IN THE COURT OF APPEAL OF NEW ZEALAND

CA 120/98

THE QUEEN

v

PETER HUGH McGREGOR ELLIS

AFFIDAVIT OF CONSTANCE J DALENBERG

INDEX

I, **CONSTANCE J DALENBERG**, PH D, of San Diego, California, United States of America, swear as follows

A. INTRODUCTION

Qualifications

1 I completed a master's degree in child clinical psychology, a four-year doctoral training in clinical psychology, and a second four year training in experimental social psychology, all at the University of Denver My final degree, a Ph D in social psychology with additional specialty ("tool") in clinical psychology was completed in 1983 Thus, I am fully trained at the doctoral level in both experimental methodology and in clinical application I find that this dual base of expertise is invaluable in evaluating child abuse claims

2 I have focused on child abuse both clinically and experimentally since the early days of my graduate training As an experimentalist, I have served as the Research Director for the Institute of Child Abuse and Neglect in Denver, Colorado from 1982 to January of 1985, and the Director of the Trauma Research Institute from 1984 to the present I have supervised over 30 trauma-related doctoral dissertations, and have designed and completed over 150 trauma-related research projects and research overviews

3 As an educator, I have taught courses on the treatment of child abuse and the forensic evaluation of child abuse in graduate institutions since 1983, and now head the Forensic Emphasis of the California School of Professional Psychology in San Diego I also have taught over 50 workshops specifically to child abuse investigators and forensic interviewers I am a special consultant to the Center for Child Protection in San Diego, which conducts the majority of child abuse evaluations in San Diego and the adjacent areas

4. My papers on child abuse evaluation and treatment have been published widely I have presented over 100 papers, symposiums and posters to national and international conferences on the subject, and I have been invited onto the

editorial board of the Journal of Child Sexual Abuse, Child Maltreatment, and (as a statistical reviewer) the Journal of Traumatic Stress. I was one of 100 international scholars invited to the NATO Symposium on trauma and memory in Port de Bourgenay, France in 1996; a greater number of my research papers were accepted for presentation at this symposium than any other single researcher. In recognition of my work in the area, I have been asked to write (and have written) the chapters on ethical issues in the treatment and assessment of victims of child abuse for the psycholegal text on Law and Ethics edited by the President of California's Psychological Association. I also have been appointed to the task force that will develop the standards of practice for assessment and treatment of child abuse for California's Board of Control.

5 Much of my research focuses on memory accuracy, short-term and long-term, in the arena of child abuse. My paper on the long-term recovery of memory of child abuse, presented at NATO, was also published both by Plenum Press as part of the NATO summary and Psychiatry and the Law. This study was named by the International Society for the Study of Dissociation as one of the most important papers of 1996, My research on understanding child fantasy and fantastic elements in child abuse allegations was presented at NATO, at three other national and international conferences, and published as a chapter in the Handbook of Interviewing in 1999 (edited by Memon and Bull) and as the featured research in the APS AC Advisor in 1996. In the most prestigious review of child abuse related research for attorneys (by Jon Myers), my research on fantasy is the most prominently cited study.

6. Clinically, I have been treating victims of child abuse since my early graduate school internships in 1975. I have a private clinical practice, and also interview for the county in cases in which very difficult decisions must be made. I testify almost solely on trauma-related matters, and have been accepted as an expert in American courts on child sexual abuse and on false allegations. I am one of few researcher/clinicians who have been specifically accepted by American courts as an expert on the meaning of recantation in child sexual abuse

7. Combining my research, clinical, and forensic experiences, I have evaluated over 1000 alleged victims of child sexual abuse.

Overview

8. I have been asked by counsel for the respondent to review and comment upon the literature reviews and summaries contained in the affidavits submitted by Dr Lamb and Dr Parsonson. I have not been provided with the video interviews or transcripts, and have not been asked to comment on specific interviews.

General Assessment

9 Both of these briefs make many noncontroversial points: that leading questions might affect the outcome of interviews, that memory is not infallible and fades over time, that repeated questions can be problematic in some cases, etc. However, there are a number of conclusions and statements within the briefs that I consider to be misleading or controversial, although this is decidedly more true of the Parsonson brief.

10 The general issues to be discussed further within the present document are the following:

10.1 At times the implication is made that children are qualitatively different than adults on some dimension, eg that children are prone to problems in memory for abuse-related material and that adults are not so affected. It is not stated, and should be stated, that the differences on some dimensions are very small (although quite real). Older adults are also 7-20% poorer at many of these tasks than young adults, but it is unlikely that the testimony of a half dozen adults in their 60's would be discounted due to the fact that statistically they tend to be slightly less accurate than adults in their 20's.

10.2 Relevant contextual information about the studies that leads children to appear more competent is often absent from the briefs, while data that leads to a conclusion that they are incompetent is highlighted. Caveats that should have been offered in order to fairly represent the state of the

literature are not given For instance, Lamb and Parsonson rarely mention that in the studies that show that children are quite suggestible, the children are often told that a trusted adult knows for certain that the "perpetrator" committed the specific act, or that another adult was there at the time and knows that an event occurred This is quite different than merely asking questions in a misleading or biased way

10.3 Certain conclusions are highly exaggerated, with a minority position among scientists presented as truth For instances, the use of props and anatomical dolls is presented as highly problematic when in fact they are quite well accepted (in the United States)

10.4 Important age distinctions are not always made In many studies reported in these briefs, only results for the 2-3 year olds are given (who make errors in the 20-25% range) Lamb and Parsonson do not note that the 5-6 years olds do far better, despite the fact that the Lamb brief states that most children were five at the time they last saw Ellis In reviewing this literature, one cannot make general statements about "children" or "preschoolers" Some risks are applicable only to children under six, some only to children under three, and some to all children

10 5 Fantasy statements and fantastic allegations within testimony are presented in these affidavits as signs of a false allegation, when this is decidedly not the consensus of researchers in the field No evidence is cited for the Parsonson conclusion that bizarre statements are indicative of the overall falsity of the allegation. Evidence that fantastic allegations are related to severity of trauma in known true child abuse cases is ignored Neither author acknowledges the mainstream conclusion that fantastic allegations are to be expected when children are traumatised

10.6 Parsonson's representation in his affidavit that recantation is rare in the true child abuse case is radically different from the mainstream point of view In fact, this affidavit is the single most extreme professional

statement regarding recantation that I have ever read, and decidedly does not represent the most consensual professional view Recantation should not be taken as evidence for either the truth or falsity of the original claim, but it is not rare among false accusing or accurately accusing children

10.7 The documents fail to reflect the complexity of the clinical thought of the field during the 1980's and 1990's Specifically, the affidavits are filled with general conclusions and implications that are unidimensional - for example, open questions are always to be preferred, leading questions are never justified, or that the encouragement to say "I don't know" is always important with young children Considering each point isolated from the others, most of these conclusions are quite justified (although, again, the dangers that are being cited of disobeying the injunctions are exaggerated) However, the real clinical world of child abuse interviewing continually requires tradeoffs in meeting one's goals Do we ask open questions only, and risk getting almost no information from the very young child (leaving the child potentially in danger) or do we ask direct questions and risk the child saying "yes" at times in order to please? If there is intense internal pressure toward silence, as can be true in embarrassing or upsetting child abuse experiences never to press the child to talk, and risk collusion with the silence, or should we occasionally suggest that the child may have a secret, and risk contamination or leading? Do we train a child to say "I don't know", and risk that he or she will use it as an excuse not to answer emotionally difficult questions, or should we not train, and risk that the child will not feel permission to state that he or she does not know the answer to the question at hand

10.8 Lamb and Parsonson are quite right that child abuse interviewers as a group tend to ask fewer open questions, lead more, and prepare the child less than would be effective for the elementary school teacher preparing a child to accurately answer questions for a history test However, they do not consider the question why these knowledgeable professionals might be less persuaded by the research on laboratory

eyewitness interviewing than are non-clinicians Lamb and Parsonson fail to take into account that interviewing regarding trauma in the real world is not identical in all ways to interviewing about a neutral laboratory event The risk/benefit ratio is different

10 9 There is no real risk, for the adults, if a preschool child fails to give adequate detail to a researcher in describing the visit of a clown to his or her preschool class There is a risk to failing to elicit accurate details of a sexual crime against the child - it damages ability to understand and/or prosecute the case, and further endangers this child and others Thus, while a cognitive researcher concentrates virtually solely on decreasing the risk of false or inaccurate disclosure, they do not often consider the risk of the child's inability to disclose at all This is quite appropriate in neutral laboratory settings However, since the child abuse interviewers must keep in mind both sets of risks (the risk of nondisclosure and the risk of inaccurate disclosure), and strive to balance them in cases when they suggest opposing interview strategies, the task is not so clear as critics might suggest Interviewing children about trauma is a balancing act, and the task of minimising errors of commission while simultaneously minimising errors of omission is not so straightforward as one might think I will articulate this point in more detail within specific examples in the document

B. SPECIFIC COMMENTS

The Use of Free Recall v Directed or Leading Questions

11 The two briefs forwarded appear to disagree as to the extent of the leading questions used in the interviews with the creche children Lamb does not find the interviewers to be overly leading However, both individuals state that there should have been greater use of open questions Given the tradeoffs in the use of direct and open questions, however, this conclusion is more controversial than one might think

12. More specifically, imbedded in both documents, although much more fairly presented in the Lamb brief, is the acknowledgment that if free recall and open questions alone are used, very young children give almost no information. In fact, when the interview is being conducted after a delay, it is not unusual for the young child to contribute no information at all in free recall. Taking simply the articles that the Parsonson and Lamb briefs cite, it might be helpful in understanding interviewer behaviour (and failure to use open questions alone) if the reader were made aware of the actual figures reported in these documents.

13 For instance, the Salmon et al 1995 article, used in the Parsonson brief to argue against use of props, was followed up in 1997. The three and five year old children were interviewed about an event after three days (results published in 1995), and then again after one year (results published in 1997). In the analysis of the children's free recall, the authors noted that it was possible for the children to note up to 112 details or features of the event. Three days after the event, the mean number of event details given by the three and five year old children were 2.71 and 16.43 respectively. One year later, the means were 1.12 and 3.46. Thus, the average four year old was reporting one unit of information if asked only open-ended questions. Logically, since 1.12 was a mean, up to half of the children were completely non-disclosing.

14 When the interviewers introduced verbal and visual prompts (props), the four year olds produced an additional 12.16 units of information at the three day point, and 9.34 units of information at the one year mark. For the six year olds, the corresponding figures were 20.67 and 16.56. Further, the new information given at the one year mark was 87% correct for the six year old (five at the time of the last event) and 73% correct for the younger children. Overall, considering the one year data, moving from free recall to prompted recall (or behavioural re-enactment - asking the child to act out the event with props) decreased accuracy by 5% for the age group similar to the children at issue in the Ellis case. It is thus true that open ended questions would have yielded more accurate percentage of information overall. The move away from free recall also increased total units of accurate disclosure by 443%

15. A second example cited in both briefs without much detail is the Saywitz et al (1991) research. Saywitz et al found that without direct questions, 22% of the children (who had been touched in a medical exam) mention genital touch. With a direct question, 86% disclose However, 3-8% in the control condition (where children had not been touched) falsely assert abuse when directly questioned (although again, typically without detail). Therefore, interviewers who use direct questions with young children risk increasing the false allegation rate by 3-8% (in this study - a relatively common finding) and increase the true allegation rate by more than 60%.

16 The conclusion of this research, and others like it, is that one uses open questions when one can, but that it is unrealistic to assume that a large portion of any successful interview with a young child will use free recall. With a child as young as 3-6, it would be highly inappropriate to use only or even predominantly open questions. The research clearly suggests (a) that most children would not disclose true abuse if this method alone were used, (b) that direct questions substantially increase correct disclosure, and (c) that total units of information may double, triple or more if direct questions are included in the interviewing protocol

17 But what of that 3-8% increase in risk? Dr Lamb notes that it is difficult to undo the damage done by leading questions. He states that once contamination has occurred, "it is typically impossible to reverse its effects, and children will be unable - except in rare circumstances unlike those at issue in the Ellis case - to distinguish between details that were "real" and those which were suggested" (paragraph 99). Contrast this with a study that tried to reverse the contamination, using very mild challenge to the implanted material. The study (Leichtman and Ceci, 1995) is cited in both briefs as supportive of the suggestibility of the creche children The five year olds in this study were interviewed four times in a very suggestive manner, eg, "remember the time that Sam Stone visited your classroom and spilled chocolate on that white teddy bear?" Interviewers told the children the false allegation they wanted to implant, making it clear that they knew it to be true Further, the children were told multiple stories of other negative deeds that Sam Stone had committed, supposedly while the interviewer was present. With these highly suggestive

features present, about 35% of children falsely accused Sam Stone of the misdeed when questioned a month later (did anything happen to a teddy bear?). However, when the children was gently asked if they actually saw the event occur, two thirds of the accusers recant. Of the remaining group, another gentle challenge, asking if the child "really" saw the event drops the figure in half again. After the two challenges, 5% maintain their accusation, even in this very extreme scenario It was not "impossible" to reverse contamination.

18. The brief description of the Ricci et al (1996) (Lamb, paragraph 37) study is another instance of presentation of material in a way that is most prejudicial to the child subjects, failing to reflect the complexity of the true findings of the research. Dr Lamb states "Most alarming, as far as its relevance to the Ellis case is concerned, however, is Ricci, Beal, and Dekle's (1996) report that five year old children were more likely to acquiesce to suggestions provided by their parents and were, in fact, most inaccurate when interviewed by them rather than by unfamiliar interviewers (page 27-28)". Note that there are two conclusions here - that children acquiesce more to suggestions by parents and that they are generally more inaccurate when interviewed by parents. There is also an implication that the age of five is of significance.

18.1 The Ricci et al (1996) work is actually a complicated design, in which children are interviewed by parents or strangers about a brief movie and then asked to pick the criminal in the movie out of a lineup. All children in the study were four to six years old, and no age comparisons were made. Thus we know nothing of the relative susceptibility of the five year old from this research.

18.2 The children knew that the mothers had been told about the movie, and approximately 75% of the subjects knew that their mothers had seen the movie. Two studies are reported, one in which the parents were given a scripted set of questions to use, and the other allowing the parent to freely interview the child. Three types of suggestibility are tested, the child's response to suggestive questions by the interviewer, the child's response to gentle confrontations regarding their truthfulness, and the child's response to the interviewer's challenge of

the lineup choice The former, of course, is more relevant to the Ellis case

18.3 Ricci et al do say that they expected to find that children were more suggestible when interviewed by parents overall, but conclude "in the present research children who were interviewed by their parent were not more likely to make errors in response to suggestive questions" (page 497) In fact, one study found equal suggestibility, and the other found less suggestibility when parents asked the questions Thus, the most relevant suggestibility findings in Ricci et al are exactly opposite to Lamb's statements

18 4 The second type of suggestibility, response to challenges to truthfulness, yielded no differences between parents and interviewers In general, children resisted these challenges

18 5 Lamb's conclusion, then, could only refer to one subarea in which parents did produce more suggestibility That is, when their child identified someone in a lineup, and interviewers questioned that identification (whether right or wrong), children changed the identification more often with a parent than they did when interviewed by a stranger Remember, however, that most children know that the parent was present during the movie, and may be deferring to a parent on a specific task (facial identification of an adult) in which they trust their own memories less than those of their parents Thus, Lamb's conclusion regarding suggestibility are non-representative of the full results of the study, and obvious caveats to the generalisability of the results to the Ellis case are not mentioned (eg that the creche children's parents were not present during the alleged molestations) Children were not more suggestible overall when interviewed by parents

18.6 Lamb's second point on this research, that the children were less accurate when interviewed by the parents, is similarly misleading. First, this effect appeared in only one of the two studies Study 2 found equal accuracy Second, there was less than one point difference between the

two groups in Study 1 Third, the inaccurate information could have been on the peripheral versus the central detail Fourth, Ricci et al note that the inaccuracy occurs in the parent condition only when parents used interview techniques that they judged as "non-supportive", and not when parents used a standard interview protocol. It is quite likely that parents in the Ellis did not interview in the skeptical and impatient way that Ricci et al (and others) found to be related to inaccuracy. Again, it is not a fair statement to suggest that parents will elicit less accurate information in general.

19 Trusting the data from the Lamb charts for the moment (Exhibits M and N) (which may not be reasonable), the Ellis interviewers appear to be typical of interviews conducted at the time in the United States and United Kingdom Lamb argues that they differed from "best practice" norms, and this is no doubt true. I have never seen a case in which real world interviews did not differ as a group from some standard of professional perfection. The issue here is whether that standard of perfection is a fair one (that is, should we really be arguing so strongly for open questions - see above) and whether the deviation from standard is a damaging one. I cannot make this judgment, since I have not seen the interviews

20 I question the Lamb chart only in the sense of wondering at the generalisability of the sample. My laboratory was one of the sites that contributed interviews to Dr Lamb, although I do not know which studies specifically did or did not use our sample. In choosing these tapes, Dr Lamb asked that tapes be chosen in which the child was alleging only one incident, and was being interviewed for the first time. If this is the sample used to generate the table, then it is not applicable to instances in which children are alleging multiple incidents of abuse and is not likely to be applicable to second or third interviews

21 Finally, the reader may need a clarification in understanding the Lamb chart fully. "Leading," in Lamb's language on the chart, is not equal to "misleading" in the language of the rest of the Lamb and Parsonson documents. The "risky" or "misleading" questions studied by Ceci, Goodman, and others in the literature review are equivalent to those Lamb calls "suggestive" in his chart.

The New Zealand interviewers asked significantly less of this type of question than did the US and UK samples, and less (perhaps even significantly less - it depends on the number of interviews represented) than did those trained in the "consensus expert guidelines " This speaks extremely well of the interviewers

I Don't Know Training

Not having reviewed the transcripts, I cannot comment on whether the children were prepared in the best possible manner for their interview Some experts now argue that such preparation might ideally include training in the use of "I don't know". Although failure to train the children in use of "I don't know" is presented as a serious error (in the Parsonson brief), it should be acknowledged that many interviewers do not yet incorporate such training in the worldwide trauma community In part, this is due to the fact that training has not been adequately researched, and interviewers fear that permission to say "I don't know" would lead to children shutting down in arenas in which they have incomplete but accurate memories Again, the Parsonson brief ignores this tradeoff

23 Parsonson cites only one study to support his conclusion that the failure to teach I don't know to children was very harmful to testimony This is a recent study by Mulder and Vrij (1996), which did in fact find that such training helped accuracy (and by the way, it helps with adults too) I would add that I know of two other published studies with children, and have conducted studies in my own lab to try to decide the pros and cons of the intervention The two other published studies (Moston, 1987, Vrij and Winkel, 1992) showed no effect on children's accuracy of I don't know training, and my own study showed very complex interactions with age, trauma history, and the nature of the questions being asked This study, conducted by one of my graduate students, Judith Shields, for her dissertation, showed that children with child abuse histories tend to say "I don't know" more than other children (We believe this may be due to the effect of trauma on sense of reality) Training the children (in a careful way over several sessions) did increase the number of I don't know responses but did not reliably increase accuracy It is far too early to make such dogmatic statements about the necessity of the use of these

strategies.Instead, I would simply state that research implies that these preparation techniques are promising and deserve our considered attention.

Repetition

24. Multiple interviews are a fact of life in forensic situations, but do have some risks. In these briefs, the repetition is again presented solely as a negative force, always to be avoided. However, even using the studies cited in the briefs, the conclusions of the research studies are not so dire as Parsonson claims. Poole and White, for instance, conclude that children six years old or older "were able to maintain good within-session consistency" on yes/no questions if asked repeatedly They also conclude that subjects of all ages average "slightly less accurate information" if the questioning continues (page 851). Repetition is also informative for diagnostic purposes. While information given for the first time may be less accurate (but not always, if there are reasons related to shame for the delay in disclosure), information that is repeated across interviews tends to be extremely accurate, even for very young children. If extreme accusations were confabulations dictated by a peculiar combination of props and interviewer suggestions, they often will not be well remembered by the children. Parsonson notes that children interviewed with repeated questions and repeated interviews may change their answers, which is quite true, and that they often become less accurate, which is also true. He fails to note that of the children who changed their answers after Poole and White's two year delay, 90% retracted an accurate accusation rather than introducing a false statement.

Props and Anatomical Dolls

25. Parsonson is very emphatic on the dangers of the use of anatomical dolls He makes some reasonable points, such as the fact that anatomical dolls cannot be used to make a differential diagnosis of sexual abuse, that is, to determine with accuracy (absent the child's verbal report) that a child has been sexually abused. However, he also states that "it is now clear that the use of anatomical dolls is to be actively discouraged" (paragraph 4.38) He lists use of the dolls in the "errors" made by investigators, and concludes (a) that both non-abused and

abused children show increased sexual interest with the dolls, (b) that there is little evidence that anatomical dolls assist in recall and (c) that there is evidence that their use for children under five reduced accurate recall. The conclusions are presented strongly and with exaggerated statements about the nature of the evidence, and the degree to which the professional community rejects use of the dolls There also are errors here in the reporting of results of research

26. Parsonson reports as fact that abused and non-abused children do not differ on their behaviour with the dolls. He cites one study, Mclver, Wakefield and Underwager, 1989, that bears directly on this point. This study is perhaps the most highly criticised of all such studies, and appears in a self-published vanity journal owned and edited by Wakefield and Underwager. Among the problems with the study are (a) the unusually low n, 10 abused children of varying ages, (b) nonmatched groups, eg, children under 3 in one group and not the other, (c) "convenience" sampling, ie, a control group made up of friends, relatives and patients of the researchers, and (d) important methodological confounds, eg, leading questions asked of some children and not others Wakefield and Underwager also have published an interview suggesting that it might be useful to pair children with willing pedophiles to find out if child sexual abuse is actually harmful. To say the least, their conclusions and position are not mainstream.

27. It is true that the Mclver et al, 1989, and a few others (eg Kenyon-Jump et al, 1991) do not find differences among abused and non-abused children However, a number of studies have found such a difference (August and Forman, 1989); DeVoss, 1987; Jampole and Weber, 1987; White et al, 1986). These studies tend to have small n's (eg nine sexually abused children in the Kenyon-Jump study and 10-25 abused children in the others); therefore it is extremely likely that there would be variation in the degree of difference between groups in different studies. The just conclusion here is not that the dolls are useless and damaging, but instead that they are not the powerful aid in diagnosis that we hoped they might be. They do, however, serve other purposes.

28 Again, while it is true that the dolls should not be used as a single diagnostic tool according to consensual opinion, it is not true that they are discounted entirely In 1988, Boat and Everson found that dolls were used by 67% of mental health professionals and 94% of child protection service agencies In a sample of 200 American professionals by Conte et al (1991) anatomical dolls were the most common tool noted, 92% of professionals used the dolls Kendall-Tackett and Watson (1992) report a similar very high figure Most of the relevant American professional organisations have published guidelines supporting use of the dolls for some purposes (eg the American Psychological Association, the American Professional Society on the Abuse of Children, and the American Academy of Child and Adolescent Psychiatry) On this point, I find it extremely puzzling that Parsonson states that there were no published guidelines for interviewing children in 1992 In the specific case of interviewing using anatomical dolls, the American Academy of Child and Adolescent Psychiatry published their guidelines in 1988, the American Psychological Association in 1991, the American Professional Society on the Abuse of Children (the most well-known of the specialty organisations) in 1990, etc Everson and Boat (1994) review over 20 such guideline statements published in the relevant time period for your case I know of no organisation that suggested that the dolls not be used Instead, the controversy is about the nature of appropriate use

29 Very strong statements are made that the use of the dolls increased the danger of abuse misreport No positive value of doll use is offered One study is offered on point by Parsonson He states that three year olds, after being interviewed five minutes after a pediatric exam by Ceci and Bruck, often misreport when the dolls are used "Some, when given a small spoon (a spoon had not been used in the examination) inserted the spoon into the dolls anus or vagina or tapped the doll's genitalia with it" (paragraph 4387, page 28). Such a statement is extremely misleading First, the study itself is inaccurately reported Actually, no child performed this action when asked to demonstrate what the doctor had done with the spoon, although some did incorrectly say that the doctor gave them medicine The behaviour that Parsonson reports occurred when the interviewers asked the child to imagine what the doctor might do with the object

30 Second, the majority of studies on the dolls find that use of these aids helps accuracy and completeness of recall, although a few find no differences and a minority find harm to recall and accuracy Briefly, Gordon et al (1993) found enhanced recall with five year olds (but no difference with three year olds, as did Goodman and Aman (1990). Britton and O'Keefe (1991) found that neither neutral dolls or anatomical dolls increased sexual behaviour in kids, but both were useful in helping the children clarify and express accurately their experience. Leventhal et al (1989) also found increases in detail using the dolls and a doubling of the accurate identification of the "perpetrator" Thus, Parsonson seems to have chosen the single most negative result published in the American literature on use of anatomical dolls, exaggerated the results further, and presented it as the norm. In fact, of the 10 studies reviewed by Aldridge (1998) that assessed the effectiveness of identification of child sexual abuse with anatomical dolls, the Bruck et al study was the only research report with a finding of increased inaccuracy when using the dolls.

31 In summary, in the American literature, one cannot say that the majority of research finds the dolls to be harmful. Instead, the data is mixed - negative on use of the dolls as a single tool for diagnosis, fairly positive on use of the dolls as one tool among many, and positive for use of the dolls as a demonstration or communication aid

32 Parsonson's conclusions regarding use of props is even more confusing He cites one study, Salmon et al (1995), that he states justifies a conclusion that children interviewed with toys make significantly more errors than those interviewed with real items or without props. He fails to mention, even in this study, that:

32.1 Children interviewed with toys made no errors at all in free recall;

32.2 Children interviewed with toys in prompted recall are extremely accurate (over 80% correct recall),

32.3 The toys lessen accuracy significantly in the immediate condition (when children are interviewed three days after the event), but this difference is small (about a 10% decrease);

32.4 A year later, when the children are interviewed again, differences change to favour the prop condition. Children interviewed with toy or real props were about 80% accurate, while the no prop group were 65% accurate.

33 There is in general a large body of literature supporting the positive value of props in interviews with children and adults. Most of us know this to be true intuitively, having had the experiencing of a sudden recall of an experience upon seeing a relevant reminder. The inference that the props are most effective if they are similar to the real item appears to be a valid one, but claims that the existence or use of toy props "contaminates" or "undermines" etc simply cannot be supported. Examples of studies in which props were shown to enhance testimony include Jones and McQuiston (1988), Leventhal et al (1989), Smith et al (1987) and Price and Goodman (1990). In the Saywitz et al (1991) study, the presence of medical props (eg toy stethoscopes or otoscopes) doubled to tripled the number of children able to accurately report a body touch.

34 The bottom line here is that there is mixed evidence on whether dolls and props improve performance overall, but the weight of the data is in that direction. This is particularly true for the 4-6 year old, and less true for older children (who do not need the dolls) and children under three (who cannot use them well as representation of the body). Demonising the use of the dolls and making extreme statements about props, however, as the brief by Parsonson appears to do, is not justified by the evidence.

Fantasy and Source Monitoring

35. The information related 10 fantasy and bizarre accusation is particularly misleading. Dr Lamb, who worked on a project related to mine within our data set, is well aware of my large scale research on fantasy, and cites it in his brief,

although he does not cite my central conclusion, which is contrary to his point. Dr Parsonson appears to be less aware of the research on abuse and fantasy, and cites the bizarre claims of some children as evidence of interviewer misconduct and child incredibility.

36 When Dr Lamb suggests that fantasy in child abuse victims is rare, this is a fair statement in one way but an unfair one in another. In our full samples of children under 17, the rate of fantastic allegations is under 5%. However, the rate goes up to 15% in the sample most relevant to the children that concern the reader here. Since this is a key factor in the argument against the children's credibility in the Parsonson brief, and since most authors (including-Lamb) cite my own research as the most central on-point study in the area, I will describe it more fully.

37 In my sample, 284 children, age 4-9, were selected from among our 6000 tape data base of children making disclosures of child sexual abuse. Half of the cases were those in which we had a strong evidence base for a true finding. In all of these "gold standard" cases, the perpetrator eventually confessed, and the medical evidence was compatible with the testimony. In 80% of the cases other physical or testimonial evidence was given. A group of 152 child matched in severity of alleged trauma, gender, age and race was collected from among those children whose testimony was judged as questionable, and who had no supporting evidence. The group was further divided into a severe and mild trauma sample.

38 Bizarre and impossible detail was most common in children who were known to have been violently abused or severely frightened than in children who suffered known milder trauma, or who may have suffered no trauma at all. In samples in which the degree of pain or the degree of fear was known to have been high, the rates of fantastic elements are up to four times higher than the norm Therefore, violent fantasy production within an abuse allegation is a sign of the truth of the allegation, not the falsity. It should be clearly stated that it is not a definitive sign of the presence of abuse. However, the presence of this type of description should legitimately alert the investigator to the likelihood of some severe abuse history

39. So why do children who have been traumatised make bizarre claims'? In my study, the only large scale evaluation of this subject, fantasy production was not related either to leading questions or to the presence of props, although some children did use the props when generating the fantasy Everson (1997) has a detailed list of dozens of possible explanatory mechanisms, many of which have not been empirically tested Because the presence of these components is often so prejudicial to the listener, it is worth examining a few of the explanations more thoroughly that have been documented by research'

39.1 The experience of trauma is often described by adults as inducing a sense of unreality Clinical writers speak of trauma as the "carrying of an impossible history" I have written of this trauma characteristic in the Handbook of Interviewing (1999), but I am joining virtually every well-known trauma therapist in highlighting this phenomena. Long-term research projects also experimentally demonstrate the sense of unreality and vagueness of memory after known trauma (eg Tromp et al, 1995). Further, depersonalisation and derealisation, two forms of dissociation that relate to feelings that the self (depersonalisation) and the world (derealisation) are not real, are specifically related to trauma history in the research and clinical literature.

39.2 As the real world is becoming less believable and more "unreal", the child's dream world is becoming more real Trauma increases both the frequency and the vividness of nightmares (Mannarino and Cohen, 1986). In my research, I found that the group most likely to show fantastic elements in their allegations (those with known severe trauma history) were also most likely to report nightmares.

40 It certainly would not have been recommended at the time that the interviewers begin to challenge the children when bizarre content arose. In fact, this is one of the common mistakes that interviewers make. Since fantastic content is more often based on child misunderstandings, memories from nightmares, incorporation of threats made by the perpetrators, and other phenomena distinct from conscious lies or fantasy, challenge is likely to be experienced as

general disbelief and nonacceptance. The content should be addressed very carefully, and we presently recommend leaving it alone in forensic interviews (and addressing it in therapy).

Demeanor Evidence

41 Parsonson's treatment of the demeanor of the children is very confusing given the clinical and research evidence. Early on in the brief, he argues that no one set of symptoms should be expected from an abused child, which is quite correct. However, when evaluating each of the children, he finds a child incredible when s/he shows "no evident indices of distress or emotion at any time, even when describing numerous events, any one of which if experienced, would be expected to cause a young child significant trauma" (paragraph 7.2.3, page 45). For Bart Dogwood he states that "almost any one of the alleged sexual and physical abuse events also would be expected to have serious physical and/or psychological effects on the children at the time. These could include depression, extreme withdrawal and apathy, and/or symptoms of post-traumatic stress disorder, including flashbacks, emotional and sleep disorders, and hypervigilance, as well as considerable resistance to attending creche or being in the proximity of allegedly abuse staff' (paragraph 7.2.5(d), page 46). In Tess's case he states that "one might have expected a traumatised child to evidence withdrawal, depression, or distress, with perhaps tears and evident reluctance when dealing with traumatic experiences, but there is no sign of this in any of the interviews" (paragraph 7.3 5(d), page 50).

42 I am quite surprised to see this type of statement from an expert on abuse. First, it should be noted that several of these symptoms are extremely rare in young children (such as flashbacks), and none characterise a large subgroup in other testimony studies. He is right that a lay person might expect these symptoms, but only 2% of our randomly selected (n=700) children cried during their testimony (and those who did cry were mostly children over 10). Wood et al's (1996) sample of videotaped abuse allegations also gives a figure for frowning and crying during testimony of 1.8% (of all coded emotional behaviours). Laughing, which Parsonson seems to believe is a sign of incredibility, is more common among accurately testifying children than is

crying in both Wood's sample and my own I would expect that this difference would become even larger as time passed The most common affective display in the Wood study was a relaxed or neutral expression

43 Again, it is understandable that a lay person might be surprised that children alleging abuse seldom cry or show extreme anxiety However, given that the foundational symptoms of PTSD involve the dissociation of emotions, this finding is well known among clinical experts I might also add that we have an unpublished study showing that children who are feigning child sexual abuse are more likely to cry and show negative affect

44 Finally, on this point, Parsonson notes that the children "rarely went beyond terms like "sad" or "mad" or "hurt", when terror, extreme distress, considerable pain and great anxiety would be expected to be the actuality of the experiences" (paragraph 7 2 5(e), page 47) I am unsure what age the children were at the time of the interview to which he refers However, I would point out that one effect of trauma is to fragment language and communication and to damage emotional recognition ability (Camras et al, 1988 is an example) Further, while one would expect a child of 4-6 to know the words mad, sad, and hurt, only about 10% of four year olds know the word terrified, and about 30% of six year olds know it About 15% of four year olds and 40% of six year olds know the word "anxious" The number who use these words in their emotion speech is very low, ranging from 0-10% in most samples The absence of these words in the children's interviews thus means very little Most of us who do quite a bit of interviewing know these norms generally, but there are also published norms that are available. One set that I know about are the Ridgeway norms, published in 1985 in Developmental Psychology

Recantation

45 Dr Parsonson's treatment of recantation is perhaps the most surprising of his conclusions I have six points to make there

45 1 First, the internal illogic of Parsonson's position should be fairly clear He is arguing that pressure by an interviewer will commonly lead a

child to lie or distort his or her report (in order to please), thus falsely accusing an innocent person Yet he also argues that pressure by a mother to retract, even if the accused is the father and the consequences will be destruction of the family, will not lead to retraction, setting the figure at 4% No theory of child abuse reporting and no empirical study of child abuse reporting has shown that errors of commission are easier to generate in a young child sample than are errors of omission Thus, Parsonson's statement is both theoretically difficult to support and contradicted by dozens if not hundreds of articles on relative rates of omission and commission in child trauma reporting, including the Poole and White studies, the Goodman series, the Steward and Steward chapters, and the Saywitz articles that are cited by both briefs

45 2 The 4% figure is the lowest figure ever reported in any study, and has been found once (by Bradley and Wood, 1996)

45 3 Parsonson does not disclose many reasons why the retraction rate in the Bradley and Wood study may be less applicable to the case at hand Among these reasons are

45 3 1 The age of the sample is not comparable Included in the sample were children who were preverbal, and thus could neither disclose or recant, and children who were much older than the creche children (43% were over age 11)

45 3 2 The only children who were included in this study were those who had physical evidence for their allegations For instance, between 63% and 81% of the perpetrators in this study had already confessed or had been convicted in court (The report states that 63% confessed and 12% were convicted, but does not state if these are independent figures) Logically, conviction or confession would reinforce a child's belief in his or her own memory and would make it less likely that others would reasonably doubt that memory

45 3 3 The study is a review of the records of the Department of Protective and Regulatory Services The children were not interviewed If the children did not retract immediately, but rather did so over time due to peer or parental pressure or memory changes, it is unlikely to be reflected in the statistics

45 4 In a survey of professionals, the view that retraction is rare in true cases, and therefore powerfully indicative of the falsity of the allegation, is almost unrepresented In a survey of 50 child abuse experts conducted by Morison and Greene, 1992, 100% disagreed with the statement that if a child were to retract an allegation, it was a reliable sign that the allegation is false. Again, Dr Parsonson's position is quite a radical departure from the mainstream

45.5 The methodological demands of the study of retraction are daunting The strongest study would demand following a child for years to check memory continuity, and controlling for the possibility that the repeated interviews themselves would remind the child of the event I therefore wish to clearly state that our estimates of the frequencies of recantation are imperfect, since the methodologies for current studies are worthy of critique Nonetheless, clinical studies who follow children in therapy find rates of denial and recantation from 25 to 75% (eg Sorenson and Snow, 1991, Gonzales et al, 1993) Survey studies of adults (asking if the victim denied or recanted an allegation when young) give comparable results (cf Jones, 1992) As stated earlier, surveys of experts also agree with this view.

45 6 American courts regularly concede that the view that recantation or delayed disclosure is rare may be held by lay persons, and that expert opinion refuting this view is acceptable

46 Parsonson is right, however, that recantation in false allegation cases appears to be extremely common It is not unusual in the research cited in both briefs to find that 80-100% of the children who make false allegations in a given study

recant if gentle requestioning or challenge is offered. I already cited the Leichtman and Ceci work, in which 85% of children recanted their false accusations, even in the circumstance of many repeated leading interviews. In the control group, where no repeated suggestion was used, all of the falsely accusing children retracted. The one recanting child described in the brief may well be a false allegation, but she also fits the clinical picture of children who recant true allegations (that is, children who have been teased, who have been accused of lying, and who wish the issue to be put to rest). We would expect the recantation rate to be quite high in the creche group if false allegations based on suggestion were the foundation for the initial reports. We would also expect a few recantations of true allegations, particularly years after the fact by those children who were shamed by others (as this child apparently was) for her accusation. Steward and Steward (1996) found that a major predictor for those children who denied a videotaped incident of medically-induced pain was the child's shame.

False Allegations to Abuse Questions

47 Finally, both affidavits give false impressions about the data regarding children's responses to misleading abuse-related questions (eg, He kissed you, didn't he? Or he took your clothes off, didn't he?) Lamb states that young children make 20-35% commission errors to such questions (paragraph 33, page 24), and cites Gail Goodman and her colleagues as his major reference. Below are some direct quotes from Goodman's work:

47.1 Rudy and Goodman, 1991: "Specifically, in response to our abuse-related questions, 7 year olds made only one commission error out of 252 opportunities. Even for four year olds, who made a total of 13 commission errors, 95% of their responses to the abuse questions were correct, and most four year olds (13 out of 18 children) did not make a single commission error to the abuse question". (Page 535).

47.2 Goodman and Aman, 1990: "Again, the most worrisome kind of error the children could make on the misleading abuse questions was a commission error. The five year olds on average made only 02 such

errors (all of which were in the no dolls-cues condition), whereas the three year olds on average made 21 such errors." (Page 1866-67)

47.3 Tobey and Goodman, 1992 [reporting on four year olds]. "However, hardly any of the children made errors of commission to the misleading abuse questions. Out of more than 70 chances, a total of only three errors of commission were made, all by boys in the observer condition, leading to an overall mean of .04.

47 4 Saywitz, Goodman, Nicholas and Moan, 1991: "Seven year olds, Mean = 99%, SD=.03, answered a higher proportion of misleading abuse questions accurately compared with five year olds, Mean=.96, SD= 10 However regardless of age, children demonstrated nearly perfect performance." (Page 688).

48. I agree with Lamb that Goodman's work is the best respected body of literature on the subject of children's reactions to open, direct, and misleading questions of abuse. However, I believe that a fair review of her work yields an estimate that children of the age relevant to the Ellis case (five year olds, according to Lamb), will make almost no false allegations with open questions, and few with direct or misleading questions. An average rate of false allegations for this age group would be no higher than 5%. If the creche children were 2-3 years of age when first interviewed, then Lamb's figures are correct.

Conclusions

49 Recantation and fantastic elements are not evidence for the falsity of a child abuse allegation by a five year old child.

50 Children of this age are quite resistant to misleading suggestions of the type allegedly posed.

51 The use of props and anatomical dolls does not undermine the credibility of the children's allegations.

52 The demeanor of the children, if accurately described, is not inconsistent with true allegations of child abuse.

SWORN by the said)
CONSTANCE J DALENBERG)
at San Diego, California)
this 7 day of June)
1999)
Before me)

Constance J Dalenberg, PhD

Notary Public

OFFICIAL SEAL
GOOL LELINWALA
NOTARY PUBLIC-CALIFORNIA
COMM. NO. 1189213
SAN DIEGO COUNTY
MY COMM. EXP. JULY 5, 2002