This article was downloaded by: [University of Arizona]
On: 15 January 2014, At: 05:45
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Child Sexual Abuse
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wcsa20

Interviewing Children Versus Tossing Coins: Accurately Assessing the Diagnosticity of Children's Disclosures of Abuse

Thomas D. Lyon [a] , Elizabeth C. Ahern [a] & Nicholas Scurich [a]
[a] University of Southern California , Los Angeles , California , USA
Published online: 16 Feb 2012.

To cite this article: Thomas D. Lyon , Elizabeth C. Ahern & Nicholas Scurich (2012) Interviewing Children Versus Tossing Coins: Accurately Assessing the Diagnosticity of Children's Disclosures of Abuse, Journal of Child Sexual Abuse, 21:1, 19-44, DOI: 10.1080/10538712.2012.642468

To link to this article: http://dx.doi.org/10.1080/10538712.2012.642468

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Journal of Child Sexual Abuse, 21:19–44, 2012
Copyright © Taylor & Francis Group, LLC
ISSN: 1053-8712 print/1547-0679 online
DOI: 10.1080/10538712.2012.642468

Routledge
Taylor & Francis Group

Downloaded by [University of Arizona] at 05:45 15 January 2014

DIAGNOSTIC UTILITY OF THE ALLEGED CHILD VICTIM'S DISCLOSURE STATEMENT AND BEHAVIOR SYMPTOMS

Interviewing Children Versus Tossing Coins: Accurately Assessing the Diagnosticity of Children's Disclosures of Abuse

THOMAS D. LYON, ELIZABETH C. AHERN, and NICHOLAS SCURICH
University of Southern California, Los Angeles, California, USA

We describe a Bayesian approach to evaluating children's abuse disclosures and review research demonstrating that children's disclosure of genital touch can be highly probative of sexual abuse, with the probative value depending on disclosure spontaneity and children's age. We discuss how some commentators understate the probative value of children's disclosures by: confusing the probability of abuse given disclosure with the probability of disclosure given abuse, assuming that children formally questioned about sexual abuse have a low prior probability of sexual abuse, misstating the probative value of abuse disclosure, and confusing the distinction between disclosure and nondisclosure with the distinction between true and false disclosures. We review interviewing methods that increase the probative value of disclosures, including interview instructions, narrative practice, noncontingent reinforcement, and questions about perpetrator/caregiver statements and children's reactions to the alleged abuse.

Submitted 22 March 2011; revised 27 July 2011; accepted 1 August 2011.

Preparation of this article was supported in part by National Institute of Child and Human Development Grant HD047290.

Address correspondence to Thomas D. Lyon, Gould School of Law, University of Southern California, 699 Exposition Boulevard, Los Angeles, CA 90089-0071. E-mail: tlyon@law.usc.edu

KEYWORDS sexual abuse, interviewing, disclosure, Bayes' theorem

Because the child victim tends to be the only eyewitness to sexual abuse (other than the perpetrator), the child's report is an extremely important piece of evidence in any abuse case. Researchers have identified a number of ways in which child interviewing can be improved, both by increasing the number of true details in reports of children who have been abused and by decreasing the likelihood that children who have not been abused make false reports. We will argue that children's disclosures of abuse can be highly probative of abuse, particularly when they are elicited using techniques supported by research.

Kuehnle and Connell (2009) brought together a diverse group of scholars with widely varying views regarding the potential to distinguish between true and false suspicions of child sexual abuse. The first three chapters, authored by David Faust and colleagues, took a Bayesian approach to understanding the process by which evaluators attempt to determine if abuse has occurred (Bridges, Faust, & Ahern, 2009; Faust, Bridges, & Ahern, 2009a, 2009b). They painted a pessimistic picture regarding the ability of research on abused and nonabused children to increase the diagnostic accuracy of sexual abuse assessments. We agree that the Bayesian approach is an excellent framework for understanding the difficulties of evaluating abuse allegations. However, we will argue that Faust and colleagues' examples may lead one to underestimate the value of disclosures of abuse.

Chapters by Brown and Lamb (2009) and Herman (2009) discussed the National Institutes of Child Health and Human Development (NICHD) protocol for interviewing children about abuse. Brown and Lamb (2009) reviewed evidence that the NICHD protocol increases the quantity and quality of information that children questioned about abuse provide. We agree and will elaborate on interviewing methods that can increase the diagnosticity of disclosures. Herman (2009) argued that the NICHD protocol does a poor job of distinguishing between true and false allegations of abuse. We believe that Herman understated the probative evidentiary value of a disclosure of abuse and the advantages of the NICHD protocol.

First, we will present a primer on Bayesian thinking about probabilities. This is intended to introduce the subject to novice readers. Second, we will critique Faust and colleagues' discussion of the probative value of sexual abuse indicators. We will explain how they fell prey to the inverse fallacy in their argument about the limited probative value of indicators. We will also argue that they underestimated the likelihood that children seen for sexual abuse assessment have been abused because of the likelihood that assessment is triggered by disclosure. Third, we will discuss the probative value of children's disclosures of genital touch and argue that such disclosures often have very high probative value, taking into account the child's

Downloaded by [University of Arizona] at 05:45 15 January 2014

age and the type of questions asked in assessing disclosure. Fourth, we will discuss innovations in interviewing that increase probative value, including interview instructions and open-ended narrative practice. Fifth, we will take issue with Herman's pessimistic conclusions about the probative value of children's disclosures.

THE BAYESIAN APPROACH: UNDERSTANDING HOW TO ASSESS EVIDENCE OF ABUSE

Before we discuss other authors' treatment of the diagnosticity of disclosures, it is important to describe the Bayesian approach (Bolstad, 2007). A mainstay of probability theory is Bayes's theorem. Bayes's theorem formally prescribes the extent to which one ought to update one's belief in a given proposition in light of receiving new information. According to the theorem, one updates one's belief by multiplying the *prior probability* of the proposition by the value of the new information (Lyon & Koehler, 1996). Let's imagine that we are attempting to determine the probability that a child was abused and that the new information is "this child disclosed sexual abuse in a forensic interview." The prior probability is the probability that the child was abused based on all we knew about the child before considering the child's disclosure.

If the child was randomly selected from the population to be questioned about sexual abuse, the prior probability would be equal to the base rate of abuse among children in the population. Faust and colleagues (2009a,2009b), for example, assumed an abuse base rate of 5%. We will return to the issue of base rates later, but for now it is important to understand that we would use the base rate as the prior probability if we had no reason to suspect abuse before we learned that the child disclosed abuse in an interview.

Bayes's theorem is easiest to understand if one speaks in terms of odds rather than percentages. For example, if the likelihood of abuse is 50%, this is even odds; the likelihood the child was abused (50%) is equal to the likelihood that the child was not abused (50%). Odds can be expressed as a ratio; even odds would be 1:1. It is not very difficult to convert percentages to odds ratios. The percentage likelihood is the first value in the ratio and 100 minus the percentage likelihood is the second value in the odds ratio. Hence, if the likelihood of abuse is 50%, 50 is the first value in the odds ratio. The second value is 100 minus 50, or 50. So for a 50% likelihood, the odds ratio is 50:50, which simplifies to 1:1. If one assumes a base rate of 5%, then the odds would be 5:(100–5) or 5:95, which simplifies to 1:19.[1]

The value of the new information (which can be called the probative value or the diagnosticity) is typically quantified as a *likelihood ratio*. The likelihood ratio is the ratio of the true positive rate and the false positive

Downloaded by [University of Arizona] at 05:45 15 January 2014

rate. The true positive rate is the likelihood that children disclose abuse when they *have* been abused. The false positive rate is the likelihood that children disclose abuse when they have *not* been abused. In other words, the likelihood ratio will be the proportion of abused children who disclose abuse divided by the proportion of nonabused children who disclose abuse. For example, assume that 40% of abused children disclose abuse, whereas 2% of nonabused children falsely claim abuse. The likelihood ratio for disclosure is the proportion of abused children who disclose abuse (40%) divided by the proportion of nonabused children who disclose abuse (2%), or 20.

The likelihood ratio indicates how much the odds of abuse are increased by the new information (the disclosure). A likelihood ratio of 20 means that the evidence increases the odds of abuse 20 times. Hence, in order to calculate the likelihood that a child who discloses abuse was in fact abused, we multiply the prior odds by the likelihood ratio. Thus, if the prior odds of abuse are 1:19 and the likelihood ratio is 20, the likelihood of abuse is 20:19, which is slightly higher than 1:1 odds.

Imagine that a child is randomly selected from a group of children, only 5% of whom were abused. This would mean that the prior probability of abuse, before we interview the child, would be equal to the base rate of abuse: only 5%, or 1:19 odds. Then imagine that the child is interviewed about abuse, and discloses abuse, and the likelihood ratio of a disclosure is 20. We would then conclude that the odds that the child really was abused was 20:19, or only slightly better than 1:1. In percentage terms, it would be 51%.[2]

Commentators sometimes describe likelihood ratios as being strong or weak evidence. For example, Wood (1996) notes that a likelihood ratio of 3 is considered weak, 5 weak to moderate, 7 moderate, 14 moderate-to-strong, and 20 strong. But even strong evidence may not be convincing if the prior odds of abuse are very low (Faust et al., 2009a, 2009b; Myers, 2005; Poole & Wolfe, 2009). Indeed, the example we just discussed is an illustration of this problem. We assumed that disclosure has a likelihood ratio of 20, which makes disclosure strong evidence. But because the prior odds of abuse were only 1:19, disclosure makes it only slightly more probable than not that the child was abused.

On the other hand, if one starts with high prior odds, even a likelihood ratio that is considered weak can convince one that abuse occurred. It is unusual for children to be plucked at random from the population and given a forensic interview for abuse. To the extent that evidence of abuse exists before the child is interviewed, then the prior probability of abuse is likely to be much higher than the base rate. For example, if the prior odds are 1:1, then a likelihood ratio of three increases the odds to 3:1 or 75%. In other words, one might be on the fence before the disclosure but firmly convinced after the disclosure.

Downloaded by [University of Arizona] at 05:45 15 January 2014

There are two other important points to make about the likelihood ratio. The reader will recall that we assumed 40% of abused children disclose abuse (the true positive rate) and 2% of nonabused children disclose abuse (the false positive rate). Forty percent is a minority of abused children. However, 2% is a very small number, so even if most abused children deny abuse, disclosure can be strong evidence of abuse. Similarly, the fact that evidence is common or rare among abused children tells us very little about the diagnosticity of that evidence. Imagine that 60% of abused children experience nightmares. This may mean nothing at all if 60% of nonabused children experience nightmares. On the other hand, imagine that 1% of abused children suffer from gonorrhea. This may be very strong evidence of abuse if the percentage of nonabused children who suffer from gonorrhea is much less than 1%. Hence, the likelihood ratio teaches us that we cannot assess the diagnosticity of any piece of evidence without knowing both the true positive rate *and* the false positive rate (Lyon & Koehler, 1996).

A second point is that it is very important not to confuse the different terms. The true positive rate and the false positive rate are quite different. Knowing the true positive rate does not tell us what the false positive rate is. Novices will sometimes assume that the true and false positive rates must sum to one, which is incorrect. They are calculated based on different groups of children: the true positive rate is calculated based on abused children, and the false positive rate is based on nonabused children.

It is also easy to confuse the true positive rate with the probability of abuse given disclosure. The true positive rate is the probability of *disclosure given abuse*. The reader should notice that this sounds similar to the probability of *abuse given disclosure*. Confusion between these two probabilities has been given different names: the inverse fallacy (Kaye & Koehler, 1991), transposing the conditional (Evett, 1995), and the prosecutor's fallacy (Thompson & Schumann, 1987). We will use the term inverse fallacy.

COIN TOSSES VERSUS TESTS THAT ARE 75 % ACCURATE

Now we are ready to consider the arguments made by Faust and his colleagues (Bridges et al., 2009; Faust et al., 2009 a, 2009b). They purported to show that a test that is 75% accurate may nevertheless be worse than a coin toss in distinguishing between true and false allegations of abuse. If their proposition is true, it would indeed undermine one's confidence in assessing child sexual abuse claims. However, the "75% accurate" argument suffers from two principal problems that may mislead the novice reader. First, the "75% accurate" test is actually only weakly probative, with a likelihood ratio of three. Understanding how Faust and colleagues define "accurate" will make this clear. Second, the "75% accurate" test is in fact stronger evidence

Downloaded by [University of Arizona] at 05:45 15 January 2014

than a coin toss. Because of looseness in the meaning of the term "accurate," Faust and colleagues fell prey to the inverse fallacy.

Faust et al. (2009a) wrote, "Assume there is an evaluative method or test with a 75% accuracy rate in separating [abused children from non-abused children]. Although 75% is far from perfect, it might well seem that the test has much to offer and could help us" (p. 10). Seventy-five percent certainly sounds like a good test. The reader might assume that if a test is 75% accurate, this means that if the test says the child is abused, it is 75% likely to be correct. Surely, if it is correct 75% of the time, then it could be usable as a means of distinguishing abused children from nonabused children.

However, Faust and colleagues used the term "accuracy" in a specific and somewhat idiosyncratic way. They described a test that has a 75% true positive rate *and* a 75% true *negative* rate. So, if the child was abused, the evidence would be present 75% of the time (the true positive rate). If the child was not abused, the evidence would be absent 75% of the time (the true negative rate). In order to assess the value of the "75% accurate" test, one needs to know the likelihood ratio. The numerator of the likelihood ratio is the true positive rate, and the denominator is the false positive rate. The true positive rate of the "75% accurate test" is 75%. The false positive rate can easily be calculated using the true negative rate. The false positive rate of the "75% accurate" test is 25%.[3] This means that the "75% accurate" test has a likelihood ratio of only three.

Translating the "75% accurate" test into a likelihood ratio gives us a better picture of how useful the test is. The likelihood ratio of three is considered weak evidence, and a false positive rate of 25% is quite high. For example, if the evidence in question was a disclosure of abuse, then a false positive rate of 25% would mean that the questioning method elicited false claims of abuse from 25% of nonabused children. Indeed, a false positive rate of 25% ensures that the evidence will be weak; the highest possible likelihood ratio is four because the true positive rate cannot be greater than 100%.

Furthermore, translating the "75% accurate" test into a likelihood ratio emphasizes the need to consider the prior odds of abuse. Faust and colleagues made the correct point that if one starts with a low probability of abuse and finds weak evidence of abuse (the "75% accurate" test), the likelihood of abuse will remain low. They discussed two different cases, either assuming a prior probability of abuse of 17% (equal to a prior odds of 1:5) or 50% (prior odds of 1:1). In the case in which the prior probability of abuse is 17%, a positive result on the "75% accurate" test would increase the odds of abuse by three, from 1:5 to 3:5, or 38%. If the prior probability is 50%, then a positive test result would increase the odds of abuse by three, from 1:1 to 3:1, or 75%.

With respect to the 38% figure, Faust and colleagues (2009a) concluded, "[W]e will be wrong far more often than we are correct: our accuracy rate will only be 38%, or worse than a coin toss, an outcome that is hardly

Downloaded by [University of Arizona] at 05:45 15 January 2014

satisfactory" (p. 12). This is an example of the inverse fallacy. The problem lies in the ambiguity of the term "accuracy rate." The authors start with the assumption that the "accuracy rate" of a coin toss is 50%. However, by comparing coin tosses to the 38% "accuracy rate," the authors confused the probability of evidence given abuse (50%, using a coin toss) to the probability of abuse given evidence (38%).

To demonstrate the inverse fallacy, consider how a coin toss can be considered both 50% accurate (likelihood of evidence given abuse) and less than 50% accurate (likelihood of abuse given evidence). Imagine that you take a group of 100 children, 10 of whom have been abused, and you toss a coin to determine who is abused and who is not abused: heads abused, tails nonabused. If you take any child who has been abused and toss a coin, the likelihood that you will call the child abused is 50%. Hence, the coin appears to be 50% "accurate." But now consider the entire group of children. By tossing coins, you will randomly select 50 of the 100 children "as abused." The proportion of the 50 children who really will have been abused will be the same as the original group: 10%. Therefore, the coin toss will only be correct 10% of the time. Hence, the probability of abuse given heads is 10%, but the probability of heads given abuse is 50%. The two probabilities are quite different and should not be confused.

By committing the inverse fallacy, Faust and colleagues were able to dismiss the use of relevant evidence. The truth of the matter is that evidence with a likelihood ratio of three increases the odds of abuse by three times. A coin toss does not affect the odds of abuse (and we do not recommend using it!). If the prior probability is low, it will remain low in the face of this weak evidence. But if the prior probability is close to 50% (or 1:1) odds, this evidence can make the difference. In sum, Faust and colleagues' argument purported to show that a test can be 75% accurate and yet worse than a coin flip. Their "75% accurate" test is only weakly diagnostic (likelihood ratio of three) and nevertheless clearly better than a coin toss.

UNREALISTIC ASSUMPTIONS ABOUT THE LIKELIHOOD OF ABUSE AMONG CHILDREN SEEN FOR EVALUATION

Faust and colleagues rightly speculated that children are typically evaluated for sexual abuse because "someone suspects abuse" (Bridges, Faust, & Ahern, 2009, p. 25). The relevance of the suspicion, however, is very important. In their hypothetical example, Faust and colleagues discussed the use of a "screening method with robust validity" (Faust, Bridges, & Ahern, 2009a, p. 11) and assumed the following numbers:

> For purposes of this example, assume that the base rate for child sexual abuse in the catchment area for Clinic 1 is 5%. Thus, there are 19 non-abused children for every sexually abused child. Assume that all children

Downloaded by [University of Arizona] at 05:45 15 January 2014

> known to display explicit sexual behavior, such as imitating sexual relations with dolls, are referred for sexual abuse evaluations. Assume further that explicit sexual behavior occurs in about 20% of sexually abused children and 5% of non-abused children (figures that roughly align with research estimates. . .). (p. 11)

The evidence—sexual play with dolls—has a likelihood ratio of 4 (true positive rate = 20%; false positive rate = 5%), which would be considered "weak" evidence. If one starts with a base rate of 5% (or 1:19 odds), multiplying this by a likelihood ratio of 4 leaves us with 4:19 or approximately 1:5 odds of abuse. This is only 17%. The 17% figure is the one that the authors use, in conjunction with the "75% accurate" test, to argue that coin tosses are better than abuse evaluations.

We will not challenge the assumption that 5% is a reasonable base rate for the population. As others in this special issue emphasize, the prevalence of abuse among children is substantially higher than 5%, particularly among girls (Everson & Faller, this issue). However, we are not sure that the prevalence rate is the correct base rate. The prevalence rate is the percentage of children who are abused at any time during childhood, but many children are questioned long before their childhood ends, and therefore the likelihood they will have been abused will be lower.

What we find questionable about the hypothetical is that it assumes communities screen for sexual abuse by giving children at large anatomically correct dolls with which to play. It is hard to imagine that anyone would ever undertake such an approach. First, we are not aware of any screening of children at large for sexual abuse. Second, if we did conduct population screening, it is unlikely that we would have children play with anatomical dolls. A fairly broad consensus has emerged that observation of doll play is not a valid means of evaluating children for sexual abuse. Rather, supporters of the use of dolls advocate their use as a demonstration aid for children who are disclosing abuse (Pipe & Salmon, 2009).

The fact is that disclosure is a common, if not the most common, reason children are evaluated for sexual abuse. For example, Heger, Ticson, Valsquez, and Bernier (2002) examined the records of 2,400 children referred for evaluation of sexual abuse and found that 70% of the children had previously disclosed abuse.

THE DIAGNOSTICITY OF THE DISCLOSURE OF GENITAL TOUCH

It is therefore most relevant to consider the diagnosticity of the disclosure of abuse in determining the likelihood that professional evaluations of abuse allegations are correct. The evidence suggests that children's spontaneous reports of genital touching are strong evidence that touching has occurred

Downloaded by [University of Arizona] at 05:45 15 January 2014

but that as the questions become more direct and potentially leading, and as younger children are questioned, the diagnosticity of a report of genital touch decreases.

For example, an oft-cited study by Saywitz, Goodman, Nicholas, and Moan (1991) examined 72 five- and seven-year-olds girls' memories of a pediatric examination. For half of the girls, the examination included genital touch, and for the other half, the examiner substituted an examination for scoliosis. When asked free recall questions about the event one month afterward, 22% of the girls who had been touched mentioned vaginal touch, and none of the girls who had not been touched did so. When asked a direct question about genital touch with the aid of an anatomically correct doll ("Did that doctor touch you there?" while pointing to the doll's vagina, p. 684), 86% of the girls who had been touched acknowledged genital touch and 3% of the girls who had not been touched falsely claimed that they had. The results suggest that a spontaneous report of genital touch (in response to a very general question about one's interactions with a person) is very strong evidence. We might say that the likelihood ratio is infinitely large, although it would be safer to say that in this study recall of genital touch provided conclusive proof that touching occurred. Even a simple "yes" response to a direct question with an anatomically correct doll constituted strong evidence of genital touch (86/3, or a likelihood ratio of 29).

Steward and colleagues (1996) conducted a similar study in which children were either touched or not touched during a physical exam. However, they included children from three to six years of age, used either anatomically detailed dolls or anatomical body diagrams, and interviewed the children one day, one month, and six months after the exam. Direct questions about genital touch asked with the assistance of a doll or drawing elicited relatively weak evidence of touching; the likelihood ratios ranged from five to nine. True positive rates ranged from 73% to 86%, whereas false positive rates ranged from 8% to 12%. In contrast, when children were asked, "Did the doctor touch you?" and then, if they answered yes, were asked, "Where were you touched?" there were no false reports of genital touch in the verbal-only interviews at any of the time periods, whereas the true positive rates ranged from 18% to 33%. Disclosure of genital touching without the use of dolls or drawings constituted conclusive evidence of touch. Three-year-olds performed worse than six-year-olds.

Indeed, as younger children are tested, the false positive rates increase, making direct questions with dolls increasingly risky. Bruck and colleagues (Bruck, Ceci, Francouer, & Renick, 1995, Bruck, Ceci, & Francouer, 2000) questioned preschool children immediately after a pediatric examination and included direct questions with the aid of an anatomically detailed doll. Some of the questions were highly suggestive, but we will only consider the questions about genital touch most analogous to those asked by Saywitz and colleagues (1991) and Steward and colleagues (1996): yes–no questions that

Downloaded by [University of Arizona] at 05:45 15 January 2014

asked whether the doctor touched the child's genitalia. Although Bruck and colleagues did not find significant age differences, their sample sizes were small and the percentages suggested improvement with age. The three-year-olds (mean age 2 years 11 months) performed close to chance, such that the likelihood ratio was not greater than one. Specifically, children exhibited a 50% true positive rate and a 42% false positive rate. The four-year-olds (mean age 4 years 1 month) exhibited a 45% true positive rate and a 14% false positive rate, which suggests a likelihood ratio of 3, or weak evidence of touching.

Critics of the Saywitz and colleagues (1991) and Steward and colleagues (1996) studies have pointed out that they did not contain a number of other suggestive elements that may occur when children are questioned about abuse, particularly when the questioner strongly suspects that abuse has occurred (Bruck & Ceci, 1996, 1999). Critics also focus on the false positive rate for anal touch, which is often inexplicably quite a lot higher (Bruck & Ceci, 1996). Indeed, because of the increased false alarm rate with the use of anatomical dolls or drawings, we have cautioned against their use (Lyon, Lamb, & Myers, 2009). Hence, this discussion is not intended as a defense of the use of dolls but rather illustrates the fact that even methods criticized as unduly suggestive will often elicit reports that increase the likelihood that genital touching occurred, particularly with children older than preschool age.

To our knowledge, researchers have not asked about genital touch when using more suggestive methods. They have, however, asked questions that would strongly intimate that genital touching may have occurred. For example, Bruck & Ceci (2004) summarize one study as finding a 68% false positive rate for "misleading questions with abuse themes (e.g., 'He took your clothes off, didn't he?')" when the questions are asked in a high-pressure interview (p. 231, citing Finnila, Mahlberga, Santtilaa, & Niemib, 2003).

Finnila and colleagues (2003) tested four- and eight-year-old children who were either the most suggestible or least suggestible in their age groups. Children participated in a 10-minute interaction with an adult male and were questioned one week later. In the "low pressure" group, children were asked questions such as, "He took your clothes off, didn't he?" In the "high pressure" group, children were first told, "I have already spoken to the big kids and they told me that he did some bad things that he shouldn't have done. Now I would like to know if you also have such a good memory and can help me, because I really need your help to find out what happened." For each question, the interviewer first told the child what other children had reported (e.g., "The other kids told me that he took their clothes off. He took your clothes off too, didn't he?" If the child said "yes" or "maybe," the interviewer "praised the child's memory." If the child said "no," the interviewer implied that the child's answer might be wrong, repeated what

Downloaded by [University of Arizona] at 05:45 15 January 2014

"the other kids" had said, and re-asked the question. Again, if the child now said "yes" or "maybe," the interviewer praised the child's memory; p. 42).

Although Bruck and Ceci (2004) report a 68% false positive rate for misleading questions generally, citing the "clothes off" question as an example, Finnila and colleagues (2003) report specific percentages for the "clothes off" question. In the "high pressure" condition, 9% of children answered "yes" to "He took your clothes off, didn't he?" In the "low pressure" condition 3% did so. The rate in the "high pressure" condition is significantly higher than in the "low pressure" condition, but at the same time, given the pressures inflicted on the child subjects, remarkably low.

We would never advocate the use of the high-pressure tactics, but if they have been used, an important question is whether the child's responses are nondiagnostic. Given a 9% false positive rate, if the high-pressure interview elicited a true positive rate of 60% or more, then a "yes" response would be considered moderately strong evidence. However, because Finnila and colleagues (2003) did not test a condition in which clothes were in fact removed, one cannot assess the diagnosticity of children's "yes" responses to the "clothes off" questions.

Similar to the anatomical dolls research, the questions are likely to elicit higher error rates in younger children. Goodman and Aman (1990) interviewed three- and five-year-old children (some with anatomical dolls) about a ten-minute interaction with a man and asked the "clothes off" question as part of a group of suggestive questions ("He took your clothes off, didn't he?" "He kissed you, didn't he?" and "How many times did he spank you?"). Whereas three-year-olds false alarmed 21% of the time, five-year-olds did so only 2% of the time (p. 1867). Similarly, the authors asked children more directly about genital touch ("Did he touch your private parts?" "Did he keep his clothes on?" and "Show me where he touched you"). Whereas 14% of the three-year-olds' responses "would be likely to raise suspicions of abuse," 5% of the five-year-olds responses were so coded (p. 1865). In contrast, Finnila and colleagues (2003) report no age differences. However, they did find that children's accuracy was related to their performance on a suggestibility scale, and that scale, in turn, was related to age. Hence, by testing the effects of both the suggestibility scale and age in the same model, the effect of age may have been captured (and thus obscured) by the effect of the scale.

In sum, the fact that a child has disclosed genital touch has substantial probative value in concluding that the child was sexually abused. If the child reports abuse in free recall, this is arguably very strong evidence of abuse, and even if the child's disclosure is only in response to more direct questions, this has some probative value. As children mature, their "yes" responses to direct questions increase in probative value.

Downloaded by [University of Arizona] at 05:45 15 January 2014

THE POTENTIAL REDUNDANCY OF DISCLOSURE

When children are referred for evaluation of sexual abuse, they are obviously questioned about abuse. If most referred children are referred because of a disclosure, this means that the evaluation will include both the child's previous statements and any disclosure the child makes during the evaluation. Faust and colleagues (Bridges, Faust, & Ahern, 2009) raised a potential problem with double-counting evidence, one that they called double-dipping. The evaluator should not assume that the child's responses during the evaluation are independent of the child's prior disclosure. As they wrote, "[T]he problem is that once the variable or variables have been used during Phase 1 as a basis for referral, then any positive qualities they might have had for the detection of abuse are neutralized when they are reapplied in Phase 2, with the result often being to reduce judgmental accuracy" (p. 29). Faust and colleagues noted that redundant information adds no incremental validity to the assessment but may lead to overconfidence in the judgment about the occurrence of abuse because of the evaluator's failure to recognize redundancy.

Faust and colleagues applied this reasoning to sexual behavior. Does it also apply to interviewing? In order for the new disclosure to add nothing to the evaluator's judgment, the second disclosure must be truly redundant. However, there are a number of reasons why the evaluation interview will not be redundant, particularly if it relies as much as possible on open-ended questions.

First, the child's initial disclosure is likely to have contained few details. Children's first reports are rarely to the authorities, and nonprofessionals are unlikely to conduct exhaustive forensic interviews when children first disclose. Second, if the child's disclosure was to an interested person (such as a parent involved in a custody dispute with the accused), then there are sure to be doubts about the credibility of that person's report. The disclosure may never have occurred at all, may have been ambiguous and subject to misinterpretation, or may have been elicited through leading questions. Third, even if the recipient of the report was unbiased, the child's report may have been elicited through direct questions, which would reduce the diagnosticity of the disclosure because responses to direct questions tend to be less accurate than free recall. To the extent that the evaluation interview is more open-ended and less leading than the interaction that led to the child's initial disclosure, the diagnosticity of the evaluation interview increases. Some suggestibility research documents that errors in response to suggestive questions may persevere or increase over multiple interviews, but this is far from inevitable; these effects emerge when each interview is itself suggestive and highly leading questions are asked (LaRooy, Lamb, & Pipe, 2009; Goodman & Quas, 2008). A large component of suggestibility is due to errors in the child's responses rather than impairment of the child's memory, which means that nonleading questions will not elicit repeated errors. On the

Downloaded by [University of Arizona] at 05:45 15 January 2014

other hand, repeated open-ended interviews are likely to reveal a great deal of new and accurate information (LaRooy et al., 2009). Furthermore, the evaluation interview need not yield any new information to be valuable. The interview is likely to be what Schum and Martin (1982) called *corroboratively redundant*. The redundant report is corroborative of abuse because it reduces worries regarding the validity of the initial disclosure.

INTERVIEWING METHODS FOR INCREASING THE DIAGNOSTICITY OF ABUSE DISCLOSURES

Although the bulk of research on children's suggestibility has emphasized the ways in which children's accuracy can be impaired, a growing number of studies have explored means of increasing the accuracy and completeness of children's reports. These techniques will increase the diagnosticity of disclosures of abuse to the extent that they increase the likelihood that abused children will disclose (true positives) and decrease the likelihood of false disclosures (false positives).

Interviewers who begin with interview instructions can increase the accuracy of children's reports. It is recommended that interviewers (a) teach children that they can say "I don't know" (Cordon, Saetermoe, & Goodman, 2005) because children may be reluctant to respond "I don't know" to yes–no questions (Poole & Lindsay, 2001) or specific wh- questions (e.g., "What color was his hat?) (e.g., Memon & Vartoukian, 1996); (b) teach children that they can say "I don't understand" (Saywitz, Snyder, & Nathanson, 1999; Peters & Nunez, 1999) because children frequently fail to ask for clarification (Carter, Bottoms, & Levine, 1996; Perry, McAuliff, Tan, & Claycomb, 1995; Saywitz, et al., 1999); and (c) teach children that they can correct the interviewer (Saywitz & Moan-Hardie, 1994; Warren, Hulse-Trotter, & Tubbs, 1991). Interviewers are encouraged to not only explain the rules of the interview but also to provide children examples, because an unelaborated instruction (e.g., "It's okay to say I don't know") is unlikely to be effective (Geddie, Fradin, & Beer, 2000; Memon & Vartoukian, 1996; Moston, 1987). Furthermore, interviewers should reinforce giving an answer when one can so that children do not overutilize the "don't know" or "don't understand" responses (Gee, Gregory, & Pipe, 1999; Saywitz & Moan-Hardie, 1994). Eliciting a promise to tell the truth from children has been found to increase honesty without increasing errors (Evans & Lee, 2010, Lyon & Dorado, 2008; Talwar, Lee, Bala, & Lindsay, 2002, 2004), even among maltreated children who have been coached to either falsely deny or falsely claim that events occurred (Lyon, Malloy, Quas, & Talwar, 2008).

Interviewers who elicit abuse details through open-ended questions increase accuracy. As noted, children's free recall reports of genital touch are much more probative than "yes" responses to closed-ended questions.

Downloaded by [University of Arizona] at 05:45 15 January 2014

In general, children's free recall reports are much more accurate than their responses to recognition questions (Dale, Loftus, & Rathbun, 1978; Dent, 1982, 1986; Dent & Stephenson, 1979; Goodman & Aman, 1990; Goodman, Bottoms, Schwartz-Kenney, & Rudy, 1991; Hutcheson, Baxter, Telfer, & Warden, 1995; Oates & Shrimpton, 1991; Ornstein, Gordon & Larus, 1992).

When children are properly questioned, questions tapping free recall can elicit surprisingly large amounts of information. In forensic interviews, children's responses to free recall prompts elicited three to five times more information than responses to more focused prompts (Lamb, Hershkowitz, Orbach, & Esplin, 2008). The key is to provide the child some guidance. Laboratory studies demonstrating very poor recall performance among younger children led some commentators to conclude that direct questions may be necessary in order to elicit abuse details (Lyon, 1999), but the limitation of those studies is that free recall questions did little more than ask "What happened?" "What else happened?" and "Tell me more." In contrast, an interviewer who asks follow-up questions in the form of "You said x; tell me more about x" or "You said x; what happened next?" elicits free recall while providing the child needed scaffolding.

Children are not accustomed to being asked open-ended questions, and they benefit from practice. A useful tool is narrative practice in which the interviewer asks the child about an innocuous event before moving to the allegation. For example, the interviewer asks the child to "Tell me everything that happened" on the child's last birthday, and seeks elaboration through "You said x, what happened next" and "You said x; tell me more about x" questions. In the field, Sternberg and colleagues (1997) found that when interviewers used narrative practice during rapport-building with open-ended rather than closed-ended questions, children provided longer and richer responses to the first substantive question about abuse, and longer responses to free recall questions throughout the interview. Laboratory research has demonstrated that children's responses are also more accurate when narrative practice is provided before the substantive interview (Roberts, Lamb, & Sternberg, 2004).

Narrative practice can profitably be combined with nonsuggestive forms of interviewer encouragement and support. Addressing the child by his or her name and providing noncontingent reinforcement (e.g., "You really help me understand") is related to greater elaboration by the child (Hershkowitz, 2009). Another form of reinforcement is the use of facilitators, also known as back-channel responses, in which the interviewer encourages additional information through simple utterances that communicate that the interviewer is listening without taking the floor (e.g., "uh huh," "okay"; Cautilli, Riley-Tillman, Axelrod, & Hineline, 2005). Laboratory research has also found that interviewer social support, such as smiling often, using warm vocal intonations, and sitting in close proximity helps children resist misleading questions (Davis & Bottoms, 2002).

Downloaded by [University of Arizona] at 05:45 15 January 2014

Reinforcement is noncontingent to the extent that the interviewer does not *selectively reinforce* desired responses. Contingent reinforcement, on the other hand, can be highly leading, particularly when the interviewer overtly praises acquiescence to yes–no questions (Garven, Wood, Malpass, & Shaw, 1998, 2000). Analogously, when examining child interviews regarding a mild transgression, general reassurance about disclosing negative information increased children's true disclosures (Lyon et al. 2008), but reassurance that specifically mentioned the transgression increased false positives (Lyon & Dorado, 2008).

We recommend that interviewers encountering reluctance inform the child that "it is really important that I know everything that happened," which is less leading than providing specific and potentially leading reasons for disclosure (e.g., "It is really important that you tell me so I can put the suspect in jail"). Simply stating that "it is really important" is analogous to "placebic" requests that increase compliance without providing explicit justification (e.g., "May I use the Xerox machine because I have to make copies?" is more effective than "May I use the Xerox machine?"; Langer, Blank, & Chanowitz, 1978).

Interviewers should strive to ask children to describe individual abusive events (Lamb et al., 2008). When children narrate individual events, they are less likely to provide skeletal and generic reports and more likely to disclose idiosyncratic details, such as interruptions (e.g., the perpetrator stopped because a parent was heard coming home). Idiosyncratic details are harder to attribute to some sort of adult coaching. On the other hand, interviewers should avoid asking children to estimate the number of times abuse occurred or temporal information about the abuse, because this is likely to make true reports more difficult to distinguish from false reports. Children have difficulty providing numerical estimates in general. Children may also have particular difficulty with time and number if they are asked to recall incidents that occurred on multiple occasions over a long period of time (Lyon & Saywitz, 2006). Moreover, children's responses to direct questions about time and number are likely to be cursory whether their report is true or false.

It is also useful to ask children what occurred before and after the abusive contact. An adult who suggestively questions a child or who coaches the child to make a false report is likely to focus on eliciting the abusive act itself rather than the context in which the abuse occurred. Children who experienced sexual abuse are frequently able to provide information about the perpetrator's preparatory actions (e.g., closing the bedroom door), the immediate after-effects of the abuse (e.g., washing up, difficulty in falling asleep), and the perpetrator's efforts to conceal what occurred (e.g., telling the child not to tell). Interviewers can also ask more direct questions about what the child and the suspect did following the abuse (e.g., "What did he do after [the touching]?" "What did you do after he left?") and what children

Downloaded by [University of Arizona] at 05:45 15 January 2014

thought or felt about the incident (e.g., "How did you feel when he x?" "What did you think when he x?"). Children who report multiple episodes of abuse can be asked about their thoughts and feelings during early and later abuse incidents. This may elicit children's naïveté regarding the perpetrator's intentions initially and their fearful expectations on later occasions. Children are often remarkably articulate with respect to describing their emotional and physical reactions to abuse (Lyon, Scurich, Choi, Handmaker, & Blank, in press).

Children often report that genital touching occurred during caretaking, such as toileting or bathing, which may make it difficult to distinguish between abusive and innocuous touch. In addition to asking questions about what the suspect said about the touching, which might support sexual intent, the interviewer can ask the child about what occurs when others bathe or care for the child, which may enable the interviewer to determine if the suspect's actions were innocent or abusive.

Interviewers can better understand the social and emotional pressures on the child by inquiring into the child's prior disclosures and the child's reasons for disclosing (or for not disclosing sooner; Hershkowitz, Lanes, & Lamb, 2007). Information about the disclosure can be elicited by continuing to ask "what happened next" questions until children report telling another person (Hershkowitz et al., 2007), or the interviewer can ask the child in a more focused manner "Who did you tell?" and "What did you say to them?" The interviewer should also ask the child about the disclosure recipient's reactions ("What did she do/say when you told her?") and what the disclosure recipient has told the child about talking to the interviewer ("What did your mom tell you about talking to me?"). The interviewer should also ask the child what the parent and other interested adults have said about the alleged perpetrator ("What has x said about y?").

The responses of the people to whom the child disclosed are very important. Children, particularly young children, are likely to disclose abuse first to a parent (Kogan, 2004; Hershkowitz et al., 2007). Children are less likely to disclose and more likely to recant when nonoffending parents refuse to believe that abuse occurred or otherwise fail to support the allegation (Lawson & Chaffin, 1992; Malloy, Lyon, & Quas, 2007). On the other hand, children's reports are often doubted because of the assertion that a parent is influencing the child to make a false claim of abuse. Hence, the parent's reaction can play an important role in determining if the child's report is consistent over time. The interviewer can also ask the child what the perpetrator has said about others (including caregivers) and about the abuse, as this may reveal threats and other inducements to keep the abuse a secret and thus help to explain delays and inconsistencies in the child's report.

In order to distinguish between events the child reports because he or she has heard about events from others and events the child has actually experienced, the interviewer can ask the child "how do you know"

Downloaded by [University of Arizona] at 05:45 15 January 2014

questions. These are called "source monitoring" questions. Although these questions are very difficult for young preschoolers, who have limited abilities to identify the source of their beliefs (Gopnik & Graf, 1988), older children can report whether they actually saw or merely heard about events. Indeed, the Sam Stone study (Leichtman & Ceci, 1995), which documented high rates of false reports in preschoolers exposed to both repeated suggestive questioning about and negative stereotyping of a stranger who briefly visited the classroom, found that the rate of false reports decreased from 72% to 44% among three- to four- year olds and from approximately 30% to 10% among five- to six- year olds when children were asked "Did you *see* Sam Stone rip the book?" and "Did you *see* Sam Stone soil the teddy bear?" (see also Poole & Lindsay, 2001).

Some source monitoring questions may be less effective, such as questioning whether a recipient "told" the child what happened (Bruck, Melnyk, & Ceci, 2000). Young children often confuse "tell" and "ask" (Walker, 1999) and may not understand the implication of a question such as "Did your mother tell you what to say?" The question implies that the mother coached the child, but speaking literally, a mother who tells the child to "tell the truth" or "tell them what really happened" is telling the child "what to say." Moreover, "how do you know" questions are preferable to "did you see" questions insofar as the latter are yes–no questions and therefore less likely to elicit less accurate responses.

All of these inquiries enable the interviewer to test alternative hypotheses regarding the child's actual experiences. Asking about others' reactions, for example, helps the interviewer explore the possibility that the child is either alleging or denying abuse because of pressures from adults close to the child. Suggestibility researchers have argued that "[w]hen an interviewer avoids confirmatory biases by posing and testing alternative hypotheses, the suggestive techniques do not seem to result in as many serious problems" (Bruck, Ceci, & Melnyk, 1997, p. 304). We would caution, however, that the questions designed to test alternative hypotheses should themselves be worded in as open-ended a fashion as possible (e.g., "Tell me about a time your mom gave you a bath" and "What did your mom say about talking to the police?"). Questions that are sometimes recommended to test one's hypotheses about the child's abuse report (e.g., "Who else beside your teacher touched your private parts? Did your mommy touch them, too?"; Bruck, Ceci, & Hembrooke, 1998) risk eliciting inaccurate responses because of their directness.

The most widely researched guideline for conducting child interviews is the NICHD investigative interview protocol (Lamb et al., 2008). Under the NICHD protocol, child interviewers administer scripted interview instructions, rapport building questions, and nonleading allegation questions. International research testing the NICHD protocol demonstrates its

Downloaded by [University of Arizona] at 05:45 15 January 2014

superiority to nonprotocol interviews (Lamb et al., 2008). For example, interviewers using the NICHD protocol use at least three times more open-ended and approximately half as many option-posing and suggestive prompts as they do without the protocol, considering comparable incidents involving children of the same age (Lamb et al., 2008).

The Ten-Step interview (Lyon, 2005) is a modified and simplified version of the NICHD protocol. Because of research warning that children may overuse responses they are taught are acceptable (such as the "I don't know" response; Gee et al., 1999; Saywitz & Moan-Hardie, 1994), the Ten-Step includes a counterexample for each instruction. For example, the "I don't know" instruction reinforces both a "don't know" response (to "What is my dog's name?") and a responsive answer (to "Do you have a dog?"). Because of research demonstrating that eliciting a promise to tell the truth from children increases children's honesty (Evans & Lee, 2010; Lyon & Dorado, 2008; Lyon et al., 2008; Talwar et al., 2002, 2004), the Ten-Step includes the promise.

In sum, children's disclosures of sexual abuse can be highly probative evidence that abuse occurred. Laboratory research supports the conclusion that if a child discloses genital touching in response to free recall questions about interactions with an individual, this is strong evidence that such touching occurred. Even "yes" responses to recognition questions that would be considered inappropriately leading by many researchers (such as questioning with the assistance of an anatomically correct doll) are often strong evidence, but the strength of the evidence diminishes as the suggestiveness increases and the age of the child decreases. Laboratory research also supports the conclusion that interviewers can elicit more accurate reports from children with the use of open-ended questions, interview instructions (including a promise to tell the truth), and narrative practice rapport-building. Interview protocols that incorporate these elements are therefore likely to produce more accurate reports, in particular by reducing the likelihood that a false allegation will either be created or perpetuated by poor interviewing.

Downloaded by [University of Arizona] at 05:45 15 January 2014

DIFFERENCES BETWEEN OUR ANALYSIS AND HERMAN (2009)

Our conclusions are quite different than one put forward by Herman (2009), who argues that "[f]alse positive error rates in forensic interviews are too high for these interviews to be used as the basis for making validity judgments about children's reports of CSA" (p. 261). Herman (2009) describes a study by Hershkowitz and colleagues (2007) that found that although trained evaluators exhibited "fairly high rates of overall accuracy compared with the accuracy of clinical judgements in other areas of psychology and medicine" when considering true and false reports elicited using the NICHD protocol,

they were unable to discriminate between true and false reports elicited with nonprotocol interviews (p. 253). Moreover, for both the NICHD protocol and the nonprotocol interviews, evaluators judged a large percentage of the false disclosures as true.

The problem with Herman's analysis is that the study by Hershkowitz and colleagues (2007) did *not* assess the probative value of disclosures of sexual abuse. Rather, they looked at whether true disclosures could be distinguished from false disclosures. This is an important issue, but it begs the question of whether disclosures increase the likelihood that abuse occurred. The Hershkowitz and colleagues (2007) study *only* examined cases in which a child disclosed sexual abuse. It then selected equal numbers of true and false disclosures for both the NICHD protocol and nonprotocol interviews. In order to assess the probative value of a disclosure, however, it is important to recognize that a disclosure is itself highly probative of abuse and that good interviewing is less likely to elicit false disclosures than bad interviewing.

An analogy might help to make this point more clear. Imagine someone claimed that DNA tests are worthless because true matches look the same as false matches. In true matches, the perpetrator really did contribute the DNA sample found at the scene of the crime. In false matches, the perpetrator didn't contribute the DNA sample found at the scene of the crime, but the test result is positive; perhaps there was a lab error (for example, we don't have the perpetrator's actual DNA) or perhaps we have a random match (in which someone with DNA identical to the perpetrator actually contributed the DNA sample found at the scene of the crime). If the study compared true matches to false matches, they would be very difficult (perhaps impossible) to distinguish. Indeed, if they are both actual matches, they should look identical. It would be wrong, however, to conclude that a DNA match means nothing. Rather, one would ask what is the likelihood of a true match compared to a false match? As long as true matches are much more likely than false matches, then a match is highly probative evidence.

Imagine a second DNA study was conducted following the implementation of improved procedures that ensured that the only false matches are random matches. But again, if the study compared true matches to false matches, they would be difficult, if not impossible, to distinguish. It would be wrong, however, to conclude that improved procedures are ineffective. Rather, one would ask whether the number of true matches compared to the number of false matches increased, thus increasing the probative value of a match.

A disclosure of sexual abuse is analogous to a match. A true disclosure is like a true match. A false disclosure is like a false match. There is good evidence that disclosure of sexual abuse is much more likely to be true than false and that improving interviewing can reduce the likelihood of false disclosures. But these virtues are not detectable if one only compares true and false disclosures.

Downloaded by [University of Arizona] at 05:45 15 January 2014

A second problem with using the Hershkowitz and colleagues (2007) study to argue that child interviews are not helpful is that the determination of whether sexual abuse occurred is not solely based on the word of the child. Herman (2009) is careful to make this point with respect to the existence of corroborative evidence, including perpetrator confessions, medical evidence, and eyewitness reports. However, he argues that unless this kind of corroborative evidence (which he calls "hard" evidence) is present, a child's disclosure (which he calls "soft" evidence) cannot suffice to substantiate an abuse report.

In a supplemental technical document (available on the Internet), Herman (2008) also acknowledges that evaluators

> might have interviews with the alleged perpetrator and other parties involved in the case, data about the context and manner in which the concern about possible sexual abuse first arose, data regarding the number of people who had talked with the child about suspected abuse before the recorded interviews occurred, as well as other case history information. It is possible that access to this additional information could improve decision accuracy across all cases. . .

Herman (2009) groups this type of information with the child's disclosure, thus categorizing it all as "soft" (p. 247). However, this is precisely the kind of information that can enable an interviewer to determine the risk that the child's prior and current disclosure are the product of suggestion, coercion, or insincerity. Moreover, as we have emphasized, the child interview itself should inquire into the child's initial disclosure and the pressures that have been placed on the child to disclose or not to disclose.

DISCUSSION

Bayesian approaches, if correctly understood, can help us understand the probative value of evidence. Children's disclosures of genital touch often constitute strong evidence that touching did in fact occur. If a grade-school child recalls genital touch in response to free recall questions, this is particularly strong evidence. More suggestive questions asked of younger children have less probative value.

Happily, researchers have moved beyond identifying methods that undermine children's accuracy and have developed positive prescriptions for effective interviewing. Interviews that utilize the tools incorporated into the NICHD structured interview protocol, including instructions, narrative practice rapport building, and open-ended questions regarding abuse, will lead to more accurate and complete reports. Furthermore, these methods can be further improved through the use of instructions with counterexamples and a promise to tell the truth (e.g., in the Ten-Step interview). Interviewers

Downloaded by [University of Arizona] at 05:45 15 January 2014

may elicit information from the child that helps the interviewer assess the likelihood that the child's report has been distorted by others, including the suspect and prior recipients of the child's disclosure.

NOTES

1. The reader might want to work through some more examples. Ten percent is equivalent to 10: (100–10), or 10:90, or 1:9. Twenty-fiver percent is equivalent to 25: (100–25) or 25:75 or 1:3. Seventy-five percent is equivalent to 75: (100–75) or 75:25, or 3:1.

2. To convert odds back to percentages we divide the first number in the odds ratio (20) by the sum of the first and the second numbers in the odds ratio (20 + 19). That is, we divide 20 by (20 + 19), which is 20/39, or 51%. Again, the reader will want to try some more examples. 1:1 converts to 50% because 1 divided by (1+1) is 1/2 or 50%. 2:1 converts to 67% because 2 divided by (2 + 1) is 2/3 or 67%. 100:1 converts to 99% because 100 divided by (100+1) is 99/100 or 99%.

3. The false positive rate is the inverse of the true negative rate; their sum equals 100%. If the child is not abused and the evidence is absent, this is a true negative. If the child is not abused and the evidence is present, this is a false positive. Hence, the likelihood that the evidence is present plus the likelihood that the evidence is absent will be 100%. If the true negative rate is 75%, the false positive rate is 25%.

REFERENCES

Bolstad, W. M. (2007). *Introduction to Bayesian statistics*. Hoboken, NJ: Wiley.

Bridges, A., Faust, D., & Ahern, D. (2009). Methods for the identification of sexually abused children. In K. Kuehnle & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 21–47). Hoboken, NJ: John Wiley & Sons.

Brown, D., & Lamb, M. E. (2009). Forensic interviews with children: A two-way street: Supporting interviewers in adhering to best practice recommendations and enhancing children's capabilities in forensic interviews. In K. Kuehnle & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 299–325). Hoboken, NJ: John Wiley & Sons.

Bruck, M., & Ceci, S. J. (1996). Issues in the scientific validation of interviews with young children. *Monographs of the Society for Research in Child Development*, *61*, 204–214 (4–5, serial no. 248).

Bruck, M., & Ceci, S. J. (1999). The suggestibility of children's memory. *Annual Review of Psychology*, *50*, 419–439.

Bruck, M., & Ceci, S. (2004). Forensic developmental psychology: Unveiling four common misconceptions. *Current Directions in Psychological Science*, *13*, 229–232.

Bruck, M., Ceci, S. J., & Francouer, E. (2000). Children's use of anatomically detailed dolls to report genital touching in a medical examination: Developmental and gender comparisons. *Journal of Experimental Psychology: Applied*, *6*, 74–83.

Bruck, M., Ceci, S. J., Francouer, E., & Renick, A. (1995). Anatomically detailed dolls do not facilitate preschoolers' reports of a pediatric examination involving genital touching. *Journal of Experimental Psychology: Applied*, *1*, 95–109.

Downloaded by [University of Arizona] at 05:45 15 January 2014

Bruck, M., Ceci, S. J., & Hembrooke, H. (1998). Reliability and credibility of young children's reports. *American Psychologist*, *53*, 136–151.

Bruck, M., Ceci, S. J., & Melnyk, L. (1997). External and internal sources of variation in the creation of false reports in children. *Learning and Individual Differences*, *9*, 289–316.

Carter, C., Bottoms, B. L., & Levine, M. (1996). Linguistic and socio-emotional influences on the accuracy of children's reports. *Law and Human Behavior*, *20*, 335–358.

Cautilli, J. D., Riley-Tillman, T. C., Axelrod, S., & Hineline, P. (2005). The role of verbal conditioning in third generation behavior therapy. *The Behavior Analyst Today*, *6*, 137–146.

Cordon, I. M., Saetermoe, C. L., & Goodman, G. S. (2005). Facilitating children's accurate responses: Conversational rules and interview style. *Applied Cognitive Psychology*, *19*, 249–266.

Dale, P. S., Loftus, E. F., & Rathbun, L. (1978). The influence of the form of the question on the eyewitness testimony of preschool children. *Journal of Psycholinguistic Research*, *74*, 269–277.

Davis, S. L., & Bottoms, B. L. (2002). Effects of social support on children's eyewitness reports: A test of the underlying mechanism. *Law and Human Behavior*, *26*, 185–215.

Dent, H. R. (1982). The effects of interviewing strategies on the results of interviews with child witnesses. In A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials* (pp. 279–297). Stockholm, Sweden: Norstedt.

Dent, H. R. (1986). Experimental study of the effectiveness of different techniques of questioning child witnesses. *British Journal of Social and Clinical Psychology*, *18*, 41–51.

Dent, H. R., & Stephenson, G. M. (1979). An experimental study of the effectiveness of different techniques of questioning child witnesses. *British Journal of Social and Clinical Psychology*, *18*, 41–51.

Evans, A., & Lee, K. (2010). Promising to tell the truth makes 8- to 16-year-olds more honest. *Behavioral Sciences & the Law*, *28*, 801–811.

Evett, I. W. (1995). Avoiding the transposed conditional. *Science & Justice*, *35*, 127–131.

Faust, D., Bridges, A., & Ahern, D. (2009a). Methods for the identification of sexually abused children: Issues and needed features for abuse indicators. In K. Kuehnle & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 3–19). Hoboken, NJ: John Wiley & Sons.

Faust, D., Bridges, A., & Ahern, D. (2009b). Methods for the identification of sexually abused children: Suggestions for clinical work and research. In K. Kuehnle & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 49–66). Hoboken, NJ: John Wiley & Sons.

Finnila, K., Mahlberga, N., Santtilaa, P., & Niemib, P. (2003). Validity of a test of children's suggestibility for predicting responses to two interview situations differing in degree of suggestiveness. *Journal of Experimental Child Psychology*, *85*, 32–49.

Downloaded by [University of Arizona] at 05:45 15 January 2014

Garven, S., Wood, J. M., Malpass, R. S., & Shaw, J. S. (1998). More than suggestion: The effect of interviewing techniques from the McMartin preschool case. *Journal of Applied Psychology*, *83*, 347–359.

Garven, S., Wood, J. M., Malpass, R. S., & Shaw, J. S. (2000). Allegations of wrongdoing: The effect of reinforcement on children's mundane and fantastic claims. *Journal of Applied Psychology*, *85*, 38–49.

Geddie, L., Fradin, S., & Beer, J. (2000). Child characteristics which impact accuracy of recall and suggestibility in preschoolers: Is age the best predictor? *Child Abuse and Neglect*, *24*, 223–235.

Gee, S., Gregory, M., & Pipe, M. E. (1999). "What colour is your pet dinosaur?" The impact of preinterview training and question type on children's answers. *Legal & Criminological Psychology*, *4*, 111–128.

Goodman, G. S., & Aman, C. (1990). Children's use of anatomically detailed dolls to recount an event. *Child Development*, *61*, 1859–1871.

Goodman, G. S., Bottoms, B. L., Schwartz-Kenney, B. M., & Rudy, L. (1991). Children's testimony about a stressful event: Improving children's reports. *Journal of Narrative and Life History*, *1*, 69–99.

Goodman, G. S., & Quas, J. A. (2008). Repeated interviews and children's memory. It's more than just how many. *Current Directions in Psychological Science*, *17*, 386–390.

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development*, *59*, 1366–1371.

Heger, A., Ticson, L., Velasquez, O., & Bernier, R. (2002). Children referred for possible sexual abuse: Medical findings in 2384 children. *Child Abuse & Neglect*, *26*, 645–659.

Herman, S. (2008). *Kuehnle/Connell book chapter supplement*. Retrieved from http://herman-research.blogspot.com/2008/03/next-two-subsections-roc-analysis-and_6775.html

Herman, S. (2009). Forensic child sexual abuse evaluations: Accuracy, ethics, and admissibility. In K. Kuehnle & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 247–266). Hoboken, NJ: John Wiley & Sons.

Hershkowitz, I. (2009). Socioemotional factors in child sexual abuse investigations. *Child Maltreatment*, *14*, 172–181.

Hershkowitz, I., Lanes, O., & Lamb, M. E. (2007). Exploring the disclosure of child sexual abuse with alleged victims and their parents. *Child Abuse & Neglect*, *31*, 111–123.

Hutcheson, G. D., Baxter, J. S., Telfer, K., & Warden, D. (1995). Child witness statement quality: Question type and errors of omission. *Law and Human Behavio*r, 19, 631–648.

Kaye, D. H., & Koehler, J. J. (1991). Can jurors understand probabilistic evidence? *Journal of the Royal Statistical Society A*, *154*, 75–81.

Kogan, S. M. (2004). Disclosing unwanted sexual experiences: Results from a national sample of adolescent women. *Child Abuse & Neglect*, *24*, 147–165.

Kuehnle, K., & Connell, M. (2009). *The evaluation of child sexual abuse allegations*. New York: Wiley.

Downloaded by [University of Arizona] at 05:45 15 January 2014

Lamb, M. E., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2008). *Tell me what happened: Structured investigative interviews of child victims and witnesses*. West Sussex, England: Wiley.

Langer, E., Blank, A., & Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: The role of placebic information in interpersonal interaction. *Journal of Personality and Social Psychology*, *36*, 635–642.

LaRooy, D., Lamb, M. E., & Pipe, M. E. (2009). Repeated Interviewing: A critical evaluation of the risks and potential benefits. In K. Kuehnle & M. Connell (Eds.), *Critical issues in child sexual abuse assessment* (pp. 327–361). Hoboken, NJ: John Wiley & Sons.

Lawson, L., & Chaffin, M. (1992). False negatives in sexual abuse disclosure interviews: Incidence and influence of caretaker's belief in abuse in cases of accidental abuse discovery by diagnosis of STD. *Journal of Interpersonal Violence*, *7*, 532–542.

Leichtman, M. D. & Ceci, S. J. (1995). The effects of stereotypes and suggestions on preschoolers' reports. *Developmental Psychology*, *3*, 568–578.

Lyon, T. D. (1999). The new wave of suggestibility research: A critique. *Cornell Law Review*, *84*, 1004–1087.

Lyon, T. D. (2005). *Ten step investigative interview*. Los Angeles: Author. Retrieved from http://works.bepress.com/thomaslyon/5/

Lyon, T. D., & Dorado, J. S. (2008). Truth induction in young maltreated children: The effects of oath-taking and reassurance on true and false disclosures. *Child Abuse & Neglect*, *32*, 738–748.

Lyon, T. D., & Koehler, J. J. (1996). The relevance ratio: Evaluating the probative value of expert testimony in child sexual abuse cases. *Cornell Law Review*, *82*, 43–78.

Lyon, T. D., Lamb, M. E., & Myers, J. (2009). Legal and psychological support for the NICHD interviewing protocol. *Child Abuse & Neglect*, *33*, 71–74.

Lyon, T. D., Malloy, L. C., Quas, J. A., & Talwar, V. (2008). Coaching, truth induction, and young maltreated children's false allegations and false denials. *Child Development*, *79*, 914–929.

Lyon, T. D., & Saywitz, K. J. (2006). From post-mortem to preventive medicine: Next steps for research on child witnesses. *Journal of Social Issues*, *62*, 833–861.

Lyon, T. D., Scurich, N., Choi, K. Handmaker, S., & Blank, R. (in press). "How did you feel?" Increasing child sexual abuse witnesses' production of evaluative information. *Law & Human Behavior*.

Malloy, L. C., Lyon, T. D., & Quas, J. A. (2007). Filial dependency and recantation of child sexual abuse allegations. *Journal of the American Academy of Child and Adolescent Psychiatry*, *46*, 162–170.

Memon, A. & Vartoukian, R. (1996). The effects of repeated questioning on young children's eyewitness testimony. *British Journal of Psychology*, *87*, 403–415.

Moston, S. (1987). The suggestibility of children in interview studies. *First Language*, *7*, 67–78.

Myers, J. E. B. (2005). *Myers on evidence in child, domestic, and elder abuse cases*. New York: Aspen Publishers.

Oates, K., & Shrimpton, S. (1991). Children's memories for stressful and non-stressful events. *Medical Science and Law*, *31*, 4–10.

Downloaded by [University of Arizona] at 05:45 15 January 2014

Ornstein, P. A., Gordon, B. N., & Larus, D. M., (1992). Children's memory for a personally experienced event: Implications for testimony. *Applied Cognitive Psychology*, *6*, 49–60.

Perry, N., McAuliff, B., Tan, P., & Claycomb, C. (1995). When lawyers serve children: Is justice served? *Law and Human Behavior*, *19*, 609–629.

Peters, W. W., & Nunez, N. (1999). Complex language and comprehension monitoring: Teaching child witnesses to recognize linguistic confusion. *Journal of Applied Psychology*, *84*, 661–669.

Pipe, M. E., & Salmon, K. (2009). Dolls, drawing, body diagrams, and other props: Role of props in investigative interviews. In K. Kuehnle & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 365–395). Hoboken, NJ: John Wiley & Sons.

Poole, D. A., & Lindsay, D. S. (2001). Children's eyewitness reports after exposure to misinformation from parents. *Journal of Experimental Psychology: Applied*, *7*, 27–50.

Poole, D. A. & Wolfe, M. A. (2009). Child development: Normative sexual and nonsexual behaviors that may be confused with symptoms of sexual abuse. In K. Kuehnle & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 101–128). Hoboken, NJ: John Wiley & Sons.

Roberts, K. P., Lamb, M. E., & Sternberg, K. J. (2004). The effects of rapport-building style on children's reports of a staged event. *Applied Cognitive Psychology*, *18*, 189–202.

Saywitz, K. J., Goodman, G. S., Nicholas, G., & Moan, S. (1991). Children's memory of a physical examination involving genital touch: Implications for reports of child sexual abuse. *Journal of Consulting and Clinical Psychology*, *5*, 682–691.

Saywitz, K. J., & Moan-Hardie, S. (1994). Reducing the potential for distortion of childhood memories. *Consciousness and Cognition*, *3*, 257–293.

Saywitz, K. J., Snyder, L., & Nathanson, R. (1999). Facilitating the communicative competence of the child witness. *Applied Developmental Science*, *3*, 58–68.

Schum, D. A., & Martin, A. W. (1982). Formal and empirical research on cascaded inference in jurisprudence. *Law & Society Review*, *17*, 105–152.

Sternberg, K. J., Lamb, M. E., Hershkowitz, I., Yudilevitch, L., Orbach, Y., Esplin, P. W., et al. (1997). Effects of introductory style on children's abilities to describe experiences of sexual abuse. *Child Abuse and Neglect*, *21*, 1133–1146.

Steward, M. S., Steward, D. S., Farquhar, L., Myers, J. E. B., Reinhart, M.,Welker, J., et al. (1996). Interviewing young children about body touch and handling. *Monographs of the Society for Research in Child Development*, *61* (4–5, Serial No. 248).

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2002). Children's conceptual knowledge of lying and its relation to their actual behaviors: Implications for court competence examinations. *Law & Human Behavior*, *26*, 395–415.

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). Children's lie-telling to conceal a parent's transgression: Legal implications. *Law & Human Behavior*, *28*, 411–435.

Thompson, W. C., & Schumann, E. L. (1987). Interpretation of statistical evidence in criminaltrials: The prosecutors' fallacy and the defense attorney's fallacy. *Law and Human Behavior*, *1*, 167–187.

Downloaded by [University of Arizona] at 05:45 15 January 2014

Walker, A. G. (1999). *Handbook on questioning children: A linguistic perspective* (2nd ed.). Washington, DC: American Bar Association.

Warren, A, Hulse-Trotter, K., & Tubbs, E. (1991). Inducing resistance to suggestibility in children. *Law & Human Behavior*, *15*, 273–285.

Wood, J. M. (1996). Weighing evidence in sexual abuse evaluations: An introduction to Bayes's Theorem. *Child Maltreatment*, *1*, 25–36

AUTHOR NOTES

Thomas D. Lyon, PhD, JD, is the Judge Edward J. and Ruey L. Guirado Chair in Law and Psychology at the University of Southern California and researches child abuse and neglect, child witnesses, and domestic violence. He is the past-president of the American Psychological Association's Section on Child Maltreatment (Division 37) and a former member of the board of directors of the American Professional Society on the Abuse of Children.

Elizabeth C. Ahern, MA, is a PhD candidate in developmental psychology at the University of Southern California. She researches children's disclosure of maltreatment, truth induction methods, and emergent lie-telling ability. She is also a child interviewing specialist and conducts trainings on child interviewing.

Nicholas Scurich, MA, is a PhD candidate in quantitative psychology at the University of Southern California. He studies normative and descriptive models of juridical decision making.

Downloaded by [University of Arizona] at 05:45 15 January 2014