

JUDGING THE CREDIBILITY OF HISTORIC CHILD SEXUAL ABUSE COMPLAINANTS How Judges Describe Their Decisions

Deborah A. Connolly
Simon Fraser University

Heather L. Price
University of Regina

Heidi M. Gordon
Simon Fraser University

Criminal prosecutions of child sexual abuse alleged to have occurred in the distant past raise myriad challenges. One significant challenge involves assessing the credibility of complainants. In the present study, 51 bench trials involving 87 complainants were coded into categories related to complainants' memory for the offense, as well as credibility of the complainant, reliability of the evidence, and judicial inferences. A total of 4,827 judicial comments were identified and categorized. Judges were more likely to describe the allegations specifically than generally; however, they were sensitive to the predictably impoverished nature of memory after such a long delay. Consistent with psychological evidence, there were more judicial comments about inconsistencies in acquit cases than in convict cases and more comments about corroboration in convict than in acquit cases, although neither inconsistencies nor corroboration were strongly associated with verdict. Of some concern was the apparent and considerable judicial interest in complainants' behavior and emotions around the time of the alleged abuse and around the time of disclosure.

Keywords: historic child sexual abuse complaints, recent complaint warning, credibility assessment

“Historically, immediate complaint was a requirement before a sexual assault or abuse prosecution could succeed” (Lewis, 2006, p. 93). This rule probably dates back to 13th-century common law, where failure to make a timely complaint was a full defense to rape (Cossins, 2002). The rule evolved such that, by the 18th century, the trier of fact was expected to draw an adverse inference with respect to credibility if a timely complaint was not made to an official or layperson (Gobbo, 1970, as cited in Cossins, 2002). Accordingly, evidence of earlier complaints, ordinarily inadmissible as hearsay or as previous consistent statements, was admitted if the earlier complaint was made at the first reasonable opportunity after the offense was committed (Lewis, 2006). For at least two reasons this was a particularly oppressive doctrine for child victims of sexual assault. First, most victims of child sexual abuse (CSA) do not report their victimization immediately. In fact, it has been estimated that approximately two-thirds of

Deborah A. Connolly and Heidi M. Gordon, Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada; Heather L. Price, Department of Psychology, University of Regina, Regina, Saskatchewan, Canada.

This research was supported by a Social Sciences and Humanities Research Council (SSHRC) operating grant to Deborah A. Connolly, a Natural Sciences and Engineering Research Council graduate scholarship to Heather L. Price, and a SSHRC graduate scholarship to Heidi M. Gordon. We thank Jocelyn Conway for her invaluable assistance with data coding.

Correspondence concerning this article should be addressed to Deborah A. Connolly, Department of Psychology, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6 Canada. E-mail: debc@sfu.ca

child victims delay reporting their abuse, and half of those victims do not report having been abused until they reach adulthood (London, Bruck, Ceci, & Shuman, 2005). Second, in many *timely* complaints of CSA, the only prosecution evidence is the allegation of the complainant (Bottoms, Golding, Stevenson, Wiley, & Yozwiak, 2007; Bruck, Ceci, & Hembrooke, 2001; Cossins, 2002; Goodman, Batterman-Faunce, & Schaaf, 2002; Lamb, 1999; McCarron, Ridgway, & Williams, 2004; Pezdek et al., 2004). A fortiori, this is true in delayed or historic CSA (HCSA) cases, in which the passage of time is likely to have depleted or eradicated any corroborating evidence that might have been present at the time of the offense. The effect, then, of the judicial warning that an adverse inference be drawn from the failure to report the offense immediately (hereinafter referred to as the *recent complaint warning*) was to prevent the prosecution of most HCSA cases: The only prosecution evidence, the testimony of the complainant, was the subject of a judicial warning.

In this introduction, we review current law relevant to HCSA prosecutions in most common-law jurisdictions and then provide evidence for the following propositions: (a) There are a significant number of HCSA criminal prosecutions in most common-law jurisdictions around the world; and (b) perceptions of credibility play a central, if not determinative, role in the outcome of such cases. We end with a discussion of legal and psychological scholarship and commentary on assessments of credibility in the context of HCSA complaints.

In most common-law jurisdictions, the recent complaint rules no longer have the unmitigated effect they once had. In Canada, the rule that a trier of fact must draw an adverse inference from failure to complain immediately was abrogated in 1983; and in 2000, the Supreme Court of Canada held that a direction may be given to the trier of fact that delayed disclosure alone is not evidence of falsity of the allegation (*R. v. D.(D.)*, 2000). In England and Wales, a warning is permissive, not prescriptive, and in a recent Court of Appeal case (*George v. R.*, 2000, cited in Lewis, 2006) the court endorsed a direction that provides reasons, other than fabrication, for delayed disclosure of CSA. Australia and New Zealand went further in passing legislation that requires the judge to inform the trier of fact that delayed disclosure alone is not evidence of the falsity of the allegation (subsequent case law held that this legislation does not preclude also issuing the recent complaint warning). In the United States, some jurisdictions have retained the recent complaint warning, others retained it in certain circumstances only, and others have left the issue of a warning to the discretion of the judge (Lewis, 2006).

Coincident with these legislative and common-law changes was a dramatic increase in criminal prosecutions of HCSA cases in many common-law countries. J. A. Southin of the British Columbia Court of Appeal in Canada stated that, “historical sexual abuse prosecutions . . . are now a thriving legal industry” (*R. v. R.(J.W.)*, (2001), ¶ 26). In the United Kingdom, the Select Committee on Home Affairs stated the following:

In the last five years, 34 of the 43 police forces in England and Wales have been involved in investigations into allegations of child abuse in children’s homes and other institutions. All of the allegations relate to historical abuse, said to have occurred several years—often decades—ago. (Home Affairs Committee, 2002, p. 5)

Lewis (2006) concluded that, although the precise number of HCSA prosecutions is not available, there are a “substantial number” in the following juris-

dictions she studied: Canada, United States, Australia, New Zealand, England, Scotland, Wales, and Ireland. Professor Graham Zellick, Chairman of the Criminal Cases Review Commission in England, Wales, and Northern Ireland, stated the following on January 27, 2004, in reference to HCSA prosecutions: “[T]his area presents possibly the greatest challenge to the criminal justice system at the present time” (cited in Lewis, 2006, p. 1).

One important factor that makes prosecutions of HCSA cases particularly challenging is that trial outcome often rests on evaluations of credibility (e.g., Lewis, 2006; *R. v. Ay*, 1994). Credibility assessments have been described as “notoriously difficult” (*R. v. Marquard*, 1993, ¶ 49), “excruciatingly difficult” (*R. v. Tzarfin*, 2005, ¶ 10), “a very onerous task” (*R. v. King*, 2005, ¶ 3), and “inherently indeterminate” (Bala, Ramakrishnan, Lindsay, & Lee, 2005, p. 996). The difficulty may be compounded in HCSA cases because of a lack of corroborating evidence. As stated in *R. v. Ay* (1994):

This case is similar to many that have come before our courts in recent years involving allegations of sexual assault, made many years after the alleged abuse took place, in which the central and most difficult issue for the jury to resolve is the credibility respectively of the complainant and the accused, neither of whose testimony is supported by any independent confirmatory evidence. (¶ 52; see also *R. v. D.R.E.*, 2002)

How do judges assess credibility? The Supreme Court of Canada stated, “Assessing credibility is not a science . . . [It is a] complex intermingling of impressions that emerge after watching and listening to witnesses and attempting to reconcile the various versions of events” (*R. v. Gagnon*, 2006, ¶ 20). These impressions are formed on the basis of both verbal and nonverbal behaviors. In *Coghlan v. Cumberland* (1898), the English Court of Appeal held the following:

When, as often happens, much turns on the relative credibility of witnesses who have been examined and cross-examined before the judge, the Court is sensible of the great advantage he has had in seeing and hearing them. It is often very difficult to estimate correctly the relative credibility of witnesses from written depositions; and when the question arises which witness is to be believed rather than another, and that question turns on manner and demeanour, the Court of Appeal always is, and must be guided by the impression made on the judge who saw the witnesses. (p. 705)

There is very little published research on evaluating credibility in the context of continuous memory in HCSA complaints. The few studies that do exist compared perceived credibility of continuous (the complainant always knew he/she had been abused) and recovered (there was a time when the complainant did not know she/he had been abused) memories and so do not inform the present discussion.¹

¹ In these studies, type of memory, continuous or recovered, was manipulated and the credibility of the complainant was measured. Two studies reported no differences in recovered and continuous memory cases (Golding, Sego, & Sanchez, 1999; Tabachnick & Pope, 1997), whereas two found that perceived credibility was higher in continuous memory cases than in recovered memory cases (Bornstein & Muller, 2001; Sugarman & Boney-McCoy, 1997). In the present research, we are interested in factors other than type of memory that influence perceptions of credibility: The finding that complaints involving continuous memories may be seen as more credible than complaints involving recovered memories does not inform this discussion.

The purpose of the present research is to study how judges adjudicate HCSA cases by studying their decisions. We might have surveyed judges about how they adjudicate these cases, and, undoubtedly, that would have provided us with invaluable information. However, it is probably true that judges are not aware of all of the factors that influence their decisions. For instance, despite their enhanced legal knowledge, magistrate judges were susceptible to five cognitive biases that influence laypersons' decision making (anchoring, framing, hindsight, heuristic, and egocentric biases) and could lead to erroneous judicial decisions (Guthrie, Rachlinski, & Wistrich, 2001). Often, when cognitive biases and irrelevant information are used in decision making, their influences are unknown to the decision maker (e.g., Kahneman, 2002; Kassin & Sukel, 1997). Thus, a direct request may not reveal some of the factors that influence judicial decision making.

Given the dearth of research on credibility in HCSA cases and the absence of research on judicial evaluations in HCSA cases, we were unable to develop particular hypotheses with respect to the types of credibility-related information that could influence case outcome. Accordingly, we used a more inductive approach to coding the decisions (e.g., grounded theory; Charmaz, 2006; Rennie, Phillips, & Quartaro, 1988). That is, rather than test a priori hypotheses, we used a careful coding system that was intended to identify themes that existed in the decisions.

Method

Full-text reports of judicial trial decisions from 1986 to 2002 involving 51 criminal cases (95 complaints) of HCSA (i.e., a delay of 2 years or more from the end of the alleged abuse to trial²) were coded. Quicklaw was used to locate HCSA cases. Quicklaw is a full-text Canadian database that contains, at least from 1986 forward, all Supreme Court of Canada decisions, decisions from provincial Courts of Appeal, written decisions from the provincial Superior Courts, and written decisions from Provincial Courts that were forwarded to Quicklaw (see Connolly & Read, 2006, for a comprehensive description of the full database). Provincial court cases were forwarded to Quicklaw at the discretion of the presiding judge (it is not known what percentage of cases were forwarded), and we have not been able to find a common set of reasons used by judges to forward a case. However, it is fair to say that those cases that judges believed should be available to the legal community were forwarded. Given this, the nature of the sampling bias cannot be precisely stated.

² Although "delayed" or "historic" prosecutions are discussed and debated across disciplines, we have not been able to find a definition of the term. We turned to statutes of limitations for guidance. In both Canada and the United States, in ordinary personal injury cases, the modal limitations period is 2 years from the time when the alleged offense ended to the date on which an action commenced. That is the period selected to describe an offense as "historic" in these data. We could not, with any consistency, obtain the date the initial complaint was made or the charge date. There is undoubtedly a delay from charge date to trial date, but the length of the delay varies widely across time and across jurisdictions. Thus, we decided to include all cases in which the delay from the end of the alleged offense to trial is 2 or more years. As can be seen in Table 1, only one complaint in the present sample involved a delay of 2 years and only 6 complaints involved a delay of 3 years.

The full database contains 1,030 HCSA criminal cases involving 2,064 complainants. Within each case, the first complainant discussed by the judge was classified as alleging high-frequency (four or more alleged instances of abuse) or low-frequency (one to three alleged instances of abuse) abuse and as resulting in a conviction or an acquittal to create a 2×2 matrix. The entire case was then assigned to that cell. When all cases were assigned to a cell, a random number generator was used to select an approximately equal number of cases from each cell. Accordingly, approximately half of the complaints were resolved with convictions and approximately half were resolved with acquittals. Second, approximately half of the complaints involved allegations of high-frequency abuse and approximately half involved allegations of low-frequency abuse. Of the original 95 complainants, no credibility information was reported for eight complainants. Seven of these complainants were in the same case that involved 17 different complainants. The other complainant was in a case involving two complainants. We excluded these complainants from all analyses. Thus, the sample described here includes judicial discussions of 87 HCSA complaints. The sample is described in Table 1.

Coding

In developing a coding scheme, we borrowed principles from the grounded theory method of qualitative analyses (e.g., Charmaz, 2006; Rennie et al., 1988). All judicial statements related to the credibility of the complainant were first identified and decontextualized, or isolated. Statements were then recontextualized around themes or categories that emerged from the data.

To identify categories for later coding, we selected several cases from the larger sample of HCSA complaints. An advanced graduate student and an advanced undergraduate research assistant reviewed the cases independently to identify categories of judicial comments. The coders then met with the first author to discuss the categories; discrepancies were discussed and resolved. The category coders then selected another set of cases and began the process again. This process was repeated several times to establish a set of categories that encompassed the majority of judicial comments. Approximately 40 cases were reviewed to identify categories. None of those cases were included in the cases used to establish intercoder reliability for coding comments or final coding for the present study. After the establishment of the coding categories, the two coders randomly selected 10 complaints, identified comments for coding, and assigned each comment to one of the 17 established categories. Coders met to discuss and resolve disagreements. Overall, kappa was .77 for assignment of comments to the 17 categories. After achieving adequate intercoder reliability, one person coded the balance of the complaints.

For each complaint, coders first located the charge and the review of applicable law to exclude both from further analyses because they did not provide information used by the judge to describe the alleged offense or evaluate credibility. Coders then read the case a second time to identify the complainant(s), the accused, and other witnesses. Next, the coders read through the case a third time to identify units of information related to credibility. Each unit of information was a comment made by a judge that was independent of other comments. Only units

Table 1
Case Description

Case	<i>M</i> (<i>SD</i>), %, and range (<i>n</i>)
Mean age (in years) of complainant:	
When alleged abuse began	9.47 (4.03); range = 2–18 (<i>n</i> = 76)
When alleged abuse ended	12.32 (3.58); range = 4–19 (<i>n</i> = 75)
At trial	25.93 (8.80); range = 12–50 (<i>n</i> = 76)
Mean age (in years) of accused:	
When alleged abuse began	34.40 (12.47); range = 12–72 (<i>n</i> = 48)
When alleged abuse ended	37.00 (12.06); range = 13–72 (<i>n</i> = 48)
At trial	50.60 (12.45); range = 20–75 (<i>n</i> = 47)
Sex of complainant = male	32.9% (<i>n</i> = 26 of 79)
Sex of accused = male	100% (<i>n</i> = 79 of 79)
Abuse frequency	
One	34% (<i>n</i> = 28 of 83)
Two or more	66% (<i>n</i> = 55 of 83)
Complainant in therapy before trial	23% (<i>n</i> = 18 of 79)
Presence of reported repression	15% (<i>n</i> = 12 of 79)
Relationship	
Parental	18% (<i>n</i> = 14 of 76)
Other relative	28% (<i>n</i> = 21 of 76)
Nonrelative family connection	14% (<i>n</i> = 11 of 76)
Community connection	40% (<i>n</i> = 30 of 76)
Allegation	
Exposure/touching	63% (<i>n</i> = 49 of 78)
Nonpenile penetration ^a	17% (<i>n</i> = 13 of 78)
Penile penetration	20% (<i>n</i> = 16 of 78)
Duration of alleged abuse (in months)	34.85 (35.34); range = 0.1–144 (<i>n</i> = 78)
Delay from end of alleged abuse to trial (in years) ^b	13.46 (7.63); range = 2–36 (<i>n</i> = 79)
Acquittals	58% (<i>n</i> = 48 of 83)

Note. Judgments did not include all information about all variables and so numbers do not always sum to 87 and percentages were not always calculated from 87. In each cell we include the number of observations used to compute a mean when there is one category and number of actual observations of number of possible observations where there are two or more categories.

^a Abuse that was more intrusive than exposure or touching, but not penile penetration (e.g., simulated intercourse, digital penetration). ^b The number of complaints that were heard 2, 3, 4, and 5 years after the alleged offense ended was 1, 6, 6, and 1, respectively. In eight complaints, the judge described the delay (e.g., “many years ago,” “a long time ago”).

of information that were relevant to complainant credibility were coded; information pertaining to the defendant’s or other witnesses’ credibility was not coded. Each unit of information was coded only once, regardless of how often it was mentioned by the judge. When there were two or more complainants in a case, comments related to each complainant were coded separately. If a comment related to more than one complainant, one code was applied to each complainant. Each unit of information was then coded into one of the 17 categories described later. Categories captured all but 10 comments.

Coders were instructed not to read the verdict until after the case was coded. Although they were probably blind to outcome in some cases, in others the verdict

may have been evident from the text of the decision. It was not possible to keep the coders blind to the frequency of the alleged abuse, as it would have been clear early in the decision. Although the coders were not always blind to frequency of the alleged abuse or verdict, we did not have particular hypotheses about the effects of these variables.

We identified four broad themes: memory for the alleged offense, reliability of the evidence, credibility of the complainant, and judicial inferences. Memory for the alleged offense refers to the way in which judges described the offense and includes specific details of a particular instance of the allegation, general descriptions of the offense(s), and comments about memory failures. Reliability of the evidence contained comments about inconsistencies, corroboration, and quality of the evidence. Three major categories emerged in judicial discussions of credibility: the behavior or emotional response of the complainant around the alleged abuse, around disclosure, and at trial. Within each of these major categories, there were minor categories as described later. The fourth theme, judicial inferences, refers to conclusions that the judge reached without apparent evidence. Below is a description and example for each of the major and minor categories. Note that examples are direct quotes from judges and may not necessarily reflect actual comments provided by the complainant. In Table 2, themes and major categories are presented in boldface type, and minor categories are presented in standard font.

Memory for the Alleged Offense

Specific Details

There was at least one detail from a specific or single incident of the alleged abuse (e.g., “one time, the accused unbuttoned his pants”). If the case involved an allegation of multiple instances of abuse, single instances were considered independent when, between instances, the complainant was physically free from the accused.

General Details

General information concerning how and when the alleged abuse normally occurred. Such details were often evinced by past continuous tense such as “he would,” and “it was always” (e.g., “he would try to kiss me”).

Memory Failures

The complainant either admitted memory failure or the judge discussed memory failures or alluded to lack of memory for details (e.g., discussed missing details).

Reliability of the Evidence

Inconsistencies

Details reported differently on different occasions (e.g., “said it happened every time she wore a skirt, but it only happened once”), details reported for the first time in later interviews or at trial (e.g., “there was no detail about the accused moving his hand in the police statement”), details not reported in later interviews but reported in earlier interviews (e.g., “previously testified to seeing his hand down her shirt”), and

Table 2
Judicial Comments in Each Theme, Major Category, and Minor Category

Judicial comment	Frequency	% of total	<i>M (SD)</i> per complaint	<i>n</i> of complaints not discussed
Memory for the alleged offense				
Overall	2,421	50.15	27.83 (20.11)	1
Specific details	1,421		16.33 (18.42)	16
General details	832		9.56 (10.24)	33
Memory failures	168		1.93 (2.45)	35
Reliability of the evidence				
Overall	776	16.08	8.92 (7.06)	6
Inconsistencies	349		4.01 (4.08)	20
Corroboration	223		2.56 (3.87)	35
Quality of the evidence	204		2.34 (2.25)	22
Evaluations of credibility				
Overall	1,384	28.67	15.91 (13.77)	0
Behavior/emotion at time of abuse	795		9.14 (9.86)	7
Conduct after the assault	678		7.79 (8.32)	12
Resistance to the offense	117		1.34 (2.42)	43
Behavior/emotion at disclosure	313		3.60 (3.60)	6
Reasons for delay	101		1.16 (1.61)	41
Possibility of collusion	83		0.95 (1.80)	57
Motive to fabricate	46		0.53 (0.85)	55
Conduct after/around time of disclosure	42		0.48 (1.32)	68
Reason for disclosure	41		0.47 (0.76)	56
Behavior/emotion at trial	276		3.17 (4.94)	25
Demeanor	215		2.47 (4.35)	32
Personality	61		0.70 (1.18)	56
Judicial inferences	246	5.10	2.83 (2.74)	20
Someone should have known	46		0.53 (1.16)	61
Other inferences	200		2.30 (2.54)	27
Total	4,827	100.0	55.48 (37.01)	0

Note. Themes and major categories are presented in boldface type, and minor categories are presented in regular type.

differences between complainant's testimony and other witnesses regarding specific details (e.g., "denied speaking with mom before police").

Corroboration

Evidence that could be used to support central or collateral details of the allegation (e.g., "a witness said she saw the girls on the accused's lap," "the accused admitted to his family that there were sexual relations").

Quality of Evidence

General comments about characteristics of the evidence (e.g., "none of the complainants gave entirely straightforward or unambiguous evidence," "appeared to recall the evidence well").

Evaluations of Credibility

Behavior/Emotion of Complainant Related to the Alleged Abuse

After an assault. This pertained to the complainant's emotional and/or behavioral state attributable to the alleged abuse (e.g., psychiatric care, running away, attempted suicide, feelings about the accused) or that was considered striking in light of the alleged abuse (e.g., "the complainant visited the accused and told the accused's mother that she was really going to miss them"). The emotion/behavior might have been reported as occurring while the alleged abuse was ongoing (e.g., between instances), shortly after it ended, or long after the alleged abuse ended.

Resistance to offense. This includes physical, verbal, or emotional resistance or nonresistance during or after the alleged abuse. (e.g., "she did not attempt to get away or resist because she did not care anymore").

Behavior/Emotion of Complainant Related to Disclosure

Reason for delay. This pertains to the complainant's reason for not disclosing the alleged abuse immediately. It may include threats by the accused, memory problems, and/or lack of a trusting individual to tell (e.g., "he told her that what he was doing was not wrong").

Possibility of collusion. The judge discussed the issue of collusion, whether acknowledging its possibility or stating that s/he believed none was evident (e.g., "complainant's accounts have merged").

Motive to fabricate. This pertains to judicial discussion of the reasons for or against the possibility of fabrication (e.g., "she told the story without alternative motive," "she is not sophisticated enough to fabricate a story").

Behavior/emotion after or around time of disclosure. This pertains to the complainant's behavior or emotions displayed to an authority figure around the time of disclosure (e.g., "they decided to proceed and her mother arranged for her to meet the detective"). An authority figure was defined as anyone who would be required by law to report known or suspected child abuse.

Reason for disclosure. The complainant's reason to disclose was discussed by the judge (e.g., "saw the accused picking up her 4-year-old son").

Behavior/Emotion of Complainant at Trial

Demeanor. This describes the complainant's attitudes, emotions, and behaviors on the stand (e.g., "mode of expression was sometimes awkward").

Personality traits. This involves enduring constructs of the complainant such as intelligence, memory, propensity to lie, education, and/or class (e.g., "does not see herself as worthwhile").

Judicial Inferences

Someone Should/Could Have Known

This includes judicial statements or evidence relating to the presumed presence or knowledge of other people at the time of the alleged abuse (e.g., "he would not have been able to see into the living room from where he was").

Other Inferences

These are inferences made by the judge for which external support is not provided (e.g., “every memory fades with time”).

Results

A total of 4,827 comments were identified and coded into one of four themes: memory for the alleged offense, reliability of the evidence, credibility of the complainant, and judicial inferences (10 comments were unclassifiable and are not discussed). Within each theme, the data are organized so that descriptive data (frequencies are in Table 2) are reported first, followed by inferential statistics (mean percentages are given in Table 3). Descriptive data reflect frequencies in themes as well as major and minor categories to provide both an overview and a detailed description of judicial comments.

Inferential statistics were used to investigate differences as a function of frequency of the alleged abuse and case outcome. In the original coding, frequency of the alleged abuse was defined as either one to three incidents (low frequency) or four or more incidents (high frequency), and cases were selected and cell sizes were fairly balanced on the basis of this definition. An anonymous

Table 3
Overall Mean Percentage of Comments (and Standard Deviations) as a Function of Alleged Abuse Frequency and Outcome

Comment	Alleged abuse frequency		Outcome	
	Once	Twice or more	Acquit	Convict
Memory				
Specific details	35.67 (21.49)	26.07 (20.53)	28.71 (21.26)	30.08 (21.45)
General details	13.05 (18.84)	21.10 (16.05)	17.01 (18.59)	20.15 (19.71)
Memory failure	2.71 (2.90)	3.28 (4.20)	3.59 (3.89)	2.43 (3.63)
Reliability				
Inconsistencies	4.98 (4.73)	7.26 (6.69)	8.02 (6.93)	4.53 (4.36)
Corroboration	5.47 (7.03)	3.34 (4.39)	2.61 (4.72)	5.92 (5.88)
Quality of the evidence	3.22 (3.72)	5.01 (4.57)	5.78 (4.49)	2.65 (3.53)
Credibility				
Conduct after assault	13.51 (9.35)	11.69 (8.40)	12.62 (8.44)	11.89 (9.16)
Resistance to offense	1.73 (2.74)	2.64 (3.76)	1.32 (2.07)	3.64 (4.38)
Reasons for delay	2.39 (2.55)	1.67 (2.05)	1.76 (2.25)	2.12 (2.24)
Possibility of collusion	2.03 (3.90)	2.76 (4.82)	1.87 (3.68)	3.34 (5.36)
Motive to fabricate	1.30 (1.88)	0.90 (1.43)	1.11 (1.86)	0.95 (1.20)
Conduct after disclosure	0.56 (1.34)	0.79 (2.11)	0.85 (2.26)	0.54 (1.24)
Reason for disclosure	1.05 (1.96)	0.78 (1.35)	0.95 (1.81)	0.77 (1.23)
Demeanor at trial	5.01 (8.08)	4.42 (7.09)	4.17 (7.25)	5.19 (7.64)
Personality traits	1.67 (2.47)	1.37 (2.59)	1.66 (2.81)	1.24 (2.02)
Judicial inferences				
Someone should have seen	0.41 (1.03)	0.87 (1.52)	1.09 (1.68)	0.23 (0.62)
Other inferences	5.19 (7.01)	6.02 (6.40)	6.86 (5.92)	4.31 (5.79)

reviewer pointed out the difficulties associated with knowing when one instance ends and another begins, as well as that event confusion and scripting may occur after just two encounters. Accordingly, for the purposes of analyses, frequency of alleged abuse definitions were changed to one instance (single) and two or more instances (repeated). The number of complainants in each cell ranged from 30 (repeated/acquit) to 12 (single/convict).

Analyses were conducted on average percentages for minor categories only (i.e., the average percentage of comments in each minor category subsumed under memory, reliability, credibility, and judicial inferences separately). Themes and most major categories are sums of minor categories and do not provide additional information. With the exception of memory for the alleged abuse, correlations between categories within themes were small or absent and did not necessitate multivariate tests. Within the memory for the alleged abuse theme, there was a substantial negative correlation between general and specific descriptions ($r = -.75, p < .001$) and a significant correlation between comments about general details and memory failures ($r = -.25, p < .05$). This probably reflects a kind of trade-off between reports of general versus specific descriptions such that an allegation is described either generally or specifically and the two kinds of descriptions do not co-occur often. Because two of the possible six correlations were significant, we conducted a multivariate test on all minor categories related to memory for the alleged abuse before interpreting the univariate tests. There were no correlations between categories of reliability. There were seven significant correlations between credibility categories, and five involved comments about the possibility of collusion. Given that judges made comments about collusion for only about one third of complainants, we did not conduct multivariate analyses on credibility categories.

With these data, it is possible that the independence assumption of an analysis of variance (ANOVA) was violated: 17 of 51 cases contained more than one complainant, and in such cases, multiple complainants were adjudicated by the same judge. Although judges are directed to evaluate the evidence related to each complainant independent of evidence related to other complainants in the same case, it would be precarious to rely on this to argue that the data are independent. To test whether our conclusions were compromised, we omitted all cases that included 2 or more complainants (leaving us with 30 cases and, of course, 30 complainants). Only one effect was lost—the effect of verdict on quality of the evidence. Given the substantial loss of complaints and the similarity between both patterns of findings, we report analyses from the entire data set.

Memory for the Alleged Offense

As can be seen in Table 2, judicial references to specific or general details of the alleged offense or to memory failures were present in discussions of all but one complaint. Comments concerning specific details about particular instances of the alleged abuse represented 60% of all comments describing the allegations. In fact, in over 80% of the complaints, judges described particular details about at least one specific incident of alleged abuse. Although comments about general details of the alleged offense were frequent, they were less common than comments about particular details. The absolute number of comments concerning

memory failures was low. However, little could have been said other than to state that, in the judge's opinion, the complainant's memory was incomplete. Given this inevitable brevity, it is noteworthy that comments about memory failures were present in almost 60% of all complaints.

The multivariate main effect of frequency of the alleged abuse was not significant, Wilks's $\lambda = 0.94$, $F(3, 74) = 1.46$, $p = .23$; nor was the multivariate main effect of verdict, $F < 1$.

Reliability of the Evidence

The absolute number of comments about reliability of the evidence was quite low, representing only about 16% of all comments. It is important, however, that only six complaints did not contain any reference to reliability of the evidence. Moreover, comments about inconsistencies, corroboration, and quality of the evidence were present in 77%, 60%, and 75% of complaints, respectively. The main effects of verdict on inconsistencies, corroboration, and quality of the evidence were significant, $F(1, 76) = 4.44$, $p < .05$, $\eta^2 = .05$; $F(1, 76) = 7.45$, $p < .01$, $\eta^2 = .09$; and $F(1, 76) = 7.80$, $p < .01$, $\eta^2 = .09$, respectively. As can be seen in Table 3, there were more comments about inconsistencies and quality of the evidence in acquit than convict cases and more comments about corroboration in convict than acquit cases. There were no other effects.

Evaluations of Credibility

As can be seen in Table 2, almost 30% of all judicial comments pertained to the credibility of the complainant and discussions of every complaint included at least one such comment. In fact, comments about credibility of the complainant were almost twice as frequent as comments about the reliability of the evidence. The most common minor category was conduct of the complainant around the time of the alleged abuse, representing almost 50% of all comments about credibility and present in over 85% of all complaints. The next most frequent minor category (a distant second) was the demeanor of the complainant at trial, representing about 15% of all comments about credibility and present in over 60% of all complaints. There were over 100 comments about each of the minor categories concerning the complainant's resistance to the alleged offense and reasons for delay. In both cases, at least one comment was present in over half of the complaints.

There were no effects of frequency of the alleged abuse. There was a main effect of verdict on resistance to the offense, $F(1, 76) = 7.74$, $p < .01$, $\eta^2 = .09$. Judges made more comments about resistance to the offense in convict than acquit cases.

Judicial Inferences

About 5% of the judicial comments coded were inferences, either that someone should have seen something or some other type of inference. There was at least one judicial inference in 75% of all complaints. There was a main effect of verdict on the judicial inference "someone should/could have known," $F(1, 76) = 6.27$, $p < .05$, $\eta^2 = .08$. There were more such comments in acquit than convict cases.

Discussion

How do judges describe their reasons in criminal cases involving HCSA? We organize our discussion of the results around four themes. First, we discuss how judges described the complainants' memory of the alleged offense(s). Our second theme concerns judicial comments about reliability of the evidence (i.e., inconsistencies, corroboration, and quality of the evidence,). Third, we discuss comments concerning the conduct (behavior and/or emotion) of the complainant at three points in time. Finally, we discuss judicial inferences.

Memory for the Alleged Offense

Although judicial comments about general details constituted a substantial minority of most descriptions, comments about specific details were far more common—almost twice as common as comments about general details and present in over 80% of all complaints. General descriptions were present in allegations of both single and repeated abuse, suggesting a loss of memory for particular details for all complainants. However, there were almost twice as many general details in allegations of repeated versus single abuse, suggesting a greater loss of particular details for complainants in the former group compared with the latter group. Indeed, general details constituted only about 21% of the memory details reported in complaints of single-frequency abuse and 43% of the memory details reported in complaints of repeated abuse. Because we used judicial decisions rather than trial transcripts, we coded how the *judge* described the allegation, not necessarily how the *complainant* described the allegation. Although there is likely to be some consensus between the two, consensus would not be perfect. When there is disparity, it is more likely that a specific allegation was described in general terms than that a general allegation was described in specific language: A judge can modify a complainant's specific report and describe it generally but is less likely to have the necessary particular information to describe a general memory report with specific details. Accordingly, the complainant's general memory may be overstated and specific memory may be understated in the present study.

In most nonforensic contexts, we would not have predicted this dominance of specific over general details. Autobiographical memory becomes more general and less particular over time (e.g., Brewer, 1986; Fivush, 1997; Hudson & Nelson, 1983; Myles-Worsley, Cromer, & Dodd, 1986), and this process may be particularly pronounced for memory for repeated events (e.g., Fivush, 2002; Slackman & Nelson, 1984). In the present study, all trials occurred at least 2 years after the final alleged abuse instance and half of the complainants alleged repeated abuse (i.e., two or more instances). This combination of a long delay in all cases and allegations of repeated abuse in most of the cases would have led us to expect that general details would be more common than specific details. However, in the relatively singular context of a criminal trial, the importance of specific details related to the allegation may prevail.

The sufficiency principle, articulated in *Brodie v. The King* (1936) and affirmed in *R. v. B. (G.)* (1990) states that, "the indictment must describe the offense so as to lift it from the general to the particular." This may not mean that particular details about time and place must be recalled with exactitude (*R. v.*

W.(R.), 1992). However, some particular details about individual instances are likely to be required (see also Connolly, Price, Lavoie, & Gordon, 2008; Connolly & Read, 2003; Guadagno, Powell, & Wright, 2006). Certainly, we do not suggest that, in the context of a criminal proceeding, complainants are better able to recall particular details of an event or events that happened in the distant past. It seems more likely that those who could not recall particular details of at least one instance of the alleged abuse were less likely to proceed to trial (for a consistent conclusion with different data, see Guadagno et al., 2006). On the one hand, this seems to place those who are victims of multiple assaults and/or who wait a long time to report the abuse at a disadvantage with respect to “having their day in court.” On the other hand, it would be contrary to the principles of fundamental justice to require an accused person to defend himself against an allegation that is vague and imprecise.

We were surprised that judicial comments about specific details were not more common in cases that ended in convictions than acquittals. Scholars have found that a report that contains particular details is often seen as more credible than a report that does not contain particular details (Bell & Loftus, 1989; Borckardt, Sprohge, & Nash, 2003). If, as suggested above, all complainants had to meet a threshold level of memory for specific details to proceed to trial, those complainants who recalled more details than required to meet the threshold may not have increased their chances of success at trial. Given the constraints of our data, this is speculative and requires further investigation to support.

Reliability of the Evidence

In all three measures of reliability of the evidence, the percentage of comments varied as a function of verdict and in a way that is consistent with psychological research. Comments about inconsistencies in the complainants' evidence were relatively infrequent (only about four comments per complaint). It is important, however, that judges made at least one comment about inconsistencies when discussing over 75% of complaints, and there were more such comments in cases that ended in acquittals than convictions. Mock jurors see report consistency, at least with moderately confident witnesses, as indicative of accuracy (Berman & Cutler, 1996; Berman, Narby, & Cutler, 1995; Brewer, Potter, Fisher, Bond, & Luszcz, 1999; Conte, Sorenson, Fogarty, & Rosa, 1991). Similarly, some legal scholars believe that an inconsistent witness lacks credibility. For instance, Brewer and colleagues quoted from a chapter written to assist legal professionals with cross-examination, “A true inconsistency can effectively destroy a witness, and sometimes a whole case If you find a true inconsistency, or if you can manufacture one, then use the deposition of previous testimony to sheet it home” (p. 298). Notwithstanding this intuitively reasonable belief that an inconsistent witness is an inaccurate witness, it appears that the actual relationship between consistency and accuracy is small. In Brewer et al. (1999), 62 participants watched a videotape of a crime and were interviewed 1 hr later and again 2 weeks later. Only 2 participants were entirely consistent across reports. Of note, there was a small but significant relationship between accuracy and consistency: Consistency accounted for no more than 10% of the variance in accuracy (see also Fisher & Cutler, 1995). The judges in the present study appeared to be sensitive

to the small but reliable relationship between consistency and accuracy: Verdict accounted for only 5% of the variability in judicial discussion of inconsistencies.

Although judicial comments about corroboration were infrequent, at least one relevant comment was made in almost 60% of complaints, and judicial discussions of corroboration were more frequent in convict than acquit cases. The conviction rate among complaints if corroboration was discussed was 54%, whereas it dropped to 24% among complaints if corroboration was not discussed. As discussed in the introduction, often, there is no corroborating evidence in CSA cases (e.g., Bottoms et al., 2007). Given the long delay to prosecution, this paucity of supporting evidence is likely to be exacerbated in HCSA cases. The criminal justice system has a very old tradition of interest in and, in some cases, a requirement for corroboration (Connolly & Read, 2003; Meyers, 1996). For most of the 20th century, in the United States, a conviction in a CSA case could not be entered without corroboration (Meyers, 1996). In Canada, until 1988, courts were required to warn triers of fact of the dangers of convicting on the uncorroborated evidence of an unsworn child (most children in Canada provide unsworn testimony; Bala, 1999). Other members of the justice system also place substantial weight on corroboration. Platt (2006) interviewed social workers in 23 cases of alleged child maltreatment. Typically, social workers use one of, or a combination of, five criteria to decide whether to proceed with a full investigation: specificity of the allegation, severity of the allegation, risk, parental accountability, and corroboration. Today in Canada and the United States corroboration is not necessary for a conviction. This does not, of course, mean that corroboration is not valuable. In the present study, there was more discussion of corroboration in convict than in acquit cases; however, verdict was associated with only 9% of the variability in comments about corroboration. As stated in the legal rule, *singularis testis*, it is possible to convict a person on the basis of the uncorroborated evidence of a single witness (in the present data, in eight convictions, there was no mention of corroboration). Nonetheless, our data are consistent with the notion that some corroboration can increase one's chances of conviction.

Recall that judicial comments about the quality of the evidence were general comments about the evidence without referring to particular details. For instance, "it is the central act about which there is inconsistent evidence," "evidence is clear," "the evidence creates a doubt," or "she presented her evidence in a credible way." There were more comments about the quality of the evidence in acquit than in convict cases. In fact, the average percentage of such comments in acquit cases was twice the percentage of such comments in convict cases. General comments about quality were more likely to be used to support a conclusion that the allegation did not happen than to support a conclusion that it did happen.

Evaluations of Credibility

Conduct of the complainants around the time of the alleged abuse represented almost 50% of all comments related to credibility. The more specific judicial comments concerning the complainants' resistance to the offense were more common in convict than in acquit cases. In spite of the apparent judicial interest in complainants' reactions to the alleged abuse, there is no empirical evidence to establish a constellation of behaviors that can reliably inform the presence or

absence of sexual abuse. There was a time after Summit (1983) published his seminal paper on CSA accommodation syndrome (CSAAS) that some scholars and practitioners took his five stages (secrecy; helplessness; entrapment and accommodation; delayed, conflicted, and unconvincing disclosures; and retraction of disclosure) as diagnostic of CSA. In a recent review, however, London et al. (2005) concluded that, with the exception of delayed disclosure (common among CSA victims), there is no clear empirical evidence for the other four stages. Malloy, Lyon, and Quas (2007) reported that recantation rates may be higher than suggested in London et al. (2005), but they are far from diagnostic of CSA. In a very thorough review of the literature, Kendall-Tackett, Williams, and Finkelhor (1993) concluded that there is not a constellation of symptoms that can reliably discriminate child victims of CSA from children who had been exposed to other forms of trauma (see also Brilleslijper-Kater, Friedrich, & Corwin, 2004 for a more recent assertion of this conclusion). Reactions to abuse are varied, and victims' reactions cannot be taken as evidence for or against an allegation of sexual abuse. Despite this empirical evidence, it is notable that judges appear to be quite interested in complainants' reactions to the alleged abuse, as evidenced by the substantial amount of related rhetoric and the impact on case outcome. Given these data and relevant scholarship, we propose that evidence concerning the conduct of the complainant around the time of the offense should be presumptively inadmissible; that is, neither side would be allowed to introduce such evidence unless the judge is convinced that such evidence is necessary in the interests of justice. This is not unlike the rule that makes previous consistent statements presumptively inadmissible—the fact that the complainant told several people about the offence does not make it more true; to hear such evidence would be a waste of the court's time and, furthermore, it could be misused.

Although the frequency of comments concerning the circumstances of disclosure was relatively small, at least one such comment was present in about 90% of complaints. The most frequent minor category was comments about the reason for the delay, present in over half of the complaints. As discussed earlier, Summit's (1983) CSAAS and the excellent review by London et al. (2005) converge on the conclusion that delayed disclosure is disturbingly common: The percentage of children who delay disclosing abuse may be as high as 66%. Most common-law jurisdictions have recognized this by abrogating recent complaint rules. Delayed disclosure alone should be neither probative nor prejudicial to the prosecution. Ideally, then, there should be no need to adduce evidence to explain the delay. The problem with this recommendation is that delayed disclosure is patently obvious, and triers of fact may draw adverse inferences in the absence of explanations for the delay. To deal with this, some jurisdictions (e.g. Canada, Australia, New Zealand, England, Wales), have either recommended or mandated a judicial direction to the effect that delay alone is not evidence of falsity of the allegation (Lewis, 2006). We know, however, that jury instructions may be ignored, particularly if the instruction is inconsistent with jurors' beliefs (e.g., Ogloff & Rose, 2005). Is delayed disclosure inconsistent with laypersons' beliefs? Desmarais, Read, Rojas, and Vander Veen (2005), reported that 82% of community participants ($n = 210$) agreed with the statement "a delay in reporting abuse is expected when the victim is a child." This suggests that a judicial direction may be sufficient to defend against adverse inferences that may arise from an unex-

plained delay in reporting CSA. However, to the best of our knowledge, this is the only study to investigate this question, and further research is needed before a more definitive recommendation can be made.

The conduct of the complainant at trial represented about 12% of the comments concerning credibility. The most common minor category was the demeanor of the complainant, representing over 75% of all comments about the conduct of the complainant at trial. Demeanor comments concerned primarily complainants' affective state, as expressed in their apparent emotion or the manner with which they testified. Much of the extant literature supports the notion that an "appropriate" amount of emotion expressed by a complainant increases perceptions of credibility (Golding, Fryman, Marshall, & Yozwiak, 2003; Kaufman, Drevland, Wessel, Overskeid, & Magnussen, 2003; Regan & Baker, 1998; Rose, Nadler, & Clark, 2006; Salekin, Ogloff, McFarland, & Rogers, 1995; Wiener, Bornstein, & Voss, 2006). Regan and Baker asked mock jurors what they expected of a child complainant when she first confronts an accused in the courtroom. The most common response (81%) was crying. On the basis of the current study and one by Wessel, Drevland, Eilersten, and Magnussen (2006), the conclusion may not generalize to judges' decisions. Fifty-three judges from Norway watched a video of an actress reporting details of a rape in one of three conditions: congruent emotion (she tried to maintain control), neutral (flat with little emotion), and incongruent (she held a positive expression and even smiled occasionally). There was no effect of emotion on judges' credibility judgments (Wessel et al., 2006). In the present research, although judges appeared to view complainants' demeanor as relevant enough to describe it in their decisions, there were relatively few such descriptions (an average of 2.47 comments per complainant), there was no discussion in almost 40% of complaints, and comments did not vary as a function of case outcome. On the basis of the study by Wessel et al. and our analysis of judicial decisions, we tentatively conclude that demeanor does not have a substantial effect on judicial assessments of credibility in historic sexual assault cases.

Judicial Inferences

At least one judicial inference or comment that someone should have seen the abuse was made in over 75% of the reports. Judicial inferences were most likely to involve either comments about memory or expected conduct at the time of abuse or disclosure. These differ from the other categories in that these comments did not deal with actual memory or reported conduct but what the judge believed should have been true. It is interesting that there were more judicial inferences and comments that someone should have seen the alleged abuse in cases that ended in acquittals than convictions. Both types of comments were more likely to disadvantage the complainant than to advantage him or her. Our data do not allow us to provide even a moderately compelling explanation for this finding. We do note, however, the challenges associated with supporting a null hypothesis (i.e., that a crime did not occur). In the present data, judges were more likely to support this judgment with judicial inferences than to use judicial inferences to support a judgment that a crime had occurred.

Limitations

There are several limitations related to this research. First, as with most archival and field-based research, it was not possible to use strict control, to manipulate variables, or to randomly assign complainants to conditions. Accordingly, no causal conclusions can be drawn from these data. Second, we investigated only criminal decisions in HCSA cases. We have no data that would allow us to generalize our conclusion to other criminal offenses or to civil trials.

Third, we coded the written decisions rather than trial transcripts. There are both practical and methodical reasons for this. By using written decisions rather than trial transcripts, we were able to investigate a much larger sample of cases (the cost and time involved in evaluating 87 complaints based on transcripts would have been prohibitive). More important, in our view, the information of interest is more likely to be present in decisions than in trial transcripts. We were interested in studying the information that judges use to adjudicate HCSA cases. The mere presence of evidence at trial does not mean that it contributed to a judge's decision (trial evidence may be accepted in whole, in part, or not at all). We were more confident that judges would report the evidence they used to render a verdict in summaries of their decisions. This methodological decision, however, places a particular restriction on inferences we can draw concerning the complainants' memory. We based our inferences on the judges' descriptions of the allegations rather than on complainants' descriptions. Our conclusion, therefore, that memory for specific details is paramount in criminal trials is tentative.

Fourth, we coded the presence and classification of comments: We did not code judicial evaluations (i.e., the weight placed by the judge) of each comment. Most often, judges' comments were not individually evaluated, and so most would have been coded as neutral. This restricts the interpretations we can draw from the data. Throughout the article, we discuss judicial interest in particular kinds of evidence rather than judicial evaluations of such evidence.

Conclusion

Overall, these data lead us to the conclusion that, for the most part, factors that influence verdicts in HCSA criminal cases are supportable with empirical research and consistent with high-court rulings. We were very impressed with how well judges balanced, on the one hand, the capacities of complainants to remember an event that occurred in the distant past (sometimes repeatedly) when he/she was a child and, on the other hand, the right of the accused to be charged with a "specific" offense. Judges appeared to be sensitive to the predictably impoverished nature of memory for particular details in that they did not ignore general details. In fact, general details represented a substantial minority of comments concerning descriptions of the alleged offenses. At the same time, sensitivity to the requirement that an allegation be specific enough for the accused to raise a defense was evidenced by our finding that descriptions of specific details were more common than descriptions of general details. These data also demonstrate, for the first time, the importance of report inconsistencies and corroboration to verdict in a large sample of judicial decisions. Legal directives and psychological literature have made it clear that the presence of inconsistencies and/or a lack of corroboration should not be determinative of case outcome. Our data are entirely

consistent with this: Verdict was associated with only 5% and 9%, respectively, of variability in inconsistencies and corroboration.

The only qualification to our general conclusion concerns judicial interest in the conduct of the complainant, around the time of the alleged abuse, disclosure, and trial, despite a lack of evidence that the conduct of the complainant is diagnostic or even helpful. Although we did not find an association between verdict and comments about the conduct of the complainant, it would be wrong, in our estimation, to ignore the judicial interest. We suggest that evidence concerning the conduct of the complainant around the time of the alleged abuse be presumptively inadmissible. A similar recommendation cannot be made for conduct of the complainant around the time of disclosure until it becomes generally understood that delayed disclosure is common and delayed reporting alone cannot be used to either enhance or damage the credibility of the complainant.

References

- Bala, N. (1999). Child witnesses in the Canadian criminal courts. *Psychology, Public Policy, and Law*, 5, 323–354.
- Bala, N., Ramakrishnan, K., Lindsay, R., & Lee, K. (2005). The judicial assessment of the credibility of child witnesses. *Alberta Law Review*, 42, 995–1017.
- Bell, B. E., & Loftus, E. F. (1989). Trivial persuasion in the courtroom: The power of (a few) minor details. *Journal of Personality and Social Psychology*, 56, 669–679.
- Berman, G. L., & Cutler, B. L. (1996). Effects of inconsistencies in eyewitness testimony and mock-juror judgments. *Journal of Applied Psychology*, 81, 170–177.
- Berman, G. L., Narby, D. J., & Cutler, B. L. (1995). Effects of inconsistent eyewitness statements on mock-jurors' evaluations of the eyewitness, perceptions of defendant culpability and verdicts. *Law and Human Behavior*, 19, 79–88.
- Borckardt, J. J., Sprohge, E., & Nash, M. (2003). Effects of the inclusion and refutation of peripheral details on eyewitness credibility. *Journal of Applied Social Psychology*, 33, 2187–2197.
- Bornstein, B. H., & Muller, S. L. (2001). The credibility of recovered memory testimony: Exploring the effects of alleged victim and perpetrator gender. *Child Abuse & Neglect*, 25, 1415–1426.
- Bottoms, B. L., Golding, J. M., Stevenson, M. C., Wiley, T. R. A., & Yozwiak, J. A. (2007). A review of factors affecting jurors' decisions in child sexual abuse cases. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *The handbook of eyewitness psychology, Vol. I: Memory for events* (pp. 509–543). Mahwah, NJ: Erlbaum.
- Brewer, N., Potter, B., Fisher, R. P., Bond, N., & Luszcz, M. A. (1999). Beliefs and data on the relationships between consistency and accuracy on eyewitness testimony. *Applied Cognitive Psychology*, 13, 297–313.
- Brewer, W. P. (1986). What is autobiographical memory? In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 25–49). New York: Cambridge University Press.
- Brilleslijper-Kater, S. N., Friedrich, W. N., & Corwin, D. L. (2004). Sexual knowledge and emotional reaction as indicators of sexual abuse in young children: Theory and research challenges. *Child Abuse & Neglect*, 28, 1007–1017.
- Brodie v. The King, [1936] S.C.R. 188.
- Bruck, M., Ceci, S. J., & Hembrooke, H. (2001). Reliability and credibility of young children's reports: From research to policy and practice. In R. Bull (Ed.), *Children and the law: The essential readings* (pp. 87–123). Boston: Blackwell.

- Charmaz, K. (2006). *Constructing grounded theory*. Thousand Oaks, CA: Sage.
- Coghlan v. Cumberland, 1 Ch. 704 (1898).
- Connolly, D. A., Price, H. L., Lavoie, J. A., & Gordon, H. M. (2008). Perceptions and predictors of children's credibility of a unique event and an instance of a repeated event. *Law and Human Behavior, 32*, 92–112.
- Connolly, D. A., & Read, J. D. (2003). Remembering historical child sexual abuse. *Criminal Law Quarterly, 47*, 438–480.
- Connolly, D. A., & Read, J. D. (2006). Delayed prosecutions of historic child sexual abuse: Analyses of 2064 Canadian criminal complaints. *Law and Human Behavior, 30*, 409–434.
- Conte, J. R., Sorenson, E., Fogarty, L., & Rosa, J. D. (1991). Evaluating children's reports of sexual abuse: Results from a survey of professionals. *American Journal of Orthopsychiatry, 78*, 428–437.
- Cossins, A. (2002). The hearsay rule and delayed complaints of child sexual abuse: The law and the evidence. *Psychiatry, Psychology, and Law, 9*, 163–176.
- Desmarais, S. L., Read, J. D., Rojas, E., & Vander Veen, S. (2005, June). *Juror knowledge of memory processes in forensic contexts: Which topics require expert opinion evidence?* Paper presented at the meeting of the Canadian Psychological Association, Montreal, Quebec, Canada.
- Fisher, R. P., & Cutler, B. L. (1995). The relation between consistency and accuracy of eyewitness testimony. In G. Davies, S. Lloyd-Bostock, M. McMurrin, & C. Wilson (Eds.), *Psychology, law, and criminal justice: International developments in research and practice* (pp. 21–28). Berlin, Germany: De Gruyter.
- Fivush, R. (1997). Event memory in early childhood. In N. Cowan (Ed.), *The development of memory in childhood* (pp. 139–161). Hove, England: United Kingdom Psychology Press.
- Fivush, R. (2002). Scripts, schemas, and memory of trauma. In N. L. Stein, P. J. Bauer, & M. Rabinowitz (Eds.), *Representation, memory, and development: Essays in honor of Jean Mandler* (pp. 53–74). Mahwah, NJ: Erlbaum.
- Golding, J. M., Fryman, H. M., Marshall, D. F., & Yozwiak, J. A. (2003). Big girls don't cry: The effect of child witness demeanor on juror decisions in a child sexual abuse trial. *Child Abuse & Neglect, 27*, 1311–1321.
- Golding, J. M., Segó, R. P., & Sanchez, S. A. (1999). The effect of multiple childhood sexual assaults on mock-jurors' perceptions of repressed memories. *Behavioral Sciences & the Law, 17*, 483–493.
- Goodman, G. S., Batterman-Faunce, J. M., & Schaaf, J. M. (2002). Nearly 4 years after an event: Children's eyewitness memory and adults' perceptions of children's accuracy. *Child Abuse & Neglect, 26*, 849–884.
- Guadagno, B. L., Powell, M., & Wright, R. (2006). Police officers' and legal professionals' perceptions regarding how children are, and should be, questioned about repeated abuse. *Psychiatry, Psychology, and Law, 13*, 251–260.
- Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2001). Inside the judicial mind. *Cornell Law Review, 86*, 777–830.
- Home Affairs Committee. (2002). *The conduct of investigations into past cases of abuse in children's homes: Fourth report of Session 2001–02*. London: Author. Retrieved May 1, 2008, from <http://www.publications.parliament.uk/pa/cm200102/cmselect/cmhaff/cmhaff.htm>
- Hudson, J., & Nelson, K. (1983). Effects of script structure on children's story recall. *Developmental Psychology, 19*, 625–635.
- Kahneman, D. (2002). *Maps of bounded rationality: A perspective on intuitive judgment and choice*. Princeton, NJ: Department of Psychology, Princeton University.

- Retrieved October 17, 2008, from http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahnemann-lecture.pdf
- Kassin, S. M., & Sukel, H. (1997). Coerced confessions and the jury: An experimental test of the "harmless error" rule. *Law and Human Behavior, 21*, 27–46.
- Kaufman, G., Drevland, G. C. B., Wessel, E., Overskeid, G., & Magnussen, S. (2003). The importance of being earnest: Displayed emotions and witness credibility. *Applied Cognitive Psychology, 17*, 21–34.
- Kendall-Tackett, K. A., Williams, L. M., & Finkelhor, D. (1993). Impact of sexual abuse on children: A review and synthesis of recent empirical studies. *Psychological Bulletin, 113*, 164–180.
- Lamb, M. E. (1999). Child witnesses: Recent research on children's accounts of forensically relevant experiences. *Applied Developmental Science, 3*, 2–5.
- Lewis, P. (2006). *Delayed prosecution for childhood sexual abuse*. New York: Oxford University Press.
- London, K., Bruck, M., Ceci, S. J., & Shuman, D. W. (2005). Disclosure of child sexual abuse: What does the research tell us about the ways that children tell? *Psychology, Public Policy, and Law, 11*, 194–226.
- 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 & Adolescent Psychiatry, 46*, 162–170.
- McCarron, A. L., Ridgway, S., & Williams, A. (2004). The truth and lie story: Developing a tool for assessing child witnesses' ability to differentiate between truth and lies. *Child Abuse Review, 13*, 42–50.
- Meyers, J. E. B. (1996). A decade of international reform to accommodate child witnesses. *Criminal Justice and Behavior, 23*, 402–422.
- Myles-Worsley, M., Cromer, C. C., & Dodd, D. H. (1986). Children's preschool script reconstruction: Reliance on general knowledge as memory fades. *Developmental Psychology, 22*, 22–30.
- Ogloff, J. R. P., & Rose, V. G. (2005). The comprehension of judicial instructions. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 407–444). New York: Guilford Press.
- Pezdek, K., Morrow, A., Blandon-Gitlin, I., Goodman, G. S., Quas, J. A., Saywitz, K. J., et al. (2004). Deception detection in children: Event familiarity affects criterion-based content analysis ratings. *Journal of Applied Psychology, 89*, 119–126.
- Platt, D. (2006). Threshold decisions: How social workers prioritize referrals of child concern. *Child Abuse Review, 15*, 4–18.
- R. v. Ay (1994) B.C.J. No. 2024. (B.C.C.A.)
- R. v. B.(G) (1990) 2 S.C.R. 30.
- R. v. D.(D) (2000) S.C.C. 43 S.C.R. 275.
- R. v. D.R.E. (2002) B.C.J. No. 2879. (B.C.C.A.)
- R. v. Gagnon (2006) 1 S.C.R. 621, 2006 SCC 17.
- R. v. King (2005) O.J. No. 3546. (Ont. C.A.)
- R. v. Marquard (1993) 4 S.C.R. 223.
- R. v. R. (J. W.) (2001) 151 C.C.C. (3d) 236; BCJ No. 39.
- R. v. Tzarfin (2005) O.J. No. 3531. (Ont. C.A.)
- R. v. W.(R.) (1992) 2 S.C.R. 122.
- Regan, P. C., & Baker, S. J. (1998). The impact of child witness demeanor on perceived credibility and trial outcome in sexual abuse cases. *Journal of Family Violence, 13*, 187–195.
- Rennie, D. L., Phillips, J. R., & Quartaro, J. K. (1988). Grounded theory: A promising approach for conceptualization in psychology? *Canadian Psychology, 29*, 139–150.

- Rose, M., Nadler, J., & Clark, J. (2006). Appropriately upset? Emotion norms and perceptions of crime victims. *Law and Human Behavior, 30*, 203–219.
- Salekin, R. T., Ogloff, J. R., McFarland, C., & Rogers, R. (1995). Influencing jurors' perceptions of guilt: Expression of emotionality during testimony. *Behavioral Sciences and the Law, 13*, 293–305.
- Slackman, E., & Nelson, K. (1984). Acquisition of an unfamiliar script story form by young children. *Child Development, 55*, 329–340.
- Sugarman, D. B., & Boney-McCoy, S. (1997). Impact of expert testimony on the believability of repressed memories. *Violence and Victims, 12*, 115–126.
- Summit, R. C. (1983). The child sexual abuse accommodation syndrome. *Child Abuse & Neglect, 7*, 177–193.
- Tabachnick, B. G., & Pope, K. S. (1997). Therapist responses to recovered and never-forgotten memories of child sex abuse. *Violence Against Women, 3*, 348–360.
- Wessel, E., Drevland, G. C. B., Eilertsen, D. E., & Magnussen, S. (2006). Credibility of the emotional witness: A study of ratings by court judges. *Law and Human Behavior, 30*, 221–230.
- Wiener, R. L., Bornstein, B. H., & Voss, A. (2006). Emotion and the law: A framework for inquiry. *Law and Human Behavior, 30*, 231–248.

Received May 19, 2008

Revision received January 5, 2009

Accepted January 9, 2009 ■