.288-89.

137 This is Ceci's personal observation from discussion with clinical experts.

138 *See, e.g.*, Lyon, *New Wave*, *supra* note 2, at 1036 ("In sum, the limited observational research on real-world interviews demonstrates that interviewers ask few open-ended questions, many specific questions and some leading questions." (footnote omitted)).

139 *See, e.g.*, *id.* at 1064-68.

140 *Id.* at 1042.

141 *See supra* notes 121-23 and accompanying text.

142 *See supra* note 96 and accompanying text.

143 Myers, *New Era*, *supra* note 2, at 395 n.53 (citations omitted).

144 Some of the evidence on multiple interviews is summarized in Stephen J. Ceci et al., *Children's Allegations of Sexual Abuse: Forensic and Scientific Issues: A Reply to Commentators*, 1 PSYCHOL. PUB. POL'Y. & L. 494, 500-01 (1995). Neither Myers nor anybody else presents any evidence suggesting that multiple interviews are *not* the norm.

one conversation that the child had with a parent, teacher, or another familiar adult.[145] Indeed, Myers contends that "[i]n quite a few cases, multiple interviews are necessary," precisely because "[m]any children who have been abused do not disclose their abuse during the first interview."[146]

Now consider the substance of the interviews. Ceci and Bruck discuss some notorious cases involving allegations of abuse in day-care centers, in which the persistent use of highly suggestive techniques is apparent beyond question.[147] Lyon's response is essentially one of confession and avoidance: he contends that investigators are more likely to use highly suggestive techniques in the day-care cases than in the more typical case charging abuse against only one child.[148] He is right, of course, that if a case involves only one alleged victim—the most common type of case, but not the only significant type, and not clearly the type that covers a majority of the alleged victims[149]—peer pressure tech niques will ordinarily not be available.[150] Beyond that, however, his argument is very thin.[151] He asserts speculatively that "[i]nterviewers who are confident that the children have suffered abuse are more likely to question

---

Myers does report one study of a hospital-based, multidisciplinary child interview center that succeeded in reducing multiple interviews, so that 80% of the children interviewed there were interviewed only once. MYERS, *supra* note 49, § 1.31, at 73-74. A study of another multidisciplinary investigation program, however, revealed no change in the number of interviews. *See id.* § 1.31, at 72-73; *see also* Myers, *New Era*, *supra* note 2, at 395 n.53 (citing a report by the California Attorney General's Office that "describ[es] *pilot projects* that succeeded in *reducing* the number of interviews" (emphases added)). Such joint investigations may eventually become a major factor in reducing multiple interviews, but they have not yet. And, even if only 20% of children were subjected to multiple interviews, that would still be a significant problem.

145 *Cf.* MYERS, *supra* note 49, § 1.27, at 63 ("When a child finally discloses abuse, the child may tell a friend, parent, teacher, school counselor, or other professional." (footnotes omitted)).

146 *Id.* § 1.32, at 78 & n.367.

147 *See* CECI & BRUCK, *supra* note 11, at 178-80 (giving one of many documented examples presented in that volume).

148 Lyon, *New Wave*, *supra* note 2, at 1032.

149 According to figures Lyon cites, about 85% of criminal sexual abuse cases charge the defendant with abusing a single child, another 11% of the cases name two victims, and 5% of the cases name three or more victims. *Id.* at 1031 n.141. Even if all 5% of the cases in the last (three-or-more) category involved three victims rather than four or more, this would nevertheless mean that 37 of 122 alleged victims (30%) were involved in multivictim indictments. But of course a few of the cases in the three-or-more category involve very large numbers of children, which means that the total percentage of alleged victims involved in multivictim indictments is over 30%. Moreover, Lyon acknowledges that, according to his source, in 22% of the cases in which a defendant was charged with assaulting a single child, there were allegations of abuse against other children that were not charged. *Id.* Thus, it appears from these figures that an actual majority of the suspected victims of abuse are involved in cases in which there is at least one other suspected victim.

150 Lyon cites a study by Amye Warren and her colleagues as showing that only three of forty-two interviews used peer pressure techniques. *Id.* at 1033. Most of Warren's cases involved only a single child, however, and so no peer claims about the alleged abuse were available to interviewers. *See* Amye R. Warren et al., *"It Sounds Good in Theory, But . . .": Do Investigative Interviewers Follow Guidelines Based on Memory Research?*, 1 CHILD MALTREATMENT 231, 233 (1996). Of those cases involving a peer, the use of peer pressure tactics was no doubt far higher.

151 It is in this context that Lyon contends that "[p]arents are unlikely to pursue the hypothesis that a spouse or a brother has abused their child," Lyon, *New Wave*, *supra* note 2, at 1032, an argument to which we have responded above. Similarly, he contends that "interviewers are not likely to paint negative stereotypes of those with whom the child may wish to maintain an ongoing relationship." *Id.*

extensively a child in a multivictim case than in a single-victim case."[152] That is not at all clear; if more children are potential sources of information, the time spent with each child might diminish, even if the case becomes more important. Moreover, one might think that in light of the bizarre allegations made in some cases involving multiple children, fair-minded investigators would not have been so confident that the children suffered abuse.[153] Lyon also argues that "[t]he median age of a sexual abuse victim in criminal court is thirteen years of age, while the day-care cases predominantly involved preschool children."[154] But this misses the point. No one contends—and surely we do not—that the research on preschool children is a guide to suggestibility in adolescents. The preschool research is of substantial use only in cases involving preschool children—but those cases are very numerous.[155] In fact, three- to five-year-olds are the single highest incidence group for substantiated sexual abuse (more than five out of every 1000).[156] In sum, a very large number of cases are multivictim cases involving preschool-aged children, notwithstanding Lyon's suggestions to the contrary.

Now narrow the focus to the question of principal interest to Lyon, the use of leading questions in cases involving a single alleged victim. Lyon perceives a whipsaw manipulation of the term "leading question": "When discussing the real world, the new wave uses the term broadly [but] in describing their own research, the new wave uses the term quite narrowly."[157] In other words, he suggests that some researchers have used narrow definitions of leading questions to show that they have large consequences, and then broad definitions to show that they are used frequently in the real world.

It is true that there is no consistent taxonomy in the research literature; different studies attach different meanings to terms such as "leading," "suggestive," "focused," "open-ended," "closed," and "direct." Ultimately, however, this definitional matter is of little importance. Our analysis of Goodman's studies has shown that even mild suggestiveness—which is plainly a pervasive interviewing technique[158]—can have more than a trivial impact.[159] Furthermore, research on the actual practices of interviewers has shown that some frequently used techniques are highly suggestive by any definition and raise a particularly high danger of false positives in light of suggestibility research. Examples of highly suggestive interviewing techniques may be

---

152 *Id.* at 1031.

153 *See* CECI & BRUCK, *supra* note 11, at 10 (detailing such allegations as "being taken by boat and thrown overboard into a school of circling sharks").

154 Lyon, *New Wave*, *supra* note 2, at 1032 (footnote omitted).

155 *See supra* note 1.

156 SEDLAK & BROADHURST, *supra* note 1, at 4-13 fig.4-6.

157 Lyon, *New Wave, supra* note 2, at 1037.

158 *See infra* notes 177-82 and accompanying text.

159 *See supra* Part I.B.3.

found very readily, even in the pages of the *United States Reports*.[160] We will concentrate, however, on three sets of studies Lyon discusses.[161]

In analyzing these studies, Lyon focuses on the ratio of suggestive questions to all questions asked by an interviewer.[162] This ratio, however, is of very limited significance. After all, a single strong suggestion can damage the accuracy of a child's report. Furthermore, surrounding it with a long series of non-suggestive questions is not likely to diminish its deleterious impact. Even if only 5% of all questions asked by a real-world interviewer are suggestive, this may nevertheless mean that all interviews contain at least one strong suggestion. Thus, the ratio of suggestive to nonsuggestive questions is far less useful as a metric than the total proportion of interviews in which at least some serious suggestion was made or the average number of suggestive questions per interview. We will nevertheless discuss the former metric to some extent, for two reasons. First, it is one that some of the studies assess. Second, if this ratio is more than trivial, it clearly indicates a high probability that an interview includes one or more suggestive questions.

a. *Warren*

Lyon argues that Amye Warren and her colleagues, who studied transcripts of interviews conducted by child protection service workers in a southern state, found only relatively innocuous leading questions.[163] But, according to Warren herself, this is a misinterpretation of her findings.[164] Warren says that she wrote the article Lyon cites in a

---

160 Idaho v. Wright, 497 U.S. 805, 808, 810-11 (1990) (relating an interview in which a pediatrician asked a 2 1/2-year-old child: "Do you play with daddy?" "Does daddy play with you?" "Does daddy touch you with his pee-pee?" "Do you touch his pee-pee?").

161 Lyon also cites a study, Barbara W. Boat & Mark D. Everson, *Concerning Practices of Interviewers When Using Anatomical Dolls in Child Protective Services Investigations*, 1 CHILD MALTREATMENT 96 (1996), which he says reports that interviewers using anatomical dolls never "asked the child to show how abuse might have occurred" in ninety-seven interviews. *See* Lyon, *New Wave*, *supra* note 2, at 1034 & n.156; Boat & Everson, *supra*, at 101 & tbl.4. This citation is at least potentially misleading. True, the interviews—only eighty-one of which actually included use of dolls—contained no instances of certain troublesome practices, such as the introduction of fantasy or conjecture by the interviewer (e.g., "If someone has hurt a little girl's bottom, show me how he would do it.") *when using the dolls as an icebreaker*. *Id.* This may be the basis for Lyon's contention that this study supports the proposition that guided imagery is uncommon in practice. But in fact most of the interviewers asked the children to use the doll to demonstrate how abuse occurred—not in itself a troublesome practice—and a high percentage of the interviews did include practices that the authors regarded as troublesome. For example, among the forty-five interviews with two- to five-year-old children in which the interviewers used the dolls, 29% used the dolls prematurely, and 20% used play language that might encourage fantasy (*e.g.*, "We can play like this is you."). In one of the forty-five cases, the interviewer asked for a demonstration with the aid of the doll before the child had made a verbal disclosure. *Id.* at 100-01 & tbl.6. Also in one of the cases, the interviewer introduced fantasy or conjecture to assess a child's labels for or understanding of anatomy. *Id.* at 101 tbl.5. Overall, Boat & Everson express a tone of concern; although the frequency of troublesome practices does not support discontinuing use of the dolls, they believe that the frequency of improper use indicates the need for better training. *Id.* at 103.

162 *See* Lyon, *New Wave*, *supra* note 2, at 1036.

163 *Id.* at 1034-35.

164 *See* E-mail from Amye Warren to Stephen J. Ceci (Sept. 7, 1999, 12:01:44) (on file with authors) [hereinafter Warren E-mail]. Warren said that she was "distressed" on reading the section of Lyon's work describing her findings and Lamb's:

> I think his description of our findings is a misinterpretation. We (Nancy Walker and I) tried to be so cautious in our description of the interviews so that we

cautious manner; published in the journal of the American Professional Society on the Abuse of Children (APSAC),[165] it attempted to avoid sensationalism so as not to put its audience on the defensive, in the hope that some mild suggestions for improving practices would be adopted. Moreover, the article was primarily devoted to an examination of interviewers' usage of recommended practices rather than their avoidance of improper techniques. The article directly analyzed only one particular problem, the provision of new, potentially leading information that the child had not disclosed in the same interview.[166] Thus, the authors did not intend the article to be a comprehensive listing of their findings, nor a list of the most serious errors these researchers have found. Even in this article, however, Warren and her colleagues gave examples of suggestive questioning that could hardly be deemed innocuous. Thus, they reported that "three interviewers on their second attempt used a potentially highly suggestive method by stating that someone else had told them that something had happened to, or that someone had touched or hurt, the child, and then followed with `Do you remember?' or `Can you think about that?'"[167] And they concluded that "[f]ewer than half of our interviewers introduced the abuse-questioning phase in a general, open-ended fashion designed to elicit a narrative response, and some introductory techniques could be perceived as highly suggestive."[168]

Elsewhere, Warren and her associates have described even more serious errors made by their interviewers than the examples listed in the article Lyon discusses.[169] To summarize, Warren's interviewers

---

> would not be perceived as attacking—we did not want the audience to become defensive—our hope being that they might be motivated to put some of these better practices to use. So in fact we never even attempted to classify all the questions as leading versus not, nor did we provide some of the worst examples of leading questions. We purposely avoided being "sensationalistic". Instead, Nancy and her student looked at the interviewers' "provision of new, POTENTIALLY leading, information" that had not been previously disclosed by the child in the same interview. And we acknowledged that sometimes that information was innocuous—like the child's name or the names of her brothers and sisters, etc. We also said that referring to what the child said in a prior interview IS problematic—because if the child was led in a prior interview, then asking them to say what they said again would be repeating the errors. But I do NOT believe, by any means, that this is a "worst case scenario" of leading questioning, as Lyon has represented this example. Having looked through these interviews again, I have found worse, obviously suggestive questions.

Id.

165 *See* Warren et al., *supra* note 148. APSAC describes itself as "the nation's largest interdisciplinary professional society for those who work in the field of child maltreatment," with its aim being "to ensure that everyone affected by child maltreatment receives the best possible professional response." MYERS, *supra* note 49, § 1.38, at 119 n.*.

166 *See* Warren E-mail, *supra* note 162.

167 Warren et al., *supra* note 148, at 238.

168 *Id.* at 239.

169 The following are two examples of the many suggestions that these interviewers employed:

> Interviewer (I): Who else's wiener have you seen?
> Child (C): Mine.
> I: You don't have one. You're bein' silly again. . . . Did S. touch (you) with his wiener?

Amye Warren, *Interviewing Child Witnesses: Some Linguistic Considerations* #21, at 8-9, invited presentation at the NATO Advanced Studies Institute on the Child Witness in Context: Social, Cognitive, and Legal Perspectives, Il Ciocco, Italy (May, 1992).

> I: We're gonna go see him sometime, Policeman Bob. Do you know why we're gonna go see him?

provided details not previously provided by the child, including ones that concern the alleged abuse; they created an atmosphere of accusation by alluding to the bad nature of the defendant; they invoked peer pressure; and they sometimes raised negative consequences, such as calling a child silly, when her answers deviated from expectation. Nearly all the interviews in this sample contained suggestions, and the majority of the interviewers asked questions that could be considered problematic.[170]

C: [Nods]
I: Why?
C: [Shrugs]
I: `Cause I want you to tell him about Mr. L. Do you know why I want you to tell him?
C: [Nods]
I: Why?
C: [Shrugs]
I: `Cause I don't want Mr. L. doin' that to anybody else. And we have to tell a policeman. Do you know that?
C: [Nods]
I: We have to tell a policeman what Mr. L. did. But my question is do we need to tell the policeman about anybody else?
C: [Shakes head no]
I: Do we need to tell the policeman about Miss D.?
C: Mm-hmmm [yes].
I: What do you want to tell him?
C: About [unintelligible]
I: Well now they already know about her, so we don't need to talk any more about her. She did bad things, and then we told the police, and we've told the police about the bad things Mr. L. did to your tootie. But is there anything else we need to tell him?
C: [Shrugs]
I: `Cause we're not gonna tell him about Ro. cause he's a little kid and we're not gonna get him in trouble.

Id., #24, at 3. Some of the interviewer behaviors revealed in these exchanges are clearly troublesome, but Warren and her associates do not discuss them in the article to which Lyon refers, in part because suggestive questioning was not the primary focus of the analyses conducted for the article. Additional examples of highly suggestive techniques occur in other of Warren's works. *See id.*; Amye R. Warren et al., *Interviewing Children: Questions of Structure and Style*, presented at the 3d National Colloquium of the American Professional Society on the Abuse of Children, Tucson, AZ (June 1995); Amye R. Warren et al., *Interviewing Child Witnesses: Beyond Leading Questions*, Biennial Meeting of the American Psychology-Law Society, San Diego, CA (March 1992).

170 Recently, Warren and her colleagues have presented a paper, based on the same interview transcripts, that in their view provides some support for Lyon's contention that the most egregious practices are used only rarely in "typical" cases. Amye R. Warren, et al., *Setting the Record Straight: How Problematic Are "Typical" Child Sexual Abuse Interviews?* (unpublished manuscript, March 2000, on file with authors). This paper certainly supports the proposition that their typical interview contained fewer egregious practices than the most notorious interviews in the McMartin and Kelly Michaels cases, a point that we have never disputed. But the data revealed in this paper indicate that some particularly troublesome techniques, though usually constituting a small part of the interaction in any given interview, are extremely common. Perhaps most strikingly, the interviewers invoked negative consequences, such as telling the child "you haven't told us anything," in twenty-eight of the forty-two interviews, or 67%. In forty of the interviews, or 95%, the interviewer repeated a question, in an attempt to elicit a new answer, even though the child had unambiguously answered the question in the immediately preceding portion of the interview. In five of the interviews, or 12%, the interviewer told the child about information received from another person. In 37 interviews, or 88%, the interviewer invoked positive consequences on at least one occasion for an answer, though the researchers report that this occurred mainly in the context of the early, rapport-building part of the interview. And in fourteen interviews, or 33%, the interviewer invited the child to speculate about past events or to use imagination or solve a mystery. All of these techniques have been shown by research to be extremely problematic, and this new data by Warren and her colleagues reveals that they are present in anywhere from 12%-95% of all interviews.

b. *Lamb*

In one of a series of studies, Michael Lamb and his colleagues examined interviews by Israeli investigators with alleged sexual abuse victims.[171] They found that only about 2% of the utterances invited open-ended responses from the child.[172] About 25% of the utterances were leading questions, defined as questions that "focus the child's attention on details or aspects of the account that the child has not previously mentioned, but do not imply that a particular response is expected."[173] About 9% of the utterances—an average of more than fourteen per interview—were "suggestive" in that they implied a desired response or assumed details that the child had not provided.[174] An example from their research of a suggestive question is, "He forced you to do that, didn't he?"[175] Lamb and his colleagues obtained similar results in studying interview samples from the United States,[176] as well as from other Israeli samples.[177] The exact results differed from study to study and within each study depending on the conditions under which the interviews were held. But, as Lyon acknowledges, a typical interview across these studies contained between five and ten suggestive statements.[178]

Furthermore, these authors offer several caveats: the interviews, provided by an agency, may not be typical; the coding rules were strictly applied, and while they may have registered some violations that were probably harmless "conversational conventions," they also missed some potentially more harmful errors; and, most importantly, as we have argued above, the frequency or infrequency of a particular problematic technique in a given interview does not mean that the use of that technique had little impact. Indeed, a single use of some of the techniques listed above is capable in some circumstances of distorting the entire interview. Finally, nothing in the recent paper undercuts the findings described above that Warren and her colleagues made in earlier papers concerning the use of other suggestive techniques.

171 Michael E. Lamb et al., *Effects of Investigative Utterance Types on Israeli Children's Responses*, 19 INT'L J. BEHAV. DEV. 627 (1996).

172 *Id.* at 631-32, 633 tbl.1.

173 *Id.* at 631, 633 tbl.1.

174 *Id.* at 631, 633 tbl.1.

175 *Id.* at 631.

176 *See* Michael E. Lamb et al., *Investigative Interviews of Alleged Sexual Abuse Victims with and Without Anatomical Dolls*, 20 CHILD ABUSE & NEGLECT 1251 (1996); Kathleen J. Sternberg et al., *The Relation Between Investigative Utterance Types and the Informativeness of Child Witnesses*, 17 J. APPLIED DEV. PSYCHOL. 439 (1996) [hereinafter Sternberg et al., *Investigative Utterance Types*]; Kathleen J. Sternberg et al., *Using a Scripted Protocol in Investigative Interviews: A Pilot Study*, 3 APPLIED DEV. SCI. 70 (1999) [hereinafter Sternberg et al., *Scripted Protocol*].

177 *See* Irit Hershkowitz & Aline Elul, *The Effects of Investigative Utterances on Israeli Children's Reports of Physical Abuse*, 3 APPLIED DEV. SCI. 28 (1999); Yael Orbach et al., *Assessing the Value of Structured Protocols for Forensic Interviews of Alleged Child Abuse Victims*, 24 CHILD ABUSE & NEGLECT 733 (2000); Sternberg et al., *supra* note 51.

178 Lyon, *New Wave, supra* note 2, at 1036. Kathleen Sternberg and her colleagues conducted a pilot study analyzing investigatory sexual abuse interviews conducted by a sheriff's office in an urban area of a southern state in the United States. Sternberg et al., *Scripted Protocol*, *supra* note 174, at 71. Investigators who were already experienced in interviewing child sexual abuse victims participated in intensive training seminars, lasting approximately forty hours, to learn one of two types of protocols. One of these relied on open-ended questions in the opening phases of the interview and the other used directed questions. *Id.* at 71-72. Sternberg then compared interviews conducted by these interviewers before and after the training and found that, although training reduced the number of directive and leading utterances used in the substantive phase of the interview, neither type of training significantly reduced the number of suggestive utterances used. *See id.* at 74-75. Prior to training, an average of 9.33 of the interviewers' substantive

Lyon belittles the significance of this research on several grounds. First, he suggests that leading questions in these studies are not troublesome because they do not imply that the interviewer expects a particular response.[179] But given that a leading question, by definition, articulates a proposition that the interviewee had not previously articulated, it almost inevitably suggests to the interviewee that the questioner regards the proposition as plausible.[180]

Second, Lyon appears to argue that the relatively low proportion of suggestive questions means that the use of these questions is not a troubling factor.[181] Once again, however, the ratio of suggestive questions to all questions asked is a statistic of very little significance. Rather, the critical factor is that, both before and after being trained in the protocol, the interviewers studied by Lamb and his colleagues used a significant number of suggestive questions per interview.[182] Given Lamb's narrow definition of suggestion as a statement implying a favored reply from the child,[183] even a single instance could call into doubt the validity of an interview—and all the more so the five to ten such instances that are typical of the interviews Lamb studied.[184]

Third, Lyon says that "[t]he data do not reveal the extent to which the suggestive questions elicited details of the alleged abuse."[185] We confess to being mystified by this assertion. Not only do Lamb and his colleagues state the number of suggestive inquiries per interview in most of their studies, but they also report explicitly the average number of

---

utterances per interview, or 8.0% of all their substantive utterances, were suggestive. *Id.* at 74 tbl.2. Following training in the directed protocol, they used an average of 8.5 suggestive utterances (11.7% of all their substantive utterances) per interview, and with the open-ended protocol the average number of suggestive utterances was 7.33 per interview (9.9%). *Id.* The researchers found that interviewers who did not adhere to a protocol "obtained most of the information from the children using leading and suggestive questions" and that about 50% of the post-training interviews "could not be included because the interviewers did not adhere to the protocol." *Id.* at 74-75.

In another study, interviewers using anatomical dolls made an average of 8.00 suggestive utterances per interview (8.2% of the total), while interviewers not using the dolls made an average of 4.88 suggestive utterances per interview (7.2%). Lamb et al., *supra* note 174, at 1255 tbl.1.

Other studies by Lamb and those who have worked with him also have found that interviewers make several suggestive utterances per interview, between 5 and 10% of all their utterances. *See, e.g.*, Hershkowitz & Elul, *supra* note 175, at 31 tbl.2 (3.2; 6.7%); Orbach et al., *supra* note 175, at 741 tbls.1 & 2 (9.1; 10%); Sternberg et al., *Investigative Utterance Types*, *supra* note 174, at 446 tbl.2 (9.5; 8.7% for multiple incident abuse investigations; 9.9% for single incident).

179 Lyon, *New Wave*, *supra* note 2, at 1035.

180 The same point applies to questions Warren and her colleagues classify as specific because they do not require a narrative response. In the studies by Lamb and his colleagues, leading questions only focused on new topics. *See, e.g.*, Lamb et al., *supra* note 174, at 1254.

181 *See* Lyon, *New Wave*, *supra* note 2, at 1036 ("In sum, the limited observational research on real-world interviews demonstrates that interviewers ask few open-ended questions, many specific questions, and some leading questions." (footnote omitted)).

182 *See, e.g.*, Sternberg et al., *Scripted Protocol*, *supra* note 174, at 74 tbl.2.

183 Throughout their writings, Lamb and his colleagues define suggestive utterances as those "stated in such a way that the interviewer strongly communicates what response is expected (e.g., `He forced you to do that, didn't he?')." Lamb et al., *supra* note 169, at 631; Sternberg et al., *Scripted Protocol*, *supra* note 174, at 73.

184 *See, e.g.*, Lyon, *New Wave,* supra note 2, at 1036.

185 *Id.* (footnote omitted).

details related to the alleged abuse that children provided in response to each type of utterance.[186] In a study on Israeli children, for example, in which they report an average of more than fourteen suggestive interviewer utterances per interview,[187] they also report an average of 2.02 details per suggestive utterance[188]—a figure that is typical across their studies.[189] In addition, Lamb and his colleagues make very clear that suggestive questions were a principal basis on which the interviewers in these studies elicited abuse-related information.[190]

Finally, Lyon contends that the Israeli investigators involved in Lamb's study were not adequately trained;[191] his implication is that American interviewers are better trained, and so the rate of error in the United States would be lower. But Lyon bases this assertion on an essay by Sternberg, Lamb, and their colleagues that discusses the evolution of investigative procedures in Israel.[192] This essay has no bearing on the actual research studies from which Sternberg and her colleagues took their data. The specific Israeli investigators who participated in the research studies were better educated, better trained, and more experienced than any sample that has been studied in the United States of which we are aware.[193] One cannot shrug Lamb's results aside by suggesting that they are a result of the inexperience of the interviewers.

---

186 *E.g.*, Sternberg et al., *Investigative Utterance Types*, *supra* note 174, at 447 tbl.3. The studies speak of "details" but limit the term to those that are abuse-related. *E.g.*, *id.* at 444 ("By definition, details involved the identification of individuals or objects, descriptions of their appearance or actions, and descriptions of relevant events or actions. . . . Details were only counted when they added to understanding of the target incidents and their disclosure."); *see also* E-mail from Michael E. Lamb to Stephen J. Ceci (July 26, 1999) (confirming that suggestive questions "almost always elicit a detail" and that, per the published definition, only "pieces of information that are related to the alleged abuse" are counted as details).

187 Lamb et al., *supra* note 169, at 633 tbl.1.

188 *Id.* at 634 tbl.2.

189 Sternberg and her associates report a mean of 1.95 details per suggestive interviewer utterance. Sternberg et al., *Investigative Utterance Types*, *supra* note 174, at 447 tbl.3. Given the average of 9.5 suggestive interviewer utterances per interview, this means an average of approximately 18.5 details per interview elicited by suggestive questions. In another study, they report 2.37 details per suggestive question for those interviewers who used direct questions to a significant degree near the beginning of the interview, and 2.31 details per suggestive question for those who began the interview with open-ended questions. Sternberg et al., *supra* note 51, at 1140 tbl.3. Lamb and his associates report 1.35 details per suggestive question for those interviews conducted with anatomical dolls and 1.64 details per suggestive question for those interviews conducted without the dolls. Lamb et al., *supra* note 174, at 1256 tbl.2. These figures correspond respectively to about 10.8 and 8.0 details per interview elicited by suggestive questions.

190 *See, e.g.*, Sternberg et al., *Scripted Protocol*, *supra* note 174, at 74 ("Interviewers obtained most of the information from the children using leading and suggestive questions.").

191 Lyon, *New Wave, supra* note 2, at 1035 n.167.

192 *See* Kathleen J. Sternberg et al., *Child Sexual Abuse Investigations in Israel: Evaluating Innovative Practices*, in INTERNATIONAL PERSPECTIVES ON CHILD ABUSE AND CHILDREN'S TESTIMONY 62, 69 (Bette L. Bottoms & Gail S. Goodman eds., 1996) [hereinafter Sternberg et al., *Evaluating Innovative Practices*]; *see also* Kathleen J. Sternberg et al., *Child Sexual Abuse Investigations in Israel*, 23 CRIM. JUST. & BEHAV. 322 (1996). The first paper describes the problems that the Israeli Youth Investigation had faced from its inception in 1955. *See* Sternberg et al., *Evaluating Innovative Practices*, *supra*, at 63-66.

193 *See* Lamb et al., *supra* note 169, at 629. Note also the extensive training that the investigators in the Scripted Protocol study received. *See* Sternberg et al., *Scripted Protocol*, *supra* note 174, at 71-72.

In short, Lamb and his colleagues, working across very different samples in two nations, have consistently documented the extensive use of strongly suggestive techniques by experienced investigative interviewers and the extent to which those techniques elicit statements about alleged abuse from children. Indeed, we believe that any fair-minded reader will recognize that a principal theme of these studies is that, even after extensive training, abuse interviewers tend to secure too much information from leading and suggestive questions.[194] In contrast to the impression given by Lyon, the research of Lamb and his colleagues is a source of concern rather than of solace.

c. *Yuille*

Lyon attempts to dismiss the findings by John Yuille and his colleagues of serious interviewer errors[195] by pointing out that the published account of the study that he cites does not give details of these errors.[196] But a reprint describing the study in detail fills in the gap.[197] Yuille and his colleagues asked "blind" raters to evaluate videotaped interviews of two groups of professional interviewers, one that had been trained in Yuille's own interviewing methods and one that was traditionally trained.[198] The differences between the two groups were so obvious that Yuille's blind raters knew within the first few minutes of viewing each video whether the interviewer had been trained in Yuille's techniques or not.[199] For the traditionally-trained interviewers, there were so many leading questions that it was often impossible to assess the child's credibility.[200] Thus, as with the Lamb and Warren studies, the findings by Yuille and his colleagues show not only that highly suggestive techniques are common, but that they frequently result in serious problems for those who evaluate the interviews.

C. Summary

We have shown that the dominant view for over a century has been that young children are highly suggestible. Studies on which child advocates rely do not disprove this fact. On the contrary, they reveal levels of suggestibility—with regard to abuse-related questions as well as others—that create potential difficulties for an adjudicative system assessing children's allegations of abuse. Moreover, when investigators use strongly suggestive techniques, children's suggestibility is even more

---

194 These researchers do offer hope that a scripted protocol may reduce interviewers' reliance on leading and suggestive questions. *See* Sternberg et al., *Scripted Protocol*, *supra* note 174, at 74. It is important to note that some of the papers discussed above were not available to Lyon when he wrote his article.

195 *See* John C. Yuille et al., *The Nature of Allegations of Child Sexual Abuse*, *in* TRUE AND FALSE ALLEGATIONS OF CHILD SEXUAL ABUSE: ASSESSMENT AND CASE MANAGEMENT 21 (Tara Ney ed., 1995).

196 Lyons, *New Wave*, *supra* note 2, at 1036 & n.176.

197 JOHN C. YUILLE, MINISTRY OF SOCIAL SERVICES, PROVINCE OF BRITISH COLUMBIA, IMPROVING INVESTIGATIONS INTO ALLEGATIONS OF CHILD SEXUAL ABUSE (1995); E-mail from John Yuille to Stephen J. Ceci (July 28, 1999, 13:22:01) (on file with author Ceci) [hereinafter Yuille E-mail].

198 Yuille E-mail, *supra* note 195.

199 *Id.*

200 *Id.*

marked–and research indicates quite clearly that interviewers use strongly suggestive techniques quite often.

Despite the attempt of child advocates to cast researchers such as Ceci as the benighted challengers of children's credibility, there is actually relatively little dispute over how vulnerable children are to various degrees of suggestion. Instead, the disputes tend to concern whether the emphasis should be on the degree to which children are suggestible or the degree to which they are not, how suggestive interviewers are in practice, and the proper attitude of the legal system towards suggestibility. In the next Part, we present a framework for considering the legal implications of the suggestibility research.

II

A Framework for Analysis of Legal Implications

In Part I, we discussed what scientific research shows about how plausibly a child may be led by suggestion to making a false allegation of sexual abuse. In this Part, we present a framework for thinking about the legal implications of this research.

Litigation often concerns the proof of uncertain, hotly-contested events. However satisfactorily the adjudicative system determines the facts, some uncertainty will remain when the time for decision arrives. Thus, as Lyon points out, errors of two types are inherent in litigation.[201] In the context of concern to us here, a *false positive* is the error that arises when abuse did not occur but the system concludes that it did, and a *false negative* is the corresponding error that arises when abuse did occur but the system concludes that it did not. In choosing among legal rules or systems, the principal question is not what the overall ratio of these two types of errors should be, but rather what the tradeoff between the error types should be—that is, how many additional false negatives is it worthwhile to create in order to prevent a false positive?[202]

Lyon regards it as "remarkabl[e]" to conclude that it is better to let 100 guilty people go free than to convict one innocent person.[203] In Section A of this Part, however, we argue that that there is good reason to adhere to the traditional—and constitutionally mandated[204]—principle that in the criminal context false positives are deemed far worse than false negatives. Indeed, a ratio of about 100:1 may well square with deeply-held senses of social value, as well as with constitutional standards articulated by the Supreme Court.[205] Thus, a very high standard of persuasion is appropriate. That is, for a conviction to be warranted, the fact-finder must be satisfied to a very high degree of confidence that the defendant is guilty.

---

201 *See* Lyon, *New Wave, supra* note 2, at 1013.

202 *See id.*; *cf.* Michael L. DeKay, *The Difference Between Blackstone-Like Error Ratios and Probabilistic Standards of Proof*, 21 LAW & SOC. INQUIRY 95, 131 (1996) (pointing out that convicting only on 10:1 odds of guilt will not necessarily yield a 10:1 ratio of false negatives to false positives).

203 Lyon, *New Wave, supra* note 2, at 1075-76.

204 *See* Coffin v. United States, 156 U.S. 432, 453 (1895). The Supreme Court has found this principle in the Due Process Clause. *See In re* Winship, 397 U.S. 358, 364 (1970).

205 *See In re Winship*, 397 U.S. at 364; *Coffin*, 156 U.S. at 453.

Section B of this Part presents a method, using simple Bayesian probability, for assessing the significance of individual items of evidence, in particular a child's allegation of sexual abuse. We show that if there is a substantial probability that the child would *not* make the allegation even though it were true, this *diminishes* the probative value of the evidence. Moreover, if a given item is crucial to the prosecution—as the child's allegation is in many sexual abuse cases—then even a small probability that the child would make the allegation even though it were false has great significance, as does even a small misjudgment in assessing that probability.

Section C of this Part then discusses how the choice of legal rules should take into account the respective possibilities of false positives and of false negatives. Once again, the large difference in the gravity of the two types of error plays a critical role. But, at least arguably, a rule-maker might also let the selection of cases enter into the calculation; if most of those defendants to whom a given rule is to be applied are in fact guilty, then that may be a factor weighing in favor of choosing a pro-prosecution rule. We argue that even if this is so, and even if the large majority of defendants charged with child sexual abuse are in fact guilty, the selection of rules in this area must nevertheless reflect great sensitivity to the creation of false positives.

We caution that, in making the arguments in this Part, we do not mean to suggest that a numerical analysis or any numerical formula should be presented to juries. We do believe, however, that use of Bayesian decision theory for heuristic purposes can sharpen analysis for legal decision makers.

A. Social Utility and the Standard of Persuasion

We begin with a simple model. Suppose that an adjudicator—combining the roles of lawmaker and fact-finder—has to decide whether to treat an accused as innocent or as guilty. After considering all the evidence, the adjudicator assesses the probability of guilt as some number *P* between 0 and 1; thus the probability of innocence, *P*□, is *1 - P*.[206] *O*, or the odds of guilt, is equal to *P/(1 - P)*, a positive number between zero and infinity. How confident should the adjudicator be—that is, how large should the adjudicator assess *O* to be—for her to treat the accused as guilty? It is easy to show that the adjudicator should treat the accused as guilty if and only if

$$O > \frac{Ep}{En}, \quad (1)$$

where *Ep* and *En* are the social costs, respectively, of a false positive and a false negative.[207] We will call this fraction, the standard of

---

206 We speak of the adjudicator's assessment of probabilities and odds because we are applying Bayesian decision theory, in which probabilities or odds represent an actor's subjective level of confidence in the truth of a proposition. In other words, we are not assuming that the probability of a given proportion is objectively determinable.

In accordance with convention, we are using Greek subscripts for matters over which a decision maker has no control, and Latin ones for matters over which a decision maker does have control.

207 *See* Richard D. Friedman, *Answering the Bayesioskeptical Challenge*, 1 INT'L J. EVIDENCE & PROOF 276, 277-78 (1997) (presenting a proof of this statement); *see also*

persuasion expressed in odds, *S*. In other words, if the adjudicator concludes that it is *x* times as bad to convict an innocent person as to acquit a guilty one, then *S* = *x*, and she should convict a defendant only if the odds of guilt are at least *x* to 1.

This approach closely tracks Blackstone's celebrated statement that "it is better that ten guilty persons escape, than that one innocent suffer."[208] Notice, however, that Blackstone's statement does not purport to assess the tradeoff exactly, but only to set a floor for it. Blackstone does not say, for example, that it is better to convict one innocent person than to free eleven guilty people. Blackstone's statement is similar in form to many others made over the centuries, some of which have set much higher floors than did Blackstone.[209] Indeed, in *Schlup v. Delo* the Supreme Court quoted with apparent approval the statement by Thomas Starkie, a leading treatise writer of the early nineteenth century, that "`it is better that ninety-nine . . . offenders should escape, than that one innocent man should be condemned.'"[210]

It is probably impossible to determine any number that satisfactorily expresses the appropriate criminal standard of persuasion, and in any event

---

Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021, 1035-36 (1977) (analyzing standard of persuasion for criminal cases in probablistic terms). The adjudicator assesses the social utility of four possible outcomes: *Ug*, the utility of determining that the defendant is guilty when in fact he is guilty; *Ug*□, the utility of determining that the defendant is guilty when in fact he is innocent; *Ui*, the utility of determining that the defendant is innocent when in fact he is guilty; and *Ui*□, the utility of determining that the defendant is innocent when in fact he is innocent. These assessments of social utility can take any values, provided that *Ug* is greater than *Ug*□ and *Ui*□ is greater than *Ug*□; as a convention we may assume that *Ug* and *Ui*□, representing accurate findings, are both positive, and *Ug*□ and Ui, representing inaccurate findings, are both negative, but this is not necessary. *EUg*, the expected value of a finding that the defendant is guilty, is

(P x Ug) + (1 - P) x Ug□. (N1)

EUi, the expected value of a finding that the defendant is not guilty, is:

(P x Ui) + (1 - P) x Ui□. (N2)

EUg is greater than *EUi*—that is, it is socially preferable that the defendant be found guilty—if and only if Expression N1 is greater than Expression N2. This is true if and only if:

P x (Ug - Ui) + (1 - P) x (Ug□ *- Ui*□*) > 0.* (N3)

And simple algebraic manipulation shows that this is true if and only if:

$$\frac{P}{(1-P)} > \frac{Ui\square - Ug\square}{Ug - Ui} \quad . \quad (N4)$$

The left side of Expression N4 equals the odds of guilt. The numerator on the right side represents the difference in social utility between an accurate finding and an inaccurate one, assuming that the defendant is innocent. In other words, it represents the forgone social utility of an inaccurate finding in this situation, or the cost of a false positive. Similarly, the denominator on the right side represents the cost of a false negative.

208 4 WILLIAM BLACKSTONE, COMMENTARIES *358. Actually, Expression (1) in the text would more closely track a statement that convicting an innocent person is ten times worse than freeing a guilty one. Assuming, however, that the harm caused by erroneous convictions is additive, at least at low numbers of convictions, this latter statement is equivalent to Blackstone's.

209 Perhaps nearly everything everyone has ever said on the subject is recorded in Alexander Volokh, n *Guilty Men*, 146 U. PA. L. REV. 173 (1997). This sardonic piece demonstrates that many people have said many different things on the subject (including some who have objected to the method of analysis altogether, *see id.* at 195-97); that Mr. Volokh is an extremely clever, industrious, and witty smart Alex; and perhaps not much else. He does not provide any reason to doubt that the standard of persuasion in a criminal case must be based on a set of social values under which a false conviction is far worse than a false acquittal.

210 Schlup v. Delo, 513 U.S. 298, 325 (1995) (alteration in original) (quoting THOMAS STARKIE, EVIDENCE 756 (1824)).

we shall not do so here. But we think the Supreme Court is clearly correct in perceiving a "`fundamental value determination of our society that it is far worse to convict an innocent man than to let a guilty man go free.'"[211] This perception has led the Court to conclude that the "beyond a reasonable-doubt" standard is constitutionally required in criminal cases.[212] Whatever that standard means,[213] it "should express our society's view that criminal convictions require, at the least, a high degree of certainty of guilt."[214]

Without attempting to quantify the matter precisely, we will say that it appears to us—in accordance with the Court's statement in *Schlup*—that the Blackstone ratio of 10:1 understates the matter. Indeed, the Starkie ratio of 99:1 appears closer to the mark.[215] We base this assertion largely on the clear differential in social responses between inaccurate failures to convict and inaccurate convictions. We know that most criminals escape conviction. Often they are not apprehended, sometimes their cases are not brought to trial because the evidence seems insufficient, and sometimes they are acquitted at trial. These are unfortunate failures of the system. Reducing them in significant number would be a great social good—though far less good than would have been accomplished by preventing the crime in the first place. In any event, society in fact tolerates these failures as an everyday matter. In contrast, there is widespread revulsion when it becomes apparent that a defendant was convicted and subjected to serious punishment for a crime he did not commit.[216] The injunction of the Hippocratic Oath, "Above all, do no harm," applies to the adjudicative system as well as to the medical profession. Convicting a person for a crime he did not commit—perhaps for a crime that never occurred—is an abhorrent outcome.

The "beyond a reasonable doubt" standard applies across all criminal cases, and it would be unwise—as well as unconstitutional[217]—to select a particular type of crime for a lighter standard. Even if it were appropriate to do so, it does not appear that child sexual abuse would be a particularly strong candidate for such treatment compared to other abominable crimes such as murder, rape, and kidnapping. A false conviction in a child sexual abuse case may have some particularly nasty consequences, including

---

211 *Id.* (quoting *In re* Winship, 397 U.S. 358, 372 (1970) (Harlan, J., concurring)).

212 *In re Winship*, 397 U.S. at 364.

213 Barbara Shapiro gives an historical account of the standard that shows its relationship to developing concepts of probability. *See* BARBARA J. SHAPIRO, "BEYOND REASONABLE DOUBT" AND "PROBABLE CAUSE" at xi-xv (1991).

214 Jon O. Newman, *Beyond "Reasonable Doubt,"* 68 N.Y.U. L. REV. 979, 981 (1993), *cited with approval in Schlup*, 513 U.S. at 325.

215 We do not endorse the standard advocated by Ralph Underwager that "`[i]t is more desirable that a thousand children in abuse situations are not discovered than it is for one innocent person to be convicted wrongly.'" Scott Kraft, *Careers, Reputations Damaged: False Molesting Charges Scar Lives of the Accused*, L.A. TIMES, Feb. 11, 1985, at 1. But note that Underwager weighs false convictions against failures to disclose abuse, not against failures to convict abusers.

216 *See, e.g.*, BARRY SCHECK ET AL., ACTUAL INNOCENCE: FIVE DAYS TO EXECUTION AND OTHER DISPATCHES FROM THE WRONGLY CONVICTED at xvi (2000); David Firestone, *DNA Test Brings Freedom, 16 Years After Conviction*, N.Y. TIMES, June 16, 1999, at A22.

217 *See In re* Winship, 397 U.S. 364, 358 (1970).

destruction of a family[218] and exposure of the defendant to intense public opprobrium and even physical danger. A false acquittal is very unfortunate, but that is true of other terrible crimes as well. The harm caused by a false acquittal, it must be borne in mind, is not the crime itself but failure to punish the crime—which, given the uncertain benefits of punishment, is a significantly different matter. The temptation to apply a lesser standard of persuasion in child sexual abuse cases is probably largely attributable to the pressing need to prevent recurrence of the crime. Of course, recidivism is often a serious possibility with respect to most crimes, and in some circumstances it is not so likely a possibility with respect to child sexual abuse. But in any event, there are often other ways to prevent repeated sexual abuse of a child where the evidence that the suspect has committed abuse is strong but not sufficient to support a conviction. Supervised custody arrangements and restraining orders will often be effective and, for better or worse, the Supreme Court has cleared the path constitutionally for civil commitment where that appears necessary to prevent future abuse.[219]

Two basic points discussed here sometimes seem to get lost in the analysis of errors of omission and of commission by child witnesses.[220] First, false positives are far worse than false negatives. Second, this is not an idiosyncratic value assessment, but one that has deep roots in our adjudicative system and that the Supreme Court has adopted as a matter of constitutional principle.

B. The Probative Value of Evidence

We have seen that the prosecution must satisfy a high standard of persuasion. Now we present a framework for considering the role that the child's allegation of sexual abuse may play in the prosecution's attempt to satisfy that standard. We will show that even small probabilities that a child would make a false allegation of sexual abuse, and small misjudgments in assessing that probability, may be highly significant.

We will use as a touchstone a simple form of Bayes's Theorem.[221] The theorem is a basic principle of logic that indicates how to adjust probability assessments in light of new evidence—which in this context is the child's allegation. This expression combines three basic components. First is the *prior odds* of a proposition–that is, the odds as assessed before receipt of the new evidence. The prior odds take into account all other evidence in the case. They are, in effect, the starting point for analysis of the impact of the allegation. The second component is the *posterior odds* of the proposition—that is, the odds that the proposition is true as assessed

218 *See* Matter of Smith, 509 N.Y.S.2d 962, 963 (N.Y. Fam. Ct. 1986) ("[A] child abuse finding against a parent or parents where no abuse has occurred is as harmful and as devastating *to the subject child* as is the failure to find child abuse where such has occurred." (emphasis added)), *order aff'd*, 513 N.Y.S.2d 483 (N.Y. App. Div. 1987).

219 *See* Kansas v. Hendricks, 521 U.S. 346, 371 (1997).

220 *Cf.* Lyon, *New Wave*, *supra* note 2, at 1018 n.74 (pointing out correctly that the statement by Saywitz and her colleagues calling risk of error of omission greater than that of error of commission "implicitly assumes that both kinds of risk are weighed equally" (referring to Saywitz et al., *supra* note 49, at 690)).

221 This presentation of Bayes's Theorem borrows from a previous one by one of the authors. *See* Richard D. Friedman, *A Presumption of Innocence, Not of Even Odds*, 52 STAN. L. REV. 873, 875 (2000).

after receipt of the new evidence. This is the ending point, the assessment that we wish to make in light of the allegation. The third and final component is the *likelihood ratio* of the new evidence with respect to the proposition at issue. The ratio is a fraction. In the simplest form of the ratio, the numerator is the probability that the evidence would arise if the proposition were true, and the denominator is the probability that the evidence would arise if the proposition were false.[222]

Bayes's Theorem then posits that the posterior odds of the proposition equal the prior odds multiplied by the likelihood ratio.[223] This means that, all other things being equal, the posterior odds will be higher (a) the higher are the prior odds; (b) the higher is the numerator of the likelihood ratio; and (c) the lower is the denominator of the likelihood ratio. Thus, in considering the impact of the new evidence, we must pay attention to both the numerator and the denominator of the likelihood ratio.

1. *The Numerator of the Likelihood Ratio*

The numerator of the likelihood ratio in this context is the probability that the child would make the allegation if it was true. All other things being equal, the higher this probability is, the higher the posterior odds will be that the allegation is true. That is, all other things being equal, the more the making of the allegation is the type of consequence one would expect to follow from abuse, the more probative that allegation is that abuse actually occurred.

Lyon devotes much of his article to showing that various factors inhibit children from making accurate reports of abuse.[224] We do not doubt that this is true. Nor do we deny that in some circumstances these considerations might make it reasonable to use suggestive questioning. But what is significant in the present context is that, all other things being equal, *the less probable it is that the child would report abuse if it occurred, the less probative is a report that she does make*. Thus, Lyon's demonstration *diminishes* the probative value of a child's report of abuse.[225]

Consider simple numerical examples, indicated by the first two hypothetical cases in Table 1. This table shows what the posterior odds

---

222 What Professors Lyon and Koehler have called the "relevance ratio" in child sexual abuse cases—the probability of a particular symptom given abuse divided by the probability of that symptom given nonabuse—is a form of likelihood ratio. Thomas D. Lyon & Jonathan J. Koehler, *The Relevance Ratio: Evaluating the Probative Value of Expert Testimony in Child Sexual Abuse Cases*, 82 CORNELL L. REV. 43, 46, 48-49 (1996).

223 The following is a simple proof of this form of Bayes' Theorem. If *E* is the evidence and *H* is the hypothesis in question, then *P(E & H)* equals *P(E)* x *P(HãE)* and also *P(H)* x *P(EãH)*, where, for example, *P(EãH)* means "the probability of *E* given *H*." Thus, *P(HãE) = P(H)* x *P(EãH)/P(E)*. Similarly, *P(NHãE) = P(NH)* x *P(EãNH)/P(E)*, where *NH* is the negation of *H*. Dividing these two equations by each other yields *P(HãE)/P(NHãE) = P(H)/P(NH)* x *P(EãH)/P(EãNH)*.
The fraction on the left side of this equation is the posterior odds of *H*. The first fraction on the right side is the prior odds of *H*, and the second fraction on the right is the likelihood ratio.

224 *See* Lyon, *New Wave*, *supra* note 2, at 1046-74.

225 Of course, reports of abuse are more probable given suggestive questioning, which is why investigators use suggestive questions. The point still holds, however: even under suggestive questioning, it is by no means certain that a child who has been abused will tell the truth.

would be for a variety of hypothetical sets of prior odds and numerator and denominator of the likelihood ratios. There is nothing particularly significant about these assumed values, which we present for illustrative purposes only. Case 1 shows that if the prior odds of guilt are 1:1 (corresponding to a probability of .5), the denominator of the likelihood ratio is .06, and the numerator is 1 (meaning that it is certain that if the abuse occurred the child would report it), then the posterior odds would be about 17:1. That would arguably (under a Blackstone-like standard) be enough to justify conviction. As Case 2 shows, however, if the numerator is divided in half, to become .5—meaning that the child is equally likely as not to report abuse—the posterior odds are also divided in half, and at about 8:1 are presumably not enough to warrant conviction.

2. *The Denominator of the Likelihood Ratio*

As just indicated, differences in the numerator of the likelihood ratio have proportional effects on the posterior odds. The inverse is true for the denominator: all other things being equal, doubling the denominator will halve the posterior odds. But the denominator—the probability that the child would make the allegation even though it is not true—is presumably much smaller than the numerator; otherwise the evidence does not have strong probative value. Thus, even a relatively slight probability that the child would report abuse even though it did not occur may be highly significant.

The subsequent rows of Table 1 illustrate this fact. Suppose at first that there is no evidence against the defendant other than the child's allegation. In this setting, at least two, and arguably all three, of the assumptions underlying Case 3 are highly favorable to the prosecution. The first assumption is that the prior odds that the defendant is guilty—assessed without the allegation—are 1 in 100. This seems unrealistically high, given that there is no evidence of his guilt apart from the allegation; the vast majority of people are not guilty of sexual abuse on a given occasion with a particular child, even a family member.[226] The second pro-prosecution assumption is that the numerator of the likelihood ratio is 1; as we have seen, it is almost certainly substantially less than that because many children fail to report abuse when it occurs. The third assumption is that the denominator is only .01, meaning that there is one chance in a hundred that if abuse did not occur the child would nevertheless allege it. Even if we rely on the data gathered by Goodman and her colleagues,[227] this estimate seems quite low. Now notice that even given all these assumptions, the posterior odds are only 1:1, or even odds. Plainly, they are not enough for conviction. To reach posterior odds in the Starkie range[228] given the first and second assumptions, Case 4 shows that the denominator of the likelihood ratio would have to be microscopically small. In other words, Cases 3 and 4 show that, if the child's statement is the only substantial evidence indicating that the

---

226 *See* Friedman, *supra* note 219, at 879-83 (discussing the presumption of innocence in terms of low prior odds of guilt).

227 *See, e.g.*, Rudy & Goodman, *supra* note 69, at 532; Saywitz et al., *supra* note 49, at 687.

228 *See* STARKIE, *supra* note 208, at 751.

defendant is guilty, the fact-finder could justifiably find him guilty only if it concluded that there was an almost infinitesimally small probability that the child would make an allegation even though it was false. Information suggesting that this probability, though small, is not *that* small might therefore be of great assistance to the fact-finder in avoiding a false conviction.

TABLE 1

Case No.	*Prior odds (assessed apart from the allegation) of abuse by defendant*	*Numerator of the likelihood ratio (probability of allegation given abuse by defendant)*	*Denominator of the likelihood ratio (probability of allegation given no abuse by defendant)*	*Posterior odds (taking allegation into account) of abuse by defendant (col. 1 x col. 2/col. 3)*
1	1:1	1	.06	16.67:1
2	1:1	.5	.06	8.33:1
3	1:100	1	.01	1:1
4	1:100	1	.0001	100:1
5	1:2	2/3	.01	33.33:1
6	1:2	2/3	2/36	6:1
7	3:1	2/3	.01	200:1
8	3:1	2/3	2/36	36:1
9	6:1	2/3	.01	400:1
10	6:1	2/3	2/36	72:1
11	6:1	2/3	.25	16:1

Case 3 illustrates another point as well. This case posits a highly reliable witness; she is certain to allege abuse if it happened, and the probability is only 1 in 100 that she would allege abuse if it did not happen. And yet the odds that she was abused, given her statement that she was, are just 1:1. Why are these odds so low? The answer is that the prior odds were so low. In other words, if the prior odds are low enough, then even with a reliable witness one might conclude that witness error is at least as likely as actual abuse to lead to a report of abuse. Looking at the problem more globally, suppose that many children, most of whom have not been abused but a few of whom have, are asked suggestive questions about abuse. It may be that, even if the group of nonabused children taken as a whole is reliable, a substantial number of those who have not been abused will falsely say they have been.

The remaining cases in Table 1 assume much higher prior odds. In other words, these cases assume that there is substantial evidence against the defendant apart from the child's allegation. For the sake of simplicity, all the cases also assume that the numerator of the likelihood ratio is 2/3. This is very close to the proportion of children in Saywitz's study (25/36) who, having received an anal touch in a pediatric examination, accurately

reported this in response to directed questioning.[229] We hypothesize a set of cases for each of three prior odds.

In Cases 5 and 6, the prior odds are 1:2, corresponding to a prior probability of .333. In other words, apart from the child's allegation it appears plausible, but substantially more likely not true, that the defendant committed the abuse charged. This could be the case, for example, if there is strong physical evidence that the child has been abused and, based on the defendant's relationship with her, he is one of a few likely candidates to have been the perpetrator. Or a fact-finder could assess those prior odds if the evidence of abuse is weaker but it is virtually certain that if the child was abused the defendant was the perpetrator. In this setting, a denominator of .01, as in Case 5, would mean that the prosecution's proof rather easily satisfies the Blackstone standard, though it is nowhere near the Starkie standard. Case 6 assumes a denominator of 2/36. This is the proportion of children in the Saywitz study who did not receive an anal touch but who, in response to directed questioning, inaccurately answered that they had.[230] With this denominator, the prosecution does not even come close to the Blackstone standard. Note the significance of this case: There is substantial evidence apart from the child's allegation that the defendant is guilty, the likelihood ratio is drawn from the Saywitz study, and yet the prosecution clearly does not satisfy its burden of persuasion.

Cases 7 and 8 use the same denominators as Cases 5 and 6, respectively, but with higher prior odds of 3:1, corresponding to a prior probability of .75. A case like this could arise if there is both relatively strong evidence that the child was abused and a strong indication that the defendant was the abuser. The denominator of .01 now means that, with the child's allegation, the prosecution's evidence would satisfy even the Starkie standard. With a denominator of 2/36, on the other hand, the case would satisfy the Blackstone standard but not the Starkie standard.

Cases 9 through 11 assume prior odds of 6:1, or a prior probability of 85.7%. Thus, apart from the child's allegation, the evidence of abuse is very strong, but not quite strong enough to justify conviction under the Blackstone standard. Here, a denominator of .01 means that, taking the child's allegation into account, the case would easily satisfy the Starkie standard. A denominator of 2/36 means that the case would not quite satisfy that standard. However, even with a relatively high denominator of .25, the case would satisfy the Blackstone standard by a comfortable margin.

Again, it is important to remember that these are merely hypothetical cases. Nonetheless, they vividly illustrate important points. Given the high standard of persuasion applicable in a criminal case, even a very low denominator—a very low probability that the child would make the allegation if it was false—may be high enough to defeat the prosecution. And rather small differences in the denominator may, depending on the circumstances, have great significance.

A corollary is this: If the fact-finder makes a bad judgment of the denominator, that may significantly increase the chance of an inaccurate

229 *See* Saywitz et al., *supra* note 49, at 687.
230 *See id.*

factual finding. Consequently, the question becomes how to design a fact-determination system that minimizes the costs of these errors.

C. Choice of a Legal Regime

As the previous Section demonstrated, small misjudgments by the fact-finder may lead to significant errors in applying the standard of persuasion. Now suppose that a rule maker is choosing between two rules that differ in at least one respect—evidentiary consequences, for example—and so may lead the fact-finder to different results. We need only be concerned with the situation in which one rule leads to fewer false positives than the other and the second to fewer false negatives, for otherwise the choice between them is immediately obvious, a "no brainer." Call the rule that yields fewer false positives, and so is more favorable to the defendant, Rule 1 and the other one, Rule 2. Then it becomes rather easy to show that Rule 2 is preferable to Rule 1 only if the number of false negatives avoided by shifting from Rule 1 to Rule 2 divided by the number of false positives created by that shift is greater than *S*, the standard of persuasion. Also, recall that *S* is equal to *Ep/En*, the ratio of social cost of a false positive to the social cost of a false negative—which, we have seen, our legal system treats as very high.[231] Thus, switching from Rule 1 to Rule 2 is a good choice only if the number of false negatives eliminated is far greater than the number of additional false positives created.

This proposition follows from the same assessment of social value used in setting the standard of persuasion itself. There is one difference in application, however. In considering a particular case, a fact-finder must

---

231 The expected value of choosing Rule 1 is

$[P_1(g\&) \times Ug] + [P_1(i\&) \times Ui] + [P_1(g\&\square) \times Ug\square] + [P_1(i\&\square) \times Ui\square]$, (N5)

where the *U* terms are defined as above, *supra* note 205, $P_1(g\&)$ is the probability that the defendant is guilty and will be found guilty under Rule 1, $P_1(i\&)$ is the probability that the defendant is guilty but will be found innocent under Rule 1, and so forth. The four *P* terms total 1.

Similarly, the expected value of choosing Rule 2 is

$[P_2(g\&) \times Ug] + [P_2(i\&) \times Ui] + [P_2(g\&\square) \times Ug\square] + [P_2(i\&\square) \times Ui\square]$. (N6)

Again, the four P terms (which are different from those in Expression N5) add to 1. Moreover, we will assume that the proportion of defendants who are actually guilty is the same under Rule 1 as under Rule 2. That is,

$P_1(g\&) + P_1(i\&) = P_2(g\&) + P_2(i\&)$, (N7)

and

$P_1(g\&\square) + P_1(i\&\square) = P_2(g\&\square) + P_2(i\&\square)$, (N8)

which means that

$P_2(g\&) - P_1(g\&) = P_1(i\&) - P_2(i\&)$, (N9)

and

$P_1(g\&\square) - P_2(g\&\square) = P_2(i\&\square) - P_1(i\&\square)$. (N10)

Rule 2 is preferable to Rule 1 only if Expression N6 is greater than Expression N5. Subtracting Expression N5 from N6, and rearranging in light of Equations N9 and N10, means that this criterion is satisfied if and only if

$[Ug - Ui] \times [P_1(i\&) - P_2(i\&)] + [Ui\square - Ug\square] \times [P_1(g\&\square) - P_2(g\&\square)] > 0$. (N11)

And this in turn is true if and only if

$$\frac{[P_1(i\&) - P_2(i\&)]}{[P_2(g\&\square) - P_1(g\&\square)]} > \frac{[Ui\square - Ug\square]}{[Ug - Ui]}. \quad \text{(N12)}$$

The numerator of the fraction on the left side of Expression N12 represents the false negatives prevented by the switch from Rule 1 to Rule 2; the denominator represents the false positives created by that switch. Notice that the fraction on the right side of Expression N12 is the same as that on the right side of Expression N4, *supra* note 205. This is *Ep/En*, the social cost, respectively, of a false positive and of a false negative, which we have said equals *S*, the standard of persuasion expressed in odds.

apply a presumption of innocence, which means in effect that it must begin the case, before receipt of evidence, with a very low assessment of the probability that the defendant is guilty.[232] In choosing between rules, however, a legal rule maker is at least arguably justified in taking into account the nature of the cases that enter the adjudicative system. If the defendant is guilty in virtually all of the cases entering the system, then the number of false positives, even under Rule 2, may be so low that the increase in false positives between Rule 1 and Rule 2 may not be of great concern. This line of reasoning may be controversial as a guide to a choice of rules, but it seems reasonable, and we assume it is true for the sake of argument because it cuts against the main contentions in this Article. It is, indeed, a viewpoint that Lyon adopts.[233]

This consideration may be expressed in a relatively simple expression. Rule 2 is preferable to Rule 1 if and only if

$$R \times O > S\,, \quad (2)$$

where O is the odds that the defendant is guilty, assessed without taking into account the information yielded by Rule 1 or 2, S is the standard of persuasion, expressed in odds as before, and R is a fraction defined in the following way: Its numerator is the excess of the probability, given that a defendant is in fact guilty, that he will be found innocent under Rule 1 over the comparable probability under Rule 2. And the denominator of R is the excess of the probability, given that a defendant is in fact innocent, that he will be found guilty under Rule 2 over the comparable probability under Rule 1.[234]

Suppose, for example, that, as compared to Rule 1, Rule 2 raises the probability that a defendant will be found guilty if in fact he is guilty by 30%, and the probability that he will be found guilty if in fact he is innocent by only 10%. Thus, R equals 3:1. If 99% of the defendants who enter the system are in fact guilty, so that O is 99:1, then Rule 2 would eliminate nearly 300 false acquittals for every extra false conviction created. This satisfies even the Starkie test. But if 90% of the defendants who enter the system are guilty, so that O is 9:1, then Rule 2 would eliminate only twenty-seven false acquittals for every false conviction created. And if 70% of the defendants who enter the system are guilty, so that O is 7:3, then Rule 2 saves only seven false acquittals for every false conviction created, which fails the Blackstone test.

To put it another way, for the pro-prosecution rule to be preferable, R must be greater than S/O. Of course, it is difficult to know the odds of

---

232 *See* Friedman, *supra* note 219, at 879-81.

233 *See* Lyon, *New Wave*, *supra* note 2, at 1080 ("As the number of truly abused children among those that researchers interview grows, the percentage of false allegations decreases.").

234 Begin by rewriting $P_1(i\&)$, a term used in Expression N12, *supra* note 229, as $P() \times P_1(i\_)$. In other words, the probability under Rule 1 that the defendant is guilty in fact *and* yet will be found innocent is the probability that he is guilty times the probability under Rule 1 that, *if* guilty he will be found innocent. Similarly, $P_2(i\&) = P() \times P_2(i\_)$, $P_1(g\&\square) = P(\square) \times P_1(g\_\square)$, and $P_2(g\&\square) = P(\square) \times P_2(g\_\square)$. Then, bearing in mind that the right side of Expression N12 equals S, that expression may be rewritten as

$$\frac{P()}{P(\square)} \times \frac{P_1(i\_) - P_2(i\_)}{P_2(g\_\square) - P_1(g\_\square)} > S. \quad (N13)$$

The first fraction on the left side is O, the odds of guilt. The second fraction is R.

guilt with respect to any class of cases. Nonetheless, highly cited studies by child advocates suggest that *O* is no greater than about 3:1 across a broad range of cases.[235] We suspect that *O* is substantially lower in the class of cases of principal interest in this Article, in which the child makes the allegation of sexual abuse only after being prompted by suggestive questioning. In any event, to be conservative, we will use the 3:1 figure for *O*. If the Blackstone standard of 10:1 governs for *S*, that would mean that *R* would have to be greater than 3.33:1 for the pro-prosecution rule to be preferable. Thus, if the pro-prosecution rule increased by 10% the probability that if a defendant was innocent he would be found guilty, that rule could be justified only if it increased by at least 33.33% the probability that if a defendant was guilty he would be found guilty. And if instead we apply the Starkie standard of 99:1, then *R* must be greater than 33:1 for the pro-prosecution rule to be preferable. This indicates the need for extreme caution before adopting a rule that increases the possibility of false convictions.

We caution again that there is nothing magical about these numerical examples. They are presented here solely for heuristic purposes. But they do emphasize the fallacy of assuming, when choosing between legal rules, that all errors are of equal significance.

III

Legal Implications

This Part now turns to the specifics of how the legal system should respond to the concerns we have raised.

A. Suggestive Interview Techniques

As Part I of this Article has shown, scientific research demonstrates that suggestive questioning, including techniques such as coaching, bribes, and threats, increases the probability that the child will make an allegation of abuse regardless of whether it actually occurred. If in the end the child

---

235 Nancy Thoennes and Patricia Tjaden, in their study of 9000 families in custody disputes, estimated that 33% of the allegations of child sexual abuse were false. *See* Nancy Thoennes & Patricia G. Tjaden, *The Extent, Nature, and Validity of Sexual Abuse Allegations in Custody/Visitation Disputes*, 14 CHILD ABUSE & NEGLECT 151, 151, 153 (1990). In a study of 136 cases involving divorce, Faller categorized eighty-nine (65.4%) as involving apparently true allegations of abuse and thirty-one (22.8%) as involving false or possibly false allegations; as to the remaining sixteen cases (11.8%) she said only that "other dynamics were at work." Kathleen Coulbourn Faller et al., *Research on False Allegations of Sexual Abuse in Divorce*, 6 APSAC ADVISOR (Fall 1993), at 1, 8-9; *see* Kathleen Coulbourn Faller, *Possible Explanations for Child Sexual Abuse Allegations in Divorce*, 61 AM. J. ORTHOPSYCHIATRY 86, 88 (1991). Jones and McGraw reviewed the disposition of 576 reported cases of suspected sexual abuse in Denver during a single year involving children up to age fifteen. David P.H. Jones & J. Melbourne McGraw, *Reliable and Fictitious Accounts of Sexual Abuse in Children*, 2 J. INTERPERSONAL VIOLENCE 27 (1987). They concluded that 6% were false claims, and they estimated that in an additional 17% of cases, an adult, without actually alleging abuse, reported a suspicion of abuse that turned out to be false, and another 24% were indeterminate. *Id.*; *see also* Mark D. Everson & Barbara W. Boat, *False Allegations of Sexual Abuse by Children and Adolescents*, 28 J. AM. ACAD. ADOLESCENT & CHILD PSYCHIATRY 230, 231 (1989) (using analyses of case workers' reports to estimate that 8% of reports of sexual abuse by adolescents, but only 2% of reports by children under six, are false).

would make an allegation, then for two reasons it is preferable that this occur without suggestive questioning. First, an unprompted allegation is more powerful, persuasive evidence than a prompted allegation and therefore more likely to lead to a conviction if the defendant is in fact guilty.[236] For this reason, the self-interest of the investigative and prosecutorial authorities should lead them to avoid suggestive questions when possible. Second, if the child does make an unprompted allegation, it is unlikely to result in an inaccurate conviction, because in most circumstances children are very unlikely to make a false allegation without suggestive questioning.

It is preferable, therefore, to avoid suggestive questioning until the child has told all that she is likely to tell without suggestion. But for at least two reasons we do not believe that investigators should avoid suggestive questioning altogether. First, the information that they gain through suggestive questioning may be useful for purposes other than criminal prosecution—for example, the determination of custody arrangements or the appropriateness of a restraining order. Because the governing standard of persuasion is lower in these settings than in criminal prosecutions, information obtained by suggestion is more likely to be decisive than in a criminal setting. Second, even in criminal prosecutions, an allegation procured by suggestive questioning may, depending particularly on the strength of the rest of the case, be decisive in carrying the prosecution's burden of persuasion.

We recommend, therefore, that investigators avoid suggestive questions until they are confident that the child has told all she is likely to tell without prompting. Interviewers should attempt to limit repetition of closed (i.e., yes/no) questions within the interview, and investigative authorities should, to the extent feasible, avoid multiple interviews with multiple interviewers. Furthermore, interviewers should adopt categorical rules against the use of techniques that have been demonstrated to create particularly significant risks that a child will make a false allegation. Thus, interviewers should not offer rewards or other positive reinforcement for favored answers, threaten punishment or create negative reinforcement for disfavored ones, vilify the accused, or (unless the child has raised the matter first) refer to statements by the child's peers. Though suggestive questions are sometimes useful, the use of these techniques is always improper.

---

236 If the interviewer asks only open-ended questions that contain no suggestion of the proposition at issue, and the child nevertheless asserts that proposition, a reasonable juror would probably assess the likelihood ratio of the assertion with respect to that proposition as extremely high. It is at least plausible that the child would make the assertion if it were indeed true, and it is highly unlikely, given the absence of suggestion, that she would make *this particular* assertion if it were false. *Cf.* Richard D. Friedman, *Route Analysis of Credibility and Hearsay*, 96 YALE L.J. 667, 683-84, 736-39 (1987) (analyzing the "lottery" problem). If, however, the interviewer merely asserts a proposition and asks the child to confirm or deny it, then the juror would assign a significantly lower likelihood ratio. That the question is leading presumably raises the numerator of the likelihood ratio substantially, because the question focuses the child's attention on the proposition, but this factor also tends to increase the denominator vastly: the denominator is the probability that the child would assert *this particular* proposition even though it was false, and because the question focuses the child's attention on the proposition that probability is presumably far greater than if the question were open-ended.

There is nothing particularly novel about these recommendations. Although some interviewers may ignore them in practice, they are essentially textbook principles, much elaborated in manuals for interviewers—including one by the National Center for the Prosecution of Child Abuse, in cooperation with the National District Attorney's Association and the American Prosecutor's Research Institute.[237] Interestingly, for all that Lyon and other child advocates contend that suggestive questioning is often necessary to prompt an accurate statement and that (nevertheless) troublesome questioning does not often occur in real practice, they do not argue anything different. They do not, for example, argue that investigators should feel free to ask suggestive questions without restraint.[238]

B. Witness Taint and Competence

In *State v. Michaels*,[239] the New Jersey Supreme Court held that if the defendant presents "`some evidence' that the [child's] statements were the product of suggestive or coercive interview techniques," then the prosecution must demonstrate by clear and convincing evidence at a pretrial "taint hearing" that, "considering the totality of the circumstances surrounding the interviews, the statements or testimony [of the child] retain a degree of reliability sufficient to outweigh the effects of the improper interview techniques."[240] If the prosecution fails to satisfy this burden, then the court must exclude the child's testimony, as well as her prior statements alleging abuse.[241]

Some courts outside New Jersey have occasionally followed *Michaels* in requiring taint hearings,[242] but more commonly courts simply consider these issues in determining the competency of the child to give testimony.[243] For our purposes, the difference is not particularly

---

237 *See* State v. Michaels, 642 A.2d 1372, 1378 (N.J. 1994); *see also, e.g.*, DEBRA A. POOLE & MICHAEL E. LAMB, INVESTIGATIVE INTERVIEWS OF CHILDREN: A GUIDE FOR HELPING PROFESSIONALS (1998) (reviewing interview protocol); *cf.* Commonwealth v. LeFave, 714 N.E.2d 805, 811 (Mass. 1999) (summarizing trial testimony of Dr. Daniel Schuman on questioning techniques to avoid); Gail S. Goodman & Vicki S. Helgeson, *Child Sexual Assault: Children's Memory and the Law*, 40 U. MIAMI L. REV. 181, 189 (1985) ("Mild suggestion, such as `Did Uncle Harry touch your penis?,' would be less likely to lead to an inaccurate report than a strong suggestion, such as `I bet Uncle Harry touched your penis, isn't that right?'").

238 *See* MYERS, *supra* note 49, § 1.26, at 62 ("The ultimate goal in forensic interviewing is to reduce the number of suggestive and mildly leading questions while, at the same time, respecting the need to ask such questions."); *id.* § 1.30, at 69 (noting under heading, "Interview Practices to Be Avoided" that "[i]t is difficult to envision a case in which it would be proper to ask highly leading questions like `He touched your vagina, isn't that right?'").

239 642 A.2d 1372 (N.J. 1994).

240 *Id.* at 1383 (citations omitted).

241 *See id.* at 1383-85.

242 *See, e.g.*, Commonwealth v. Callahan, 9 Mass. L. Rptr. 228, 1998 WL 808850, at *1 (Mass. Super. Oct. 19, 1998) (describing prior decision of court); State v. Carol M.D., 948 P.2d 837, 845 (Wash. Ct. App. 1997).

243 *See, e.g.*, State v. Leak, No. 16424, 1998 WL 184646, at *2 (Ohio Ct. App. 2d Dist. Mar. 27, 1998) (declining to apply the *Michaels* rule but addressing the same issue through determination of competency); Fischbach v. State, 1996 WL 145968, at *2 (Del. Mar. 15, 1996) (declining to adopt the formal *Michaels* procedures, under which, once the defendant presents sufficient evidence of reliability, the prosecution bears the burden of proving reliability by clear and convincing evidence, but holding that "if a witness's

significant. Either way, the bottom-line issue is whether the court should preclude the child from giving live testimony about the abuse because she has been subjected to a substantial degree of suggestion.

Although Ceci coauthored the amicus brief that some have credited with persuading the *Michaels* court, we agree in general with Lyon and Myers that children's suggestibility should not usually prevent them from being heard as witnesses, even if the circumstances indicate that the child was subjected to strong forms of suggestion.[244] We have two basic reasons for reaching this conclusion.

First, a child's statement alleging abuse has significant value in proving that abuse. Nothing we have said indicates the contrary. Our argument supports the proposition that the suggestibility of the child *may* account for her allegation of abuse in some circumstances. The allegation itself is not conclusive evidence that abuse occurred. But the allegation may yet be important, even decisive, evidence, at least when there is other evidence supporting it. In Bayesian terms, we have argued that the denominator of the likelihood ratio in some settings is not infinitesimal, and therefore the ratio itself is not enormously high. But we have not argued that the likelihood ratio is no greater than one. Plainly, it is often much greater, even in the face of significant suggestion.

Second, we believe that the dignity of the child is fostered by allowing her to tell her story first-hand in the proceeding that will resolve the truth of her allegation.

Against these considerations, three basic arguments may be made for excluding the testimony of the child. We will call these the reliability argument, the best evidence argument, and the wrongful conduct argument.

According to the *reliability argument*, on which *Michaels* principally depended, if the child has been subjected to significant suggestion, her testimony may be so unreliable that it should be rejected. We certainly agree that often the child's testimony may not be reliable in the sense of being virtually conclusive. Indeed, in some circumstances, the testimony may not even be reliable in the weaker sense that the denominator of the likelihood ratio—the probability that the child would testify as she has even though the testimony is false—is very small. But notwithstanding some judicial statements to the contrary,[245] reliability in neither sense is, or should be, the general standard for the admissibility of live testimony.

---

statement is obtained by use of impermissible interviewing techniques [] the trial court must determine whether the statement is reliable after considering the totality of the circumstances"); English v. State, 982 P.2d 139, 146 (Wyo. 1999) (expressing agreement with "the reasoning" of *Michaels* but adopting the holding of the Washington State Supreme Court in *In re Dependency of A.E.P.*, 956 P.2d 297 (Wash. 1998), that competence inquiry includes the question of pretrial taint; discussing varying standards for determining whether to hold a competence hearing); *cf.* 18 U.S.C. § 3509(c)(4) (1994) (requiring a showing of "compelling reasons" before a federal court can hold a hearing on competence of child witness).

244 Myers, *supra* note 88, at 944-45.

245 *See, e.g.*, *Michaels*, 642 A.2d at 1380 ("`[R]eliability [is] the linchpin in determining admissibility' of evidence under a standard of fairness that is required by the Due Process Clause of the Fourteenth Amendment." (alteration in original) (quoting Manson v. Braithwaite, 432 U.S. 98, 114 (1977))). *Manson* itself dealt with identification testimony, which we address *infra* note 253 and accompanying text.

Rather, the governing principle is that, at least within broad bounds, the credibility of witnesses is for the jury to determine.

In an earlier age, courts excluded the testimony of many potential witnesses, including the parties themselves, on the ground that bias or some other factor would make their testimony unreliable.[246] The modern, vastly preferable view recognizes that such an exclusionary approach has huge costs in loss of valuable information. Cross-examination, impeachment, rebuttal, and recognition by the fact-finder of defects of the testimony—sometimes with the assistance of expert testimony—are the mechanisms that we hope will prevent the testimony from leading the fact-finder astray. Testimony of the parties is extremely unreliable, if for no reason other than self-interest, but it is universally allowed today. Indeed, a criminal defendant has a constitutional right to present his own testimony, even, in at least some circumstances, if it has been tainted by suggestion.[247] In general, witnesses who claim firsthand knowledge do not have to pass through a reliability screen, even when testifying against a criminal defendant.[248] Witnesses with a grudge against the defendant, witnesses whose perception of the events at issue may have been impeded by stress, bad lighting, or weak eyesight, witnesses with faulty memory, and witnesses who have been offered some inducement (such as a reduction of sentence) to testify—all these are allowed to testify about what they assert they perceived, without the court first determining that their evidence is reliable. Courts should not hold the testimony of children to a more stringent standard.

A reliability standard for the admissibility of testimony misconceives the basic theory of evidence. To warrant admissibility, an individual item of evidence does not have to point reliably in the direction the proponent claims. "A brick is not a wall," and every witness need not hit a home run, in the classic aphorisms.[249] That is, a single piece of evidence, including the testimony of a witness, does not have to support the prosecution's entire case but need only provide one of the building blocks for the case. Prosecution evidence, not reliable in itself because there is a substantial probability that it would arise even if the defendant were innocent, may in conjunction with other evidence make an overwhelming case.

The better standard is whether the prejudicial potential of the evidence outweighs the probative value.[250] It must be constantly borne in mind that the child's testimony that abuse occurred does have substantial probative value. Even if the child was subjected to strong forms of suggestion, the child is significantly more likely to testify to a given proposition if that proposition is true than if it is false, and no research suggests otherwise. In some cases, that probative value may be decisive.

---

246 1 MCCORMICK ON EVIDENCE 266-80 (John W. Strong gen. ed., 5th ed. 1998).

247 *See* Rock v. Arkansas, 483 U.S. 44, 61 (1987).

248 A few decisions have set up such screens in the related context of adult witnesses who claim to have recovered long-repressed memories of abuse or violence. *See, e.g.*, State v. Walters, 698 A.2d 1244, 1246-47 (N.H. 1997); State v. Hungerford, 697 A.2d 916, 920 (N.H. 1997).

249 FED. R. EVID. 401 advisory committee's note (quoting MCCORMICK ON EVIDENCE § 152 (1st ed. 1954) and citing Judson F. Falknor, *Extrinsic Policies Affecting Admissibility*, 10 RUTGERS L. REV. 574, 576 (1956)).

250 *See* FED. R. EVID. 403 (allowing exclusion where various factors, including prejudice, substantially outweigh probative value).

What then of prejudice? The principal prejudice concern is that the jury will overvalue the testimony by so much that the truth-determination process is benefitted by exclusion. But to our knowledge, the scientific research provides no indication that juries are likely to overvalue the testimony of a child to this degree. It may well be that, especially absent explanation of the research on suggestibility, a jury would tend to underestimate the probability that the child would make the allegation if it was false (the denominator of the likelihood ratio). Such an error would tend to cause the jury to overassess the probative value of the testimony. It is much more doubtful, however, that the jury would overassess the probative value to such an extent that admission of the evidence is worse for the truth-determining process than denying the jury access to this information. After all, jurors are capable of understanding the problem of suggestibility and taking it into account in assessing the testimony, and experimental evidence suggests that they do so.[251] Excluding the evidence, which has some probative value, guarantees that the jury will underassess it. Those who argue for this result, notwithstanding the usual rule that credibility is for the jury, should have the burden of demonstrating that the uncertain prospect of jury overassessment is significant enough to warrant exclusion.[252]

Moreover, treating a witness as incompetent is a blunderbuss, which should be used only with great caution. We believe that other methods can usually limit the danger of juror overestimation without relying on this weapon. Two of these methods are discussed below. One is expert explanation of suggestibility to educate the fact-finder as to the vulnerability of the evidence. The other, for extreme cases only, is judicial refusal to enter judgment of guilt if the child's allegation provides the only substantial evidence pointing to guilt and the court concludes that there clearly is a significant danger that the allegation was the product of strong suggestion.

We acknowledge that in some contexts, such as coerced confessions and identifications made after official suggestions, courts have spoken of unreliability of testimony as a factor warranting exclusion.[253] We think, however, the argument is generally misplaced, and that, to the extent exclusion is appropriate in those contexts, it is better justified by the two other arguments discussed below.

The *best evidence argument* does not rely on the proposition that the evidence is more prejudicial than probative. Rather, it is based on the

---

251 *Cf.* Lyon, *New Wave*, *supra* note 2, at 1083 n.419 (reviewing public attitudes regarding the suggestibility of children).

252 It is by no means obvious how to compare prejudice from overevaluation against probative value. One possibility is to use the ratio of (a) incorrect verdicts of not guilty prevented by admitting, rather than excluding, the evidence to (b) incorrect verdicts of guilty caused by that evidentiary decision. Under this approach, if the ratio is greater than *S*, the standard of persuasion discussed in Part II, the probative value is greater than the prejudice from overevaluation. We know of no research currently offering significant assistance in making this assessment.

253 *See* Jackson v. Denno, 378 U.S. 368, 385-86 (1964) (holding that the use of involuntary confessions is unconstitutional "not only because of the probable unreliability of confessions that are obtained in a manner deemed coercive," but also because of the sacrifice of human values created by such coercion and the dangers created by police violation of the law); *infra* note 253 (discussing identifications made after suggestion).

"best evidence" principle, the proposition that exclusion of proffered evidence is warranted in some settings because it may induce the creation of better evidence.[254] To the extent that interviewers—whether private individuals or government agents associated with the prosecution—regularly conduct interviews of children with the anticipation that prosecutors will use them in abuse cases, the threat of exclusion of the child's testimony for undue suggestiveness may inhibit them from being so suggestive. We believe that this factor, rather than concerns about trustworthiness, underlies the doctrine—invoked often but rarely with success—that in-court eyewitness identification testimony may be so tainted by prior suggestiveness as to be constitutionally inadmissible.[255]

This consideration plays a significant role in the realm of child witnesses. Nevertheless, given the affirmative considerations weighing in favor of admissibility, we do not believe it usually suffices to justify exclusion of the child's testimony.

For one thing, many professional interviewers, even those inclined to assist the prosecution, may already have considerable incentives not to conduct interviews in an unduly suggestive manner. Strong suggestiveness, as we have pointed out, is in some circumstances counterproductive in that it reduces, rather than increases, the useful information yielded by the questioning. It also makes the child's

---

254 *See generally* Dale A. Nance, *The Best Evidence Principle*, 73 IOWA L. REV. 227 (1988) (advocating a central rule for the principle in evidentiary law).

255 In *Gilbert v. California*, 388 U.S. 263 (1967), the Court held that, in the circumstances of that case, "[t]he admission of the in-court identifications without first determining that they were not tainted by the illegal lineup but were of independent origin was constitutional error." *Id.* at 272 (citation omitted). *Gilbert* was one of a trilogy of cases on suggestive identification. The same day, in *Stovall v. Denno*, 388 U.S. 293 (1967), the Court held that the rules of *Gilbert* and of *United States v. Wade*, 388 U.S. 218 (1967), which recognized a defendant's constitutional right to have counsel present at post-indictment lineup, *id.* at 236-37, did not apply retroactively. *Stovall*, 388 U.S. at 291. The *Stovall* Court said:

> *Wade* and *Gilbert* fashion exclusionary rules to deter law enforcement authorities from exhibiting an accused to witnesses before trial for identification purposes without notice to and in the absence of counsel. A conviction which rests on a mistaken identification is a gross miscarriage of justice. The *Wade* and *Gilbert* rules are aimed at minimizing that possibility by preventing the unfairness at the pretrial confrontation that experience has proved can occur and assuring meaningful examination of the identification witness' testimony at trial.

Id. at 297. Later, in cutting back on *Gilbert*, the Court tended to place greater weight on the question of whether the identification was reliable notwithstanding the suggestiveness. *See* Neil v. Biggers, 409 U.S. 188, 196-97 (1972); Simmons v. United States, 390 U.S. 377, 384 (1968) (requiring suppression of identification testimony only if pretrial procedure was "so impermissibly suggestive as to give rise to a very substantial likelihood of irreparable misidentification"). In *Manson v. Brathwaite*, 432 U.S. 98 (1977), the Court said that "reliability is the linchpin in determining the admissibility of identification testimony." *Id.* at 114.

At the same time, however, the *Manson* Court undercut the significance of untrustworthiness. Assuming the evidence did not fail the *Simmons* standard, it was for the jury to weigh. "We are content," the Court said, "to rely upon the good sense and judgment of American juries, for evidence with some element of untrustworthiness is customary grist for the jury mill. Juries are not so susceptible that they cannot measure intelligently the weight of identification testimony that has some questionable feature." *Id.* Thus the *Wade-Gilbert-Stovall* doctrine has "largely withered on the vine" in the last two decades. Conyers v. State, 691 A.2d 802, 803 (Md. Ct. Spec. App. 1997). Indeed, the law now differs little in effect from where it stood before the 1967 trilogy. *Id.* at 804.

statements less persuasive. Moreover, strong suggestiveness opens the statements up to attack by defense experts and defense counsel. In this light, it is not clear that the threat of exclusion will add very much incremental incentive to avoid undue suggestion.

Furthermore, as we have indicated in Part I, suggestive questioning has a proper role in investigations of child abuse, because in some settings it generates reports of abuse that open-ended questions might not. Investigations often look not only towards criminal prosecutions but towards civil proceedings aimed at protecting the child and others. It may be unfair to the interviewer, and in any event it will likely chill her investigation, if she is put on a tightrope—one step too passive, and she may miss a truthful report of abuse; one step too aggressive, and the court will exclude the child's testimony.

A best evidence rule, using the harsh sanction of exclusion of evidence, depends on predictability, which requires that a rule operate in a crisp, bright-line manner. We have argued that categorical rules are possible with respect to ploys, such as bribes, threats, ridicule, and peer pressure, that research has shown to create particularly significant risks of false allegations. Generally, however, delicate, fact-based judgments are more appropriate in this area than bright-line rules. Interviewers must take the circumstances of the particular case into account in deciding the degree of suggestiveness appropriate at any given point in a given interview. The interviewer must balance the risk of losing information by remaining too open-ended against the risk of producing false information by being too suggestive.

In short, the best evidence argument may warrant excluding the child's testimony in extreme cases, in which any reasonable interviewer should know that her questioning was unduly suggestive. We believe, however, that it would be difficult or impossible to make the court's decisions both predictable and sensible if they exclude the child's testimony in less extreme cases.

The *wrongful conduct argument* contends that the prosecution should not benefit from evidence that it or those associated with it secure by acting in a reprehensible way. It thus resembles the argument made by Justice Holmes and others in support of the exclusionary rule for evidence secured by unconstitutional search, that it is "a less evil that some criminals should escape than that the Government should play an ignoble part."[256] We do not dispute the principle, but we believe it has rather narrow application in the realm of child interviewing. When an interviewer recklessly or intentionally follows a course that raises a significant risk of leading a child to a false memory of abuse, the interviewer's conduct may be deemed sufficiently wrongful to provide a strong argument for exclusion of the child's testimony. But we do not contend that this degree of irresponsibility characterizes most interviews, even most highly suggestive ones.

In sum, the arguments for exclusion of the child's testimony have substantial weight only in extreme cases, and even then only the best

256 Olmstead v. United States, 277 U.S. 438, 470 (1928) (Holmes, J., dissenting), *overruled in part by* Katz v. United States, 389 U.S. 347 (1967), *quoted in* Stone v. Powell, 428 U.S. 465, 525 (1976) (Brennan, J., dissenting).

evidence and wrongful conduct arguments carry significant force. The reliability argument, principally emphasized by *Michaels*, is unpersuasive. Thus, in extreme cases, when the interviewing technique violates clearly established norms or amounts to an intentional or reckless usurpation of the child's memory—and *Michaels* appears to have been such a case[257]—exclusion is justifiable. In other cases, it is not.

C. Hearsay

Often the child makes an allegation before trial, but does not testify at all at trial or does not testify to the full substance of the earlier allegation. If the prosecutor offers the prior statement into evidence, the defendant will likely object that it is barred by the rule against hearsay and by his right under the Sixth Amendment to the Constitution to "be confronted with the witnesses against him."[258]

In recent years, most jurisdictions have relaxed the application of the hearsay rule so far as it would exclude out-of-court statements by children that allege abuse and are offered to prove the abuse. Some courts have accomplished this end by stretching the limitations on the hearsay exceptions for excited utterances[259] and for statements made for medical diagnosis or treatment.[260] Others have invoked the residual or "catch-all" exception to the hearsay rule now expressed in Federal Rule of Evidence 807.[261] Also, some states have adopted hearsay exceptions specifically tailored for children of "tender years."[262] Because the Supreme Court has, to a large extent, conformed the confrontation right to the prevailing law of hearsay, if evidence satisfies one of these hearsay exceptions it usually will be deemed to satisfy the Confrontation Clause as well.[263]

---

257 The *Michaels* court held that the following factors, present in that case, were "more than sufficient" to justify a pretrial taint hearing: "the absence of spontaneous recall, interviewer bias, repeated leading questions, multiple interviews, incessant questioning, vilification of defendant, ongoing contact with peers and references to their statements, and the use of threats, bribes and cajoling, as well as the failure to videotape or otherwise document the initial interview sessions." State v. Michaels, 642 A.2d 1372, 1383 (N.J. 1994).

258 U.S. CONST. amend. VI.

259 *See, e.g.*, People v. White, 555 N.E.2d 1241, 1243-44, 1250 (Ill. App. Ct. 1990) (approving admissibility of statements made to police after questioning by child's babysitter and mother), *aff'd*, White v. Illinois, 502 U.S. 346 (1992); *In re* Marriage of Theis, 460 N.E.2d 912, 917 (Ill. App. Ct. 1984) (admitting evidence of a three-year-old's head nods in response to specific questions asked by her doctor two months after the alleged event).

260 United States v. Tome, 61 F.3d 1446, 1449-51 (10th Cir. 1995) (on remand) (holding that statements describing earlier abuse made to pediatrician by a girl who was four years old at time of alleged abuse and five and six years old at time of statements were admissible under this exception, because the identity of the alleged assailant, the girl's father, was objectively material to her treatment, her knowledge of materiality being deemed unnecessary); United States v. Renville, 779 F.2d 430, 435–39 (8th Cir. 1985) (emphasizing emotional and psychological injuries in applying this exception).

261 *See, e.g.*, State v. Robinson, 735 P.2d 801, 811 (Ariz. 1987).

262 The progenitor statute is that of Washington. *See* WASH. REV. CODE ANN. § 9A.44.120 (West 2000).

263 *See, e.g.*, *White*, 502 U.S. at 357-58. In *Idaho v. Wright*, 497 U.S. 805 (1990), the Supreme Court did hold inadmissible under the Confrontation Clause statements by a young child that the trial court held satisfied the state's version of the residual exception to the hearsay rule. *See id.* at 818. The *Wright* Court did not, however, suggest that the Confrontation Clause would bar statements falling within the residual exception, but not within any other exemption to the hearsay rule. Rather, the Court held that the

The Court has repeatedly stated that hearsay law and the confrontation right protect "`similar values,'"[264] and the principal value perceived is the need to weed out unreliable hearsay evidence from the reliable. According to the Court, the confrontation right is "primarily a functional right that promotes reliability in criminal trials."[265] Thus, jurisdictions taking a receptive attitude towards hearsay statements by children alleging abuse against them have done so on the grounds that the statements are reliable. In the case of a statement made by a very young child two factors have been particularly influential—first, the apparent absence of a motive for the child to lie and, second, the apparent unlikelihood in some settings that the child could develop a plan to deceive or to concoct her account if it did not in fact reflect abuse she had actually suffered.[266]

The scientific research, however, indicates that in some circumstances children's statements are not particularly reliable. Compared to general hearsay, a statement made by a child who has been subjected to strong forms of suggestion may be notably *un*reliable.[267] The apparent absence of a motive to lie is of significance only to the extent the defendant, in attempting to reconcile the fact that the child made the statement with his theory that the statement is false, contends that the child lied. The defendant may, however, contend principally not that the child lied but that suggestive questioning led her to believe honestly that the assertion was truthful. Similarly, suggestive questioning may make it far more plausible that the child would state a false account of abuse that one would not otherwise expect from a young child who was not abused. For obvious ethical reasons, researchers have refrained from trying to inculcate false memories of abuse; however, there is ample anecdotal evidence that field interviewers sometimes ply child witnesses with information that could be construed as indicative of sexual abuse. Some of this information, if later

---

residual exception is not a "firmly rooted" hearsay exception. *Id.* Thus, a statement falling within that exception and no other exemption must be supported by "particular guarantees of trustworthiness," *id.* at 828, and the Court held that this standard is satisfied only by "the totality of circumstances that surround the making of the statement and that render the declarant particularly worthy of belief," not by corroborating evidence. *Id.* at 820. It is an unresolved question whether "tender years" exceptions such as Washington's will be deemed to be "firmly rooted" and so to satisfy the Confrontation Clause without a need for individualized inquiry into trustworthiness.

264 *E.g.*, *White*, 502 U.S. at 352 (1992) (quoting California v. Green, 399 U.S. 149, 155 (1970)).

265 Lee v. Illinois, 476 U.S. 530, 540 (1986).

266 *See, e.g.*, United States v. Tome, 3 F.3d 342, 351 (10th Cir. 1993), *rev'd on other grounds*, Tome v. United States, 513 U.S. 150 (1995). According to the Tenth Circuit:

> Tome implied that A.T. fabricated the allegations about her father because she wanted to live with her mother. Although this argument does present some motive to lie, we do not believe that it is a particularly strong one. Moreover, Tome's contention would require us to believe that A.T.'s statements were the result of a calculated scheme to deceive. Yet Tome has presented no evidence that the five-year-old A.T. possessed the ability to appreciate the causal relationships inherent in the conception and implementation of such a scheme.

Id. (citations omitted).

267 *See supra* Part I.

incorporated into the child's disclosure, would be considered outside their realm of knowledge, and so viewed as highly credible by fact-finders.[268]

We emphasize two points. First, we are not arguing that all children's statements are unreliable. How reliable a statement is depends on all the circumstances, including—as we have suggested throughout this Article—the nature of the interviewing process to which the child has been subjected. For example, sometimes a child, without any prompting, articulates a detailed and plausible account of abuse soon after the alleged event and, still without prompting, consistently adheres to that account. In such a situation, the child's statement may be very reliable.

Second, even if the statement appears unreliable, that does not necessarily mean that a court should exclude it under an ideal doctrine of hearsay and confrontation. Friedman has argued for some years that the law of hearsay and confrontation is in a most unsatisfactory state.[269] The chief errors, in his view, lie in conforming the confrontation right to the

---

268 For example, in the *Michaels* case, the following exchange occurred between a child witness and an interviewer who gave the child an anatomical doll, a wooden spoon, and a crayon. *See* CECI & BRUCK, *supra* note 11, at 73.

> Interviewer: Where else are you hitting [the doll], on the legs?
> Child: On the bottom. . . .
> Interviewer: Did anything happen back here? [pointing to the doll's buttocks] (laughter) Huh? (laughter)
> Child: She doesn't hurt and do this.
> Interviewer: Stick a crayon in her butt?
> Child: Yeah. (laughter)
> Interviewer: Oh, how does that feel. How does a crayon in your butt feel?
> Child: (laughter)

Id. Elsewhere in *Michaels*, an interviewer plies a different child witness with information relevant to abuse, but the child does not assent to it:

> Interviewer: Did [the defendant] drink the pee pee?
> Child: Please that sounds just crazy. I don't remember about that. Really don't.

Id. at 73. In *State v. Robert Fulton Kelly, Jr.*, the defendant Bob Kelly had allegedly sodomized several children, including a three-year-old boy. *Id.* at 10, 104. At trial, when the boy was six, the prosecutor committed a rather spectacular error—reversing the roles allegedly played by the accused and the boy—and yet the child went along with the suggestion:

> Prosecutor: Do you remember a time where you ever had to do anything to Mr. Bob's hiney with your mouth?
> Andy: No, Ma'am.
> Prosecutor: Do you remember telling Dr. Betty that one time you had to lick Mr. Bob's hiney? Did that happen? Did you ever have to do that, that you didn't want to do it?
> Andy: Yes, ma'am.

Id. at 104. Absent a record of such suggestive questioning, one might infer that the event must have occurred, for otherwise the child could not know of such acts.

269 *See, e.g.*, Richard D. Friedman, *Confrontation and the Definition of Chutzpa*, 31 ISR. L. REV. 506 (1997) (arguing that a defendant generally should not be able to invoke the Confrontation Clause against a statement made by a potential witness—including the victim—whom the defendant caused to be unable to testify) [hereinafter Friedman, *Chutzpa*]; Richard D. Friedman, *Confrontation: The Search for Basic Principles*, 86 GEO. L.J. 1011 (1998) (advocating a reconceptualization of the Confrontation Clause); Richard D. Friedman, *Toward a Partial Economic, Game-Theoretic Analysis of Hearsay*, 76 MINN. L. REV. 723 (1992); *see also* Brief *Amicus Curiae* of the American Civil Liberties Union and the ACLU of Virginia, Lilly v. Virginia, 527 U.S. 116 (1999) (No. 98-5881) (Friedman coauthor with Margaret A. Berger and Steven R. Shapiro), *commented on favorably in Lilly*, 527 U.S. at 140-43 (Breyer, J., concurring); *The Supreme Court, 1998 Term—Leading Cases*, 113 HARV. L. REV. 200, 244 (1999) (advocating "the sort of reevaluation of the Court's hearsay-based Confrontation Clause jurisprudence that Justice Breyer imagined in his [*Lilly*] concurrence").

law of hearsay and in perceiving both as based principally on the need to improve the reliability of evidence. This conjunction results both in hearsay law that is often overly restrictive and in a confrontation right that is insufficiently protective of defendants. We do not attempt to develop this argument in full here. But a system that, according to Friedman, would be far superior to the present one could admit many hearsay statements by children without making the admissibility decision depend on a determination of reliability.[270]

D. Expert Evidence

Traditionally, courts have been loath to allow expert witnesses to testify about factors affecting the credibility of percipient witnesses.[271] Courts were afraid that experts would usurp one of the central functions of the jury, to evaluate the credibility of witnesses. In recent decades, however, courts have been more willing to allow experts to testify about factors that might affect the credibility of a witness in a given situation and that might otherwise be insufficiently understood by a jury.[272] In criminal cases, either the prosecution or the defense may urge the need for expert testimony. For example, a defendant may introduce expert testimony on the vulnerabilities of eyewitness testimony. A prosecutor might introduce expert testimony concerning rape trauma syndrome to help explain the complainant's delay in making her allegation of rape.

---

270 Consider four basic principles that, in Friedman's view, should frame the law in this area. First, the confrontation right should apply only to a statement that is testimonial in nature—in essence, one made under circumstances in which a reasonable declarant would realize that the statement would likely be used to investigate or prosecute a crime. Second, as to statements to which the right applies, a criminal defendant has a categorical right not to have the statement admitted against him unless he has had an adequate opportunity to examine the declarant. Third, this right is qualified only by the proposition that the defendant forfeits it if his own wrongful conduct causes his inability to examine the declarant. *Cf.* FED. R. EVID. 804(b)(6) (providing for forfeiture of hearsay objection when party against whom evidence is offered "engaged or acquiesced in wrongdoing that was intended to, and did, procure the unavailability of the defendant as a witness"). Fourth, if a statement is not within the scope of the confrontation right, it usually should not be excluded by hearsay law. RICHARD D. FRIEDMAN, THE ELEMENTS OF EVIDENCE 327-29, 336-38 (2d ed. 1998).

The interplay of these principles with respect to child declarants could be very complex. But a court might well hold, depending on the circumstances, that the defendant had forfeited his confrontation right by intimidating the child into silence. *See* State v. Sheppard, 484 A.2d 1330, 1345-48 (N.J. Super. Ct. Law Div. 1984) (holding the Confrontation Clause inapplicable because defendant threatened to kill child victim of sexual abuse if she revealed the abuse). Also, it may be that the confrontation right ought not to apply to a statement made by a very young child because the child lacks sufficient maturity and understanding at the time of her statement for the statement to be considered testimonial. Friedman, *Chutzpa*, *supra* note 267, at 532 n.55. Such lack of maturity and understanding might also diminish the probative value of the statement, but not necessarily enough to foreclose admissibility. Friedman has argued:

> If a dog's bark has sufficient probative value, we do not exclude it because the accused has not had a chance to cross-examine the dog. It may be that the cry for help of a young child, even if verbalized, bears a closer material resemblance to the dog's bark than to an adult's accusatory declaration.

Id.

271 *See, e.g.*, United States v. Amaral, 488 F.2d 1148, 1152-53 (9th Cir. 1973).

272 *See, e.g.*, United States v. Smithers, 212 F.3d 306 (6th Cir. 2000); *see generally* ELIZABETH F. LOFTUS & JAMES M. DOYLE, EYEWITNESS TESTIMONY: CIVIL AND CRIMINAL 273-326 (3d ed. 1997) (describing presentation of expert psychological testimony).

Similarly, in child sexual abuse cases, prosecutors often offer, and courts often admit, expert evidence to bolster the complainant's credibility. As Myers has stated,

> Courts permit expert testimony [among other reasons] to explain why sexually abused children delay reporting abuse, why children recant, why children's descriptions of abuse are sometimes inconsistent, why some abused children are angry, why some children want to live with the person who abused them, why a victim might appear "emotionally flat" following the assault, [and] why a child might run away from home . . . .[273]

Myers endorses the use of such testimony, which often fits within the rubric of child abuse accommodation syndrome, on the ground that "[t]o the untutored eye of a juror, such behavior may seem incompatible with allegations of sexual abuse."[274] We agree that such testimony on behalf of the prosecution is proper at least after the defendant attacks the child's credibility, and sometimes even before, if the grounds on which the jury might doubt her credibility are already apparent.

Often, however, it is the defense in child sexual abuse cases that wishes to introduce credibility-related expert testimony, usually to show that the child's statements may have resulted from suggestive questioning. Many courts have admitted such testimony,[275] but some courts still exclude it or confine it rather narrowly.[276] Lyon, while not expressing any opinion on the frequent use by prosecutors of expert testimony to bolster a child's credibility once it has been attacked,[277] expresses doubt about the need for defense expert testimony on suggestibility.[278]

We believe that if evidence supports the conclusion that an interviewer subjected the child to a given set of suggestive influences, then the court should allow the defense to present the testimony of a well-qualified expert as to the plausible effects of those influences.

The research on suggestibility discussed in this Article gives an expert ample basis on which to express an opinion that should easily satisfy the "gatekeeping" scrutiny of the trial court as outlined by *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[279] Indeed, if the "general acceptance" test of

---

273 MYERS, *supra* note 49, § 5.49, at 561-63 (footnotes omitted); *see* Mary Ann Mason, *A Judicial Dilemma: Expert Witness Testimony in Child Sex Abuse Cases*, 19 J. PSYCHIATRY & L. 185 (1991) (reviewing 122 appellate cases in which experts, many of them called by the prosecution, testified about the diagnostic uses of anatomical doll play by children suspected of being abused).

274 MYERS, *supra* note 49, § 5.49, at 560.

275 *See, e.g.*, State v. Kirschbaum, 535 N.W.2d 462, 466-67 (Wis. Ct. App. 1995) (collecting cases in support of the proposition that "[m]any jurisdictions also recognize the utility of expert testimony on the suggestive interview techniques used with a young child and how suggestive techniques can shape a young child witness's answers," but holding that trial court acted properly in excluding expert testimony because defendant did not make a showing of particularized need for the evidence).

276 *See, e.g.*, English v. State, 982 P.2d 139 (Wyo. 1999) (noting trial court's exclusion of expert testimony); United States v. Rouse, 111 F.3d 561 (8th Cir. 1997) (affirming the trial court's limitation on the testimony of the defense expert).

277 *See* Mason, *supra* note 271, at 186 (describing testimony of mental health experts that a particular child fits the profile of a sexually abused child); *see also* note 271 and accompanying text (discussing use of experts by prosecutors in child abuse trials).

278 Lyon, *New Wave*, *supra* note 2, at 1083.

279 509 U.S. 579, 597 (1993).

Frye v. United States,[280] which still prevails in some states, is sensibly applied, such expert opinion should easily satisfy that test as well. As Part I of this Article has shown, this research has used the scientific method of testing, has been extensively subjected to the rigors of publication and review, and has gained broad acceptance in its scientific community. Naturally, as in any area of the social sciences (and some of the hard sciences as well), there is not unanimity on all significant points, and on some points there is a range of interpretations. But a court should not exclude testimony by a qualified expert reflecting an opinion held by a clear majority, or even by substantial proportion, of professionals in the field simply because others hold divergent views. If that were the standard for exclusion, fact-finders would virtually never have the benefit of the experts' knowledge.[281] Thus, we find unpersuasive the rather mysterious opinion of the Eighth Circuit in *United States v. Rouse*, which held that the trial court had acted within its discretion in allowing the defense expert to testify on the basis of his own research, but not on the basis of the research of others.[282]

The question remains whether, and when, an expert's opinion may assist the jury sufficiently to warrant admissibility.[283] Ultimately, this question depends on an assessment of the probative value and prejudice of the expert evidence. Lyon contends that "jurors likely already know" that "children are suggestible."[284] This argument may seem odd, coming near the end of a long article contending that children are not as suggestible as some interpretations of the research indicate. But Lyon's point seems to be that, while children are indeed suggestible to some degree, jurors do not need expert advice to tell them that, and such advice may in fact cause

---

280 293 F. 1013, 1014 (D.C. Cir. 1923), *superseded by rule as stated in* Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

281 Application of a restrictive standard of admissibility for scientific evidence would probably hurt prosecution of child sexual abuse cases far more than help it. *See generally* MYERS, *supra* note 49, §§ 5.11-5.45, at 456-557.

282 According to the court:

> The district court was . . . well within its discretion in ruling that Dr. Underwager should not embellish his own research and opinions by telling the jury about the research and writings of other psychologists because these works have not produced a consistent body of scientific knowledge and therefore admission of other theories and writings would result in a battle of experts that could confuse or even mislead the jury.

Rouse, 111 F.3d at 571. Ralph Underwager is one of two veteran defense expert witnesses whose work Lyon discusses. While acknowledging the "stark contrast" between experts like Underwager and Richard Gardner on the one hand, and the researchers of the "new wave" on the other, Lyon, *New Wave*, *supra* note 2, at 1076-77, Lyon chides Ceci and Bruck for declining to characterize Underwager and Gardner as extremists. *Id.* at 1077 n.398. However, throwing labels around results in much heat and little light. Underwager and Gardner do hold some views that the research data justify. As Lyon points out, Ceci and Bruck cited them for the proposition that children are "potentially" less reliable than adults, *id.* (internal quotation marks omitted), and also for the proposition that "children are capable of high levels of accuracy, provided that adults who have access to them do not attempt to bias their reports." Ceci & Bruck, *Historical Review*, *supra* note 14, at 403 n.1. At the same time, they hold some views that are not so justified or that are based on norms that we do not accept. The same can also be said of many experts who tend to consistently support the credibility of children.

283 *See, e.g.*, FED. R. EVID. 702 ("If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.").

284 Lyon, *New Wave*, *supra* note 2, at 1083.

jurors to overestimate substantially the degree of suggestibility. Myers makes a similar point saying that "some adults" think children are more suggestible than they actually are.[285]

One can easily accept the proposition—which Lyon supports with survey evidence[286]—that many, even most, potential jurors understand that children are more suggestible than adults, and yet recognize the value of expert evidence. Two points are fairly obvious. First, the same surveys reveal that a substantial number of jurors probably do *not* recognize this suggestibility differential.[287] Second, recognizing that children are suggestible, or more suggestible than adults, says little about *magnitude*—how suggestible they are. Perhaps more fundamentally, this Article has shown that the suggestibility of children is not a one-dimensional matter that can be summarized adequately by saying that children are [pick your adjective] suggestible. How plausibly a given child might have alleged abuse even if the abuse did not occur depends on the particular situation, including the extent and nature of the suggestive influences to which the child was subjected. There is no reason to assume that the average potential juror, much less the overwhelming majority of jurors, has a good understanding of all the insights that decades of psychological research have yielded. Consider, for example, the effects of repeated questions and the plausibility of children making false statements about physical events that would be of central concern to them.[288]

Furthermore, there is little reason to assume that expert evidence on this subject will be unduly prejudicial. There is no plausible basis for believing that allowing the defense to present expert testimony will bias

---

285 MYERS, *supra* note 49, § 5.7, at 449 ("Although young children are suggestible, they are not as suggestible as some adults believe." (footnote omitted)).

286 Lyon, *New Wave*, *supra* note 2, at 1083 n.419.

287 One of the surveys cited by Lyon actually indicates rather clearly that jurors are substantially less likely than experts to regard eight-year-old children as suggestible. A. Daniel Yarmey & Hazel P. Tressillain Jones, *Is the Psychology of Eyewitness Identification a Matter of Common Sense?, in* EVALUATING WITNESS EVIDENCE: RECENT PSYCHOLOGICAL RESEARCH AND NEW PERSPECTIVES 13, 33, 35 tbl.2.15 (Sally M.A. Lloyd-Bostock & Brian R. Clifford eds., 1983). According to another of the surveys cited by Lyon, 15% of respondents asserted that children aged five to nine are less suggestible than adults when the influence agent is an adult, and 8% believed that children are about as suggestible. Michael R. Leippe & Ann Romanczyk, *Children on the Witness Stand: A Communication/Persuasion Analysis of Jurors' Reactions to Child Witnesses*, *in* CHILDREN'S EYEWITNESS MEMORY 155, 159 & tbl.9.1 (Stephen J. Ceci et al. eds., 1987). Moreover, 35% believed that children are generally more consistent in action and conversation, and 23% believed that children are about as consistent as adults. *Id.* The third survey cited by Lyon showed that college students tended on average to rate children as more suggestible than adults, but the data as reported do not indicate whether a substantial number of the respondents felt otherwise. David F. Ross et al., *Age Stereotypes, Communication Modality, and Mock Jurors' Perceptions of the Child Witness, in* PERSPECTIVES ON CHILDREN'S TESTIMONY 37, 38 (S. J. Ceci et al. eds., 1989). The authors went on to conduct an experiment, reported in the same paper, in which they presented respondents with a mock trial transcript. They were surprised to find that the mock jurors found eight-year-old witnesses no more manipulable by either prosecutors or defense lawyers than were either twenty-one-year-old or seventy-four-year-old adults. *Id.* at 46 tbl.3.2.

288 Readers can ask themselves the following questions: will nearly all typical jurors appreciate that preschoolers are disproportionately likely to change their answers to yes/no questions (e.g., "Did he touch you there?") when asked such questions repeatedly within the same interview? And will nearly all jurors likely (a) appreciate the myriad of factors that increase the reliability of children's statements (e.g., spontaneity), and then (b) apply them to the case in question?

the jury in favor of the defendant, in the sense of making the jury impose an inappropriately high standard of persuasion on the prosecution.[289] The danger to which Lyon seems to be pointing is the possibility that the jury will give excessive weight to the expert's testimony of suggestiveness. But there appears to be no sound basis for concluding that this danger is real—and that the jury will not only overvalue the expert's testimony but will do it so much that the testimony will be substantially more prejudicial than probative. Juries have convicted defendants in many cases in the face of expert testimony on suggestibility presented by the defense.[290]

In assessing the danger of overvaluation, it is important to bear in mind a major theme of this Article (and for that matter of Lyon's): the degree of a child's testimony is extremely dependent on the particular circumstances of the case. Thus, if the defense expert is performing her function properly, she will testify only to suggestive influences that the jury could reasonably conclude, on the basis of all the circumstances, were present in the case. For example, if there is no basis for concluding that the child was threatened with negative consequences for failure to describe abuse,[291] then research on the effects of such threats would be irrelevant to the case and should not be included in the expert's testimony.[292] If the

---

289 That is, it does not appear plausible that a juror, on hearing the expert testimony concerning suggestibility, will apply an improperly increased standard of persuasion because the child appears less worthy of protection or because (for any prescribed assessment of the probability of guilt) the defendant appears to be a more sympathetic figure.

290 Commonwealth v. LeFave, 714 N.E.2d 805, 811 (Mass. 1999) ("The defendant presented evidence at her trial tending to show that the child witnesses' testimony was unreliable as a result of improper interviewing techniques. The jury nevertheless believed the child witnesses, despite evidence of the use of improper interviewing techniques and the opinions of the defendant's experts.").

291 We speak of a basis for drawing the conclusion, rather than of evidence supporting the conclusion, for two reasons. First, in some circumstances, a court may deem conduct by the prosecution or investigative agents to have obstructed the presentation of such evidence. Most significantly, a court may deem a failure to videotape a formal interview held for investigative purposes improper conduct, supporting an inference that some highly suggestive influences were used. Second, the jury does not come into the courtroom as a blank slate—it is entitled to use its knowledge of the world. *See, e.g.*, John H. Mansfield, *Jury Notice*, 74 GEO. L.J. 395, 395-96 (1985). Thus, if the case arises in the context of a particularly acrimonious divorce, an expert may be allowed to testify as to the effects of suggestive parental questioning even absent proof that the child was exposed to such questioning: the hostility of the accusing parent to the defendant parent *may* be enough to fill the gap.

292 Similarly, we agree with Lyon that, because of the number and nature of suggestive techniques used in the Monkey Thief study, that study should not be the basis for expert testimony in most actual cases. Lyon, *New Wave*, *supra* note 2, at 1039.

But Lyon further claims that Ceci and Bruck advised experts "that they should learn very little about the case (save the child's age) to remain impartial," *id.* at 1027 (citing CECI & BRUCK, *supra* note 11, at 276 n.1). However, Lyon's characterization is misleading. Ceci and Bruck counseled experts "to remain uninvolved in the court proceedings and in the personal lives of the defendants." CECI & BRUCK, *supra* note 11, at 276. Thus, "[e]xperts who plan to review the scientific literature should learn *only enough about the case to assure that their testimony is relevant*." *Id.* (emphasis added). Ceci and Bruck therefore said that when they have testified they have asked that the attorney who sought their testimony "not describe any but the most global case details . . . (*e.g.*, age of child)." *Id.* at 276 n.1 (emphasis added). But this did not mean that Ceci and Bruck thought it improper to learn about the techniques actually used. In fact, Lyon quotes Bruck as saying that before testifying in one case she asked the defense attorney for "some material on the interviewing procedures used with the children so that I could be sure that the suggestibility of young children was a key issue in this case." Lyon, *New Wave*, *supra* note 2, at 1027

defense expert does not exercise self-restraint, the court can ensure that her testimony does not stray beyond the case at hand.

And, of course, the prosecution is not toothless. The prosecutor may cross-examine the defense expert. In doing so the prosecutor should attempt to expose any overgeneralizations that the expert has made or any dubious assumptions on which the materiality of her evidence depends. Moreover, as stated previously, if the defense impeaches the child's testimony, whether by expert testimony or otherwise, the court should allow the prosecution to present its own expert testimony supporting the child's credibility.[293] Likewise, this testimony should be limited to the issues made material by the setting of the case—specifically, to the grounds raised explicitly or implicitly by the defense for being skeptical of the child, or to those that would likely appear plausible to the jury even absent the defense's contention. In short, the adversarial system, through the use of cross-examination and rebuttal witnesses, is resilient and can adequately expose the weaknesses of expert opinions offered by either side.

There does not seem to be any substantial reason to assume that jurors will tend systematically to overvalue defense expert evidence significantly but undervalue prosecution expert evidence—and to do so by enough to warrant exclusion. Some jurors may be confused by the "battle of the experts," of course, and some might unthinkingly treat conflicting expert evidence as a wash, which they can safely ignore. But these are always potential problems when expert witnesses contest each other, whatever the subject. Such problems do not justify insisting that the fact-finder make decisions of enormous importance on the basis of intuition, uninformed by the insights that decades of scientific research have to offer.

E. Videotaping Interviews

The issue of videotaping interviews with a child witness has generated much discussion.[294] Myers has ably summarized many of the factors for and against videotaping.[295] On the positive side of the ledger Myers notes that videotaping gives an interviewer incentive to use proper techniques and preserves a record of such use.[296] Perhaps because he is

---

n.115 (internal quotation marks and citation omitted). *Cf. id.* (quoting Bruck "as respond[ing] negatively to the prosecutor's question [on cross-examination] as to whether she had done `any work on this case in terms of looking at the techniques used.'" (citation omitted)).

293 *See supra* text accompanying note 272.

294 *See, e.g.*, Catherine Stephenson, *Videotaping and How It Works Well in San Diego*, 7 J. INTERPERSONAL VIOLENCE 284, 284 (1992) (videotaping "enhances prosecution efforts and serves the best interests of the child"); Paul Stern, *Videotaping Child Interviews: A Detriment to an Accurate Determination of Guilt*, 7 J. INTERPERSONAL VIOLENCE 278 (1992).

295 MYERS, *supra* note 49, § 1.33, at 85-96.

296 *Id.*, § 1.33, at 86-87. He also points out that videotaping may reduce the need for multiple interviews, in part because expert witnesses may be able to view the tape rather than interview the child. *Id.*, § 1.33, at 85. Also, it preserves the child's statement, including her "emotion, demeanor, and body language," which might make the statement better evidence if it is admissible. *Id.* It may also discourage recantation and may persuade the non-offending parent that abuse occurred. *Id.*, § 1.33, at 87. Additionally, it may encourage confession or negotiation of a plea by the defendant and may refresh the child's

writing from the vantage point of the interviewer, Myer does not mention another equally important argument: if the interviewer does use suggestive techniques, the videotape will reveal it. This Article has emphasized that the degree to which a child's suggestibility accounts for her allegation of abuse depends very largely on the extent and nature of the suggestive influences to which she has been subjected. If all interviews with the child are videotaped, it will substantially reduce, and in some cases effectively eliminate, uncertainty on this score. An interviewer's notes are an unsatisfactory alternative:[297] if historical accuracy is the goal, there is no substitute for electronically recording interviews.

Of course, informal communications with the child, such as by her parents or teachers, will not ordinarily be videotaped.[298] These informal communications are often significant sources of suggestion. Similarly, though it might be feasible for a therapist to tape sessions with a child if there is suspicion of abuse, taping therapy sessions as a matter of course would probably be inappropriate.[299] Moreover, even if therapy sessions could be appropriately recorded, the patient-psychotherapist relationship is privileged,[300] which would probably preclude evidentiary use of the tape. Thus, in many cases, a practice of videotaping investigative interviews does not eliminate all serious questions of suggestiveness. But the intractability of some aspects of the problem is a weak argument against mitigating the problem where that is possible. Videotaping considerably narrows the problem of determining the extent of suggestive influences to which the child is subjected, and that is a great benefit.

The arguments on the other side of the ledger are, once again, based in large part on the fear that the jury will overvalue the evidence in favor of the defense.[301] And once again, we believe that keeping potentially useful information away from the jury is an inappropriate means of

---

recollection. *Id.*, § 1.33, at 87-88. Of course, showing the child the videotape some time later after she has failed to tell the same story is in itself highly suggestive.

297 Accuracy is a problem. Note in this context a study conducted by Bruck, Ceci, and Francoeur in which mothers asked to take careful notes on what their children told them about a surprise event were later highly inaccurate in recounting their children's disclosures. *See* Maggie Bruck et al., *The Accuracy of Mothers' Memories of Conversations With Their Preschool Children*, 5 J. EXPERIMENTAL PSYCHOL.: APPLIED 89, 102-03 (1999); *see also* Lucy S. McGough, *For the Record: Videotaping Investigative Interviews*, 1 PSYCHOL. PUB. POL'Y & L. 370, 376-77 (1995) ("[T]he meaning of an out-of-court encounter can turn on many subtleties of behavior and expression . . . . When no record has been made of the investigative interview, an accurate reconstruction of it is a formidable and perhaps impossible task.").

Furthermore, note-taking is more distracting than videotaping. *In re* R.M. Children, 627 N.Y.S.2d 869, 871, 873 (Fam. Ct. 1995) (noting that interviewer declined to audiorecord interview, and concluding that note-taking is far more distracting); *see* MYERS, *supra* note 49, § 1.33, at 96 (noting that many children forget the camera, and it can be hidden behind one-way glass).

298 MYERS, *supra* note 49, § 1.33, at 95-96.

299 Nelson v. Farrey, 874 F.2d 1222, 1229 (7th Cir. 1989) ("[S]urely [a therapist's] clients would be shocked if he had a routine practice of videotaping therapy sessions with child victims of sexual abuse."); MYERS, *supra* note 49, § 1.33, at 96.

300 *See, e.g.*, Jaffee v. Redmond, 518 U.S. 1 (1996).

301 Myers discusses the possibilities that videotaping will place exaggerated emphasis on inconsistencies in the child's descriptions of abuse and that defense counsel will take portions of the tape out of context or exaggerate interviewer error. MYERS, *supra* note 49, § 1.33, at 88-95. To prevent release of videotaped interviews to the media, he proposes a protective order. *Id.* § 1.33, at 96-97.

ensuring that the jurors will not place too much weight on it. The prosecution has ample opportunity, through the interviewer and expert witnesses, to counter any argument raised by the defense. Judge Posner has argued that the sheer length of interviews leaves an unattractive choice between presenting hours of tape to the jury and risking distortion through editing.[302] But this concern is present whenever a significant amount of evidence is scattered throughout a much larger amount of minimally probative chaff. In practice, we may expect each side to select the excerpts it feels presents its case in the most favorable light and to present evidence and arguments minimizing the importance of the excerpts used by the other side. The court has authority to restrain the parties if the process consumes too much trial time in relation to the probative value of the evidence.

Thus, in accord with most professionals in this field,[303] we believe that it is good practice for official interviewers to videotape interviews conducted with children during an investigation or prosecution of suspected child abuse.[304] Moreover, we believe that, absent exigent circumstances, interviewers should be required as a matter of law to tape such interviews. This is the standard practice in many jurisdictions, and there is no reason why it should not be made mandatory.[305]

In jurisdictions where taping is not required as a matter of law, courts may nevertheless craft evidentiary rules based on a "best evidence" principle that give interviewers strong incentives to follow the practice. The most stringent of these rules would exclude the child's statements, or even her testimony, if the interviews were not taped (again, and throughout this discussion, absent exigent circumstances). This rule, although harsh on its face, would quickly amount in effect merely to an almost absolute requirement of taping.[306] Officials would quickly learn that it is easier to

---

302 *See Nelson*, 874 F.2d at 1229.

303 Myers summarizes the results of pilot projects in California as follows: "Videotaping appeared to have no deleterious effect on investigation or prosecution. A large majority of prosecutors, police officers, and social workers involved in the pilot projects were enthusiastic about videotaping, stating that videotaping improves investigation and reduces trauma for children." MYERS, *supra* note 49, § 1.33, at 84-85 (footnote omitted); *see also* McGough, *supra* note 294, at 385 ("Pretrial videotaping of child witness's accounts is surely an idea whose time has come.").

304 *See In re* R.M. Children, 627 N.Y.S.2d 869, 873 (Family Ct. 1995) ("`As a matter of sound interviewing methodology, nearly all experts agree that' pretrial interviews of alleged child sex abuse victims should be recorded to reduce the danger that the children's recollection is tainted by suggestive questioning." (quoting State v. Michaels, 642 A.2d 1372, 1379 n.1 (N.J. 1994))).

305 *See* McGough, *supra* note 295, at 379 ("[S]tate legislatures may decide to require videotaping of investigative interviews with children, despite the lack of a constitutional imperative."). Lyon contends that "many jurisdictions require videotaping or taping of investigatory interviews," Lyon, *New Wave*, *supra* note 2, at 1026, but we are uncertain that this is technically correct. Certainly numerous states encourage the use of videotaping by making videotaped statements taken under prescribed procedures admissible evidence in some circumstances, *see* McGough, *supra* note 295, at 380 & n.43 (listing statutes), and the practice has been followed as a matter of standard protocol in some foreign jurisdictions for years. *Id.* at 380.

306 In *Idaho v. Wright*, 497 U.S. 805 (1990), the Supreme Court refused to impose an absolute rule that the admission of statements made by the child during an untaped interview violates the defendant's confrontation right. *See id.* at 818-19. We agree that such a rule ought not be imposed as a matter of national constitutional law, and in any event, the Confrontation Clause would probably be the wrong vehicle for such a rule. But

tape than to invite exclusion of evidence, and as a result, very little evidence would actually be excluded. A somewhat softer rule, followed by some courts, makes the failure to videotape the interview a significant factor in determining admissibility of the child's statements or testimony.[307] Other variations would seek to impose the costs of failure to videotape the interview on the prosecution, but without relying on exclusion. Thus, given the failure to record, a defense expert could be allowed to testify as to the potential effect of all suggestive influences to which the child *may* have been subjected. The court might also instruct the jury that the interviewer failed to follow proper practice and that the jury should take the failure into account in evaluating the possibility that the child's statement or testimony was the product of suggestion.

F. Guidance and Control of the Jury

Finally, we come to the end of a trial. Judges in criminal cases in federal court, and in some other jurisdictions, are free to comment to the jury on the weight of the evidence, including factors bearing on the credibility of witnesses.[308] Thus, if a witness is a drug or alcohol abuser, or a former accomplice of the defendant, or if she has received or hopes to receive favorable treatment in return for her testimony, the judge may comment on how these factors affect her credibility.[309] Similarly, judges often comment generally about factors that are believed to affect the credibility of eyewitnesses.[310]

Suppose, then, that a child testifies or makes an admissible out-of-court statement alleging abuse, and evidence supports the conclusions that she was previously subjected to highly suggestive influences. The question arises whether the judge should comment on these influences as potentially affecting her credibility. In most cases, we do not believe that any judicial comment—either supporting or adverse to the child's credibility—is necessary. We believe it usually suffices if the court affords the parties adequate opportunities to present expert evidence on the likely impact of these influences.[311] In an egregious case involving highly suggestive influences, some judicial comment might be appropriate.

Along with the power to comment on the credibility of witnesses, a trial court also has the authority in a criminal case to refuse to enter judgment on a verdict of guilt, and to remit the prosecution to a new trial, if it is persuaded that the verdict is contrary to the great weight of the

---

that does not mean that courts should decline to impose a rule inducing taping as a matter of their own evidentiary law.

307 *See, e.g.*, State v. Wright, 775 P.2d 1224, 1230 (Idaho 1989), *aff'd*, 497 U.S. 805 (1990); *In re R.M. Children*, 627 N.Y.S.2d at 873.

308 *See* KEVIN F. O'MALLEY ET AL., FEDERAL JURY PRACTICE AND INSTRUCTIONS § 11.06, at 89 (5th ed. 2000). *But see id.* (stating "[t]he wisdom of such comments, however, is greatly suspect," because of the difficulty of commenting equitably).

309 *See id.* § 15.03-.05.

310 *See id.* § 15.01.

311 For various forms of instruction given with respect to child witnesses, see *id.* § 15.13. The principal theme of most of these instructions is that the credibility of a child should be determined in the same manner as that of any other witness, and that age is only one factor to be considered. At least three federal courts of appeal—the Seventh, Eighth, and Ninth Circuits—recommend that, because the general standards apply, no special instruction should be given. *Id.* § 15.13, at 451 (2000 & Supp. 2000).

evidence. In making this determination, the court is free to consider the credibility of witnesses.[312] Therefore, an accused might argue that a child's statement or testimony is so tainted by suggestion that a verdict of guilty cannot stand. We believe that this argument should usually, but not always, fail.

Suppose that the case is marked by two factors. First, apart from the child's testimony or prior statement, the prosecution has insubstantial evidence as to at least one element of the charge, most likely to the fact of abuse. Second, the child was subjected to highly suggestive influences. As Part II has shown, the first factor means that the prosecution must rely heavily on the child's allegation. Indeed, the allegation must carry the prosecution's case the very large distance from the presumption of innocence to the constitutionally mandated standard of proof beyond a reasonable doubt. And, as discussed in Part II, the court might conclude, on the basis of the second factor, that the denominator of the likelihood ratio–the probability that the child would make the allegation even though it is false–cannot reasonably be perceived as minuscule. Putting these two considerations together, the court might well conclude that a jury could not reasonably find that the prosecution satisfied its standard of persuasion.[313]

If prosecutors select cases appropriately, cases with both these features will be rare. The judicial power to reject a verdict, even if usually kept in reserve, can be a powerful force ensuring that prosecutors do indeed make careful selections.

Conclusion

In this Article, we have summarized some of the research on the suggestibility of children. The research reveals that the degree to which children are suggestible depends to a large extent on how investigators conduct interviews. It also indicates that abuse investigations are often conducted in such a way as to enhance the dangers of suggestibility. We have offered an analytic framework for considering legal implications of the research findings and have presented a set of policy recommenda tions that we believe are consonant with those findings. These recommendations are, we believe, even-handed, reflecting a bias for neither the prosecution nor the defense. The proof of our even-handedness may be that we have exposed ourselves to a two-flank attack. Prosecutors may complain about our recommendations that in some circumstances children's statements regarding abuse should be regarded as unreliable for hearsay purposes, that courts should often be receptive to expert evidence emphasizing the suggestibility of children, that videotaping of interviews

---

312 *See* Tibbs v. Florida, 457 U.S. 31, 38 n.11 (1982); United States v. Lincoln, 630 F.2d 1313, 1319 (8th Cir. 1980).

313 A court has the raw power to enter an outright judgment of acquittal on the basis of doubts about the credibility of the witness. Such an order, even if illegitimate, is effectively unreviewable. *See* Fong Foo v. United States, 369 U.S. 141, 142-43 (1962). We have concentrated on a new trial order because, although it lacks the finality of an outright judgment of acquittal, it is clearly legitimate. An alternative way to achieve finality, with a greater veneer of legitimacy than a judgment of acquittal based on doubts about credibility, would be for the court to reconsider its ruling admitting the child's testimony or prior statement and then hold that there is insufficient evidence to support a conviction. The manipulativeness of this approach diminishes its attractiveness.

should be mandatory, and that occasionally the weakness of a child's statement or testimony should cause the court to refuse to enter a judgment of guilt. Defense lawyers on the other hand, are likely to complain about our recommendation that, in all but egregious cases, the child should not be rendered incompetent to testify because she was exposed to strongly suggestive interviewing techniques.

We suspect that scholars who have recently challenged the legal significance of the psychological research emphasizing children's suggestibility are not motivated principally by antipathy to policy proposals such as the ones we have presented. Rather, we suspect that they are concerned about a matter of mood. In an earlier day, children's statements were often not taken seriously. As a result, child sexual abuse was under-reported and under-prosecuted. Thus, there is a concern that scientific research emphasizing that children are suggestible will be taken for more than it is worth and lead us back to pervasive and unwarranted devaluation of children's statements and testimony.

We recognize this concern. But we balk at any approach that makes it more difficult to recognize, and thus mitigate, problems in the way children alleging abuse are interviewed. And we confess that we do have a bias of an intellectual sort, which underlies our predilection in favor of allowing both the child and experts to testify. Accurate fact-finding, we believe, is not best achieved by trying to maintain and regulate the fact-finders' ignorance. The best cure for possible misunderstanding is not to keep an area in darkness, but rather to bathe it in light.