CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

BIAS IN THE COURTROOM?
ADVERSARIAL ALLEGIANCE AND EXPERT TESTIMONY

A thesis submitted in partial fulfillment of the requirements
For the degree of Master of Arts in Psychology
General-Experimental

by

Jeana L. Arter

August 2012
The thesis of Jeana L. Arter is approved:

__________________________________________  ____________________________
Scott W. Plunkett, Ph.D.                   Date

__________________________________________  ____________________________
Andrew T. Ainsworth, Ph.D.                   Date

__________________________________________  ____________________________
Bradley D. McAuliff, J.D., Ph.D., Chair     Date

California State University, Northridge
DEDICATION

This thesis is dedicated to my family. Without all of you, I would not be where I am today. Thank you for your continued love and support.
ACKNOWLEDGMENT

I would like to thank my committee members who supported my efforts in writing this thesis.

To my chair, Dr. Bradley McAuliff: Thank you for assisting me with my thesis. More importantly, thank you for encouraging me to do better, not only in academia, but also in life. I would not be here today without your continued support and encouragement.

To Dr. Andrew Ainsworth: Thank you for every opportunity you have provided me over the past four years. It has been a wonderful pleasure working with and learning from you.

To Dr. Scott Plunkett: Thank you agreeing to serve on my committee. Although my thesis is not your specific area of expertise, you provided much-appreciated feedback and insight. Thank you for greeting everyone in Sierra Hall with a warm smile. Your kindness does not go unnoticed.

To Joshua Lapin: Thank you for your eagerness to help, not only with my thesis but also with all McLab tasks. More importantly, thank you for your friendship, or always knowing when to listen and when to make me laugh. I wish you great success at Pepperdine and in your future career.

This research was supported in part by grants from the Office of Graduate Studies at California State University, Northridge and the American Psychology-Law Society Grants-in-Aid Program.
# TABLE OF CONTENTS

Signature Page ............................................................ ii  
Dedication ................................................................. iii  
Acknowledgment ......................................................... iv  
List of Tables .............................................................. vii  
Abstract ................................................................. viii  

CHAPTER I – INTRODUCTION 1  
Expert Testimony on Children’s Suggestibility 2  
Statement of the Problem 4  

CHAPTER II – REVIEW OF LITERATURE 5  
Adversarial Allegiance 5  
Anchoring Effect 6  
Confirmation Bias 7  
Mere Exposure Effect 9  
Norm of Reciprocity 10  
Overview and Hypotheses 11  

CHAPTER III – METHODOLOGY 13  
Expert Sample 13  
Materials 15  
Dependent Measures 16  
Short Answer Coding 17  
Procedure 18  

CHAPTER IV – RESULTS 20  
Manipulation Check 20  
Section One: MANOVA Analyses 21  
Willingness to Testify 21  
Expert Testimony Focus 22  
Perceived Expert Bias 24  
Section Two: MANOVA Analyses 24  
Child Victim Accuracy 25  
Child Victim Trustworthiness 25  
Police Interview Quality 26  

CHAPTER V – DISCUSSION 27  
Willingness to Testify 27  
Actual and Perceived Expert Bias 29  
Evaluations of Child Victim and Police Interview 31  
Limitations 33  
Implications and Future Research 35  
Conclusion 37
LIST OF TABLES

Table 1 – Demographic Information for Expert Sample 14
Table 2 – Self-Reported Frequency for Experts Who Were Asked, Had Agreed, and Actually Testified 15
Table 3 – Likert-type Items for Experts’ Evaluations of Child Victim and Police Interview 17
Table 4 – Frequencies for Police Interview Variables 23
ABSTRACT

BIAS IN THE COURTROOM?
ADVERSARIAL ALLEGAINCE AND EXPERT TESTIMONY

By:

Jeana L. Arter

Master of Arts in Psychology,
General-Experimental

The use of expert witnesses in jury trials is common. Experts are assumed to be neutral, objective parties who disseminate information to jurors; however, the adversarial system may bias experts in favor of the side that retained them (prosecution or defense). This thesis examined the potential influence of adversarial allegiance on experts’ evaluations in a simulated child sexual abuse case. A national sample of 89 experts reviewed a police interview of an alleged 5-year-old female victim. Retaining party (prosecution or defense) and interview suggestibility (low or high) varied across experts. Experts were more willing to testify for the defense than prosecution when interview suggestibility was high, but equally willing to testify for either party when interview suggestibility was low. Experts focused on aspects of the evidence that favored the retaining party only when interview suggestibility was low and blind raters’ perceptions of expert bias mirrored those findings. Experts’ perceptions of child victim accuracy, trustworthiness, and interview quality only varied as a function of interview suggestibility: child victim accuracy, trustworthiness, and interview quality were lower in the high versus low interview suggestibility condition. Results suggest that expert witnesses in child sexual abuse cases may have an initial bias towards the retaining party when interview
suggestibility is low. However, experts appear to remain unbiased when evaluating the child victim and police interview. This information is valuable in the courtroom setting for attorneys and judges alike. Knowing that experts may remain unbiased supports the use of experts in cases to improve the quality and fairness of jurors’ decisions at trial.
CHAPTER I

INTRODUCTION

The use of expert witnesses in jury trials is common, especially when evidence is technical or difficult to understand (Grosocup, Penrod, Studebaker, Huss, & O’Neil, 2002). Experts are assumed to be neutral, objective parties who disseminate information to jurors; however, it is possible that the adversarial system may bias experts in favor of the side that retained them (prosecution or defense). This tendency is known as adversarial allegiance (Murrie, Boccaccini, Turner, Meeks, Woods, & Tussey, 2008).

Basic information about expert testimony, such as frequency and type, has been difficult to obtain (Krafka, Dunn, Johnson, Cecil, & Miletich, 2002). The absence of public information about expert testimony and cases in which experts have intentionally provided incorrect information has intensified lay beliefs that experts are susceptible to bias (for a review, see Hagen, 1997). Judges also struggle with issues surrounding expert testimony (Mnookin, 2008). In *Daubert v. Merrell-Dow Pharmaceuticals* (1993), the U.S. Supreme Court held that judges must evaluate the relevance and reliability of expert testimony when determining its admissibility. Not surprisingly, judges are more likely to scrutinize the quality of expert testimony now than before *Daubert* (Dixon & Gill, 2002; Merlino, Murray, & Richardson, 2008).

Despite public and legal skepticism about expert testimony, very little research has examined adversarial allegiance in expert testimony. In beginning to do so, researchers must first distinguish between an expert who *intentionally* presents false information in court versus one who *selectively* presents information that favors the retaining party (Otto, 1989). My thesis will focus on the latter by examining the
unintentional ways that adversarial allegiance may bias expert testimony on children’s suggestibility in a simulated sexual abuse case.

**Expert Testimony on Children’s Suggestibility**

Suggestibility is the extent to which certain cognitive, social and developmental factors influence an individual’s ability to encode, store, retrieve, and report an event (Ceci & Bruck, 1993). Scholars have studied suggestibility since the early 1900s; however, an unprecedented surge in research occurred after extreme allegations of child sexual abuse in preschool daycare settings surfaced in the early 1990s (e.g., McMartin Preschool case in California, Kelly Michaels case in New Jersey). Since that time, the scientific understanding of how the accuracy of memory can be influenced by suggestive questions (Ceci, Ross, & Toglia, 1987; Warren & Lane, 1995) has increased dramatically. For example, scholars in the field know that younger (versus older) children answering leading (versus open-ended questions) from a high (versus low or equal) authority interviewer are less accurate.

Psychologists have developed various protocols to overcome the limitations inherent in interviewing children about alleged abuse. One of the most well-researched protocols is the National Institute of Child Health and Human Development Investigative Interview (NICHD; Lamb, Orbach, Hershkowitz, Esplin, & Horowitz, 2007). This structured protocol has been shown to minimize the effects of interviewer authority by including an introductory phase in which the child is given permission to say, “I don’t understand” and to correct the interviewer when appropriate. The NICHD protocol also includes a rapport-building phase that is designed to create a comfortable and supportive interview environment for the child. The substantive phase of the interview minimizes
the interviewer’s reliance on direct or leading questions by presenting the child a series of non-suggestive invitations (e. g., “Tell me about why you’re here today…”) and open-ended prompts (e. g., “Tell me more about that…” and “What happened next?”).

Social scientific interest in witness suggestibility and interview protocols has spread into the legal system as experts have begun to impart their knowledge to jurors in court. Although jurors appear to understand age-related trends in suggestibility, they lack knowledge about how other factors such as leading questions and interviewer authority can increase suggestibility (McAuliff & Kovera, 2007; Quas, Thompson, & Clarke-Stewart, 2005). Expert testimony on these issues has been shown to improve jurors’ understanding and decision-making.

One recent study examined the effects of interview quality and presence of expert testimony on mock jurors’ evaluations of a child victim’s forensic interview (Buck, London, & Wright, 2011). Participants read a summary of a child sexual abuse case in which forensic interview quality (good, typical, or poor) and expert testimony on proper interviewing techniques (present, absent) varied. When expert testimony was present and interview quality was typical or poor, mock jurors acquitted the defendant more often than when expert testimony was absent. Mock jurors who did not receive expert testimony failed to consider interview quality when reaching their verdicts. Additional research has shown that expert testimony can improve jurors’ understanding of children’s testimonial demeanor (Kovera, Gresham, Borgida, Gray, & Regan, 1997), prompted reports (Laimon & Poole, 2008), hearsay testimony (Nunez, Gray, & Buck, 2012), and recovered memories (Buck & Warren, 2009).
The need for and helpfulness of expert testimony on children’s suggestibility make it an appropriate backdrop for examining the potential effects of adversarial allegiance. Do experts in a child sexual abuse case selectively focus on aspects of the child’s interview that favor the retaining party? Do prosecution-retained experts evaluate child accuracy and interview quality more favorably than experts retained by the defense or vice versa? My thesis will provide answers to these important questions.

**Statement of the Problem**

The purpose of this thesis is to examine the potential role of adversarial allegiance in expert testimony. In other words, do experts’ opinions and evaluations of evidence vary as a function of being asked to testify on behalf of the prosecution versus defense? I examined this research question by presenting a national sample of witness suggestibility experts a description of a police officer interview of an alleged child sexual abuse victim. Using a fully-crossed factorial design, I varied the retaining party (prosecution, defense) and interview suggestibility (low, high) to determine the effects of those variables on experts’ evaluations of evidence quality and other case-related judgments.
CHAPTER II
REVIEW OF LITERATURE

Adversarial Allegiance

Systematic research on adversarial allegiance is surprisingly scant and has revealed mixed findings. One of the earliest studies analyzed 42 paired mental health evaluations of flood survivor litigants at two different points in time (Zusman & Simon, 1983). Plaintiff-retained experts who evaluated litigants three years after the flood observed higher levels of psychological distress than defense-retained experts who evaluated litigants 8 years later. The authors concluded that plaintiff-retained experts were biased in favor of the plaintiff; however, the confound between retaining party and time of interview obfuscates this conclusion because psychological distress should improve over time.

Using a more sophisticated methodology, researchers in a second study surveyed Florida forensic evaluators to determine whether being asked to review a defendant’s case file by the prosecution, defense, or court affected their judgments of whether the defendant was not guilty by reason of insanity (Beckham, Annis, & Gusafson, 1989). Differences across the three conditions were not statistically significant. Otto (1989) conducted a similar experiment with clinical graduate students who evaluated a civil and criminal case. Students rendered guilty verdicts in the criminal case more frequently when asked to testify by the prosecution versus defense; however, no differences in the civil case emerged as a function of retaining party.

More recent research has focused on adversarial allegiance in sexually violent predator proceedings by comparing the scores within and between evaluators retained by
the prosecution versus defense. Using STATIC-99 scores, Murrie and colleagues (2009) observed a strong correlation between evaluators’ scores and the retaining party. For experts on the same side, the difference on the STATIC-99 was less than 1 point. Yet for experts on opposing sides, the difference on the STATIC-99 was larger by an average of 3 points. The poor interrater agreement between opposing experts suggests adversarial allegiance.

This thesis will help advance this growing body of research by examining the role of adversarial allegiance in a different type of case (child sexual abuse) and by utilizing a different type of expert evaluation (child accuracy and interview quality). Before doing so, it is imperative to consider different psychological mechanisms that may contribute to adversarial allegiance.

**Psychological Mechanisms**

**Anchoring Effect**

One potential psychological mechanism underlying adversarial allegiance is the anchoring effect, which is the tendency to rely too heavily on initial information (the anchor) and inadequately adjust a preliminary decision once additional information is introduced (Tversky & Kahneman, 1974).

Researchers have explored anchoring effects within the domains of general knowledge, probability estimates, and legal decisions. Tversky and Kahneman (1974) conducted seminal work on the effects of anchors on probability estimates. They observed that arbitrary numbers provided to students influenced their subsequent estimates of the number of African countries within the United Nations. More recently, Englich, Musswiler, and Strack (2006) examined anchoring among attorneys and judges
by providing them case files containing information about a rape trial. Later when asked whether they thought the defendant’s sentence would be higher or lower than a certain number of years (1 year versus 3 years), participants exposed to the higher number provided longer defendant sentences compared to the lower number. These studies demonstrate the potential influence of anchors on subsequent decision-making tasks.

What role might anchoring effects play in adversarial allegiance? When attorneys consult an expert regarding potential testimony, they provide case-related facts and evidence to help inform the expert’s decision whether to testify. In all likelihood, this information is biased in favor of the attorney’s client and, therefore, is not representative of the entire case. Experts may anchor on this initial information, form their opinion, and evaluate additional evidence in a biased manner. Even though the anchoring heuristic may be common and even useful, it can lead to systematic errors in the evaluation of subsequent information (Tversky & Kahneman, 1974). Once anchored on the initial case information presented by an attorney, experts may disregard or overlook additional information that fails to support their opinions.

**Confirmation Bias**

Confirmation bias is closely related to anchoring and also might contribute to adversarial allegiance. Confirmation bias is the tendency for people to seek information that confirms their pre-existing beliefs, regardless of whether those beliefs are correct (Klayman, 1995). In other words, people selectively attend to information that matches what they expect in a given situation and underutilize or even ignore information that contradicts their expectations.

Confirmation bias is present in a variety of decision making contexts. Strohmer
and Shivy (1994) demonstrated confirmatory bias in clinical judgments by creating a staged therapy interaction between a confederate and clinical doctoral students. The confederate exhibited 10 behaviors that were consistent with low self-control (e.g., lack of warmth and lack of sensitivity) and 10 behaviors that were not (e.g., social stability and rule conscience). Students who were asked to test the clinical hypothesis that the confederate “patient” lacked self-control recalled a larger number of the confirmatory versus disconfirmatory behaviors. Classic social psychological research by Kelley (1950) documented similar effects in the classroom setting. Students who interacted with a visiting professor described in advance as being “very warm” versus “rather cold” evaluated the professor more positively and remembered more expectancy-consistent behavior than their counterparts. As demonstrated in these studies, confirmation bias can affect both the encoding and recall of information.

In the adversarial legal system, each attorney has a duty to zealously advocate for his/her client. Deciding whether and who to retain as an expert is a critical part of advocacy, and attorneys are likely to contact experts who support their clients’ positions. Experts’ pre-existing beliefs can dramatically influence how they process information and the conclusions they draw (Klayman, 1995). For example, if a prosecuting attorney contacts an expert in child witness memory, it is readily apparent to the expert that the prosecution wants to bolster the child’s credibility to ensure a guilty verdict against the defendant. This knowledge, in turn, may cause the expert to seek or selectively attend to evidence that confirms the child’s credibility (and hence defendant’s guilt) while ignoring information that contradicts this conclusion.
**Mere Exposure Effect**

A third psychological mechanism related to adversarial allegiance is the mere exposure effect, which states that people tend to evaluate a stimulus more favorably the more often they are exposed to that stimulus (Zajonc, 1968). Simply put, increased exposure equals increased favorability, and this effect appears to generalize across a wide-range of stimuli. Monahan, Murphy, and Zajonc (2000) studied the mere exposure effect by using Chinese ideographs. Participants in one condition viewed five Chinese ideographs five times in random order, and participants in a second condition viewed 25 different Chinese ideographs. Participants who viewed the same five ideographs five times reported being in a better mood at the end of the experiment than those who viewed 25 different ideographs. Thus, exposure to the same neutral stimuli repeatedly can have a positive effect on one’s emotions. A follow-up study by the same researchers showed virtually identical effects of exposure frequency on participants’ favorability ratings of the stimuli as well as improved mood ratings (Monahan et al., 2000).

Another study indicated that the mere exposure effect also increases processing fluency (Reber, Winkielman, & Schwarz, 1998). The researchers exposed participants to 20 neutral, but distorted drawings (e.g., plane or horse) for 2 seconds. Right before the target drawing, participants viewed another drawing that did or did not match the target drawing. Participants who viewed the matching drawings not only recognized the targets more quickly, they also rated the targets as being prettier than participants who viewed the non-matching drawings. In summary, increased exposure can lead to improved mood, more favorable ratings, and increased processing fluency.

Even if the information given to an expert in a case is neutral and objective
(which, for the reasons stated above, is unlikely), the fact that the expert reviews this information repeatedly when preparing his/her opinion is likely to influence the expert’s mood, evaluation of the materials, and rate of processing. These effects may be further exacerbated by an expert’s lack of contact with and information from opposing counsel.

**Norm of Reciprocity**

Lastly, the norm of reciprocity also may contribute to adversarial allegiance. The norm of reciprocity is the mutual exchange of favors (Gouldner, 1960). In other words, people are nice to those who are nice to them. This norm is part of a moral code in which a person who has received a favor feels socially indebted to return the favor to restore equilibrium in the relationship (Gouldner, 1960). The effects of the norm of reciprocity are usually subtle and unconscious in nature.

Greenglass (1969) examined the effects of prior help on individuals’ willingness to help others. Female participants completed a two-part study for the chance to win a gift certificate. Part one consisted of each participant constructing boxes and being rated for motivation to complete the task; whoever constructed the most boxes and was viewed by her peers (who were actually confederates) as the most motivated to complete the task won the gift certificate. Part two involved each participant rating her “peers” on their motivation to complete a task for the chance to win a gift certificate. Participants who received help rated their peers more favorably than did participants who were not helped. Greenglass asserted this outcome occurred because participants who were helped by their peers felt obligated to return the favor by rating their peers more favorably.

More recently, researchers examined the effect of reciprocity on undergraduates (Shen, Wan, & Wyer, 2011). Undergraduates were offered candy; once the undergraduate
accepted the candy they were asked to complete a survey. The results indicated that the willingness to complete the survey had a direct correlation with the number of candies the undergraduate accepted. Shen and colleagues concluded that small gifts (candy) do affect an individual’s willingness to complete a favor due to the feeling of being indebted.

The norm of reciprocity has implications for the relationship between an attorney and expert. Experts may feel obligated to repay their debt of being asked to testify or being paid for their opinion by providing favorable evidence for the retaining party. According to Gouldner (1960), the debt to be repaid is directly proportionate to the benefit received. If experts are paid sizably to testify, this should increase their bias in favor of the retaining party. Even if experts are not paid, simply being asked to testify may be personally and professionally validating, thus prompting reciprocity as well.

**Overview and Hypotheses**

This study examined the potential influence of adversarial allegiance on witness suggestibility experts’ evaluations in a simulated child sexual abuse case. Experts were asked by the prosecution or defense to read a description of a police officer’s high or low suggestibility interview of a 5-year-old girl who alleged inappropriate sexual touching by her stepfather. Experts then indicated their willingness to testify, described the issues they would address in their testimony, and rated the child victim’s accuracy, trustworthiness, and interview quality.

Consistent with adversarial allegiance, I hypothesized that experts’ evaluations would favor the side (prosecution or defense) that retained them across the following key dependent measures: willingness to testify (Hypothesis 1), expert testimony focus (Hypothesis 2), perceived expert testimony bias (Hypothesis 3), child victim accuracy
and trustworthiness (Hypothesis 4), and interview quality (Hypothesis 5). However, I expected that the main effect on these dependent measures would vary as a function of the interview suggestibility manipulation.

Specifically, I predicted that when interview suggestibility was high, prosecution-retained experts would be less likely to testify, to provide testimony that favored the prosecution, and to rate the child victim and interview quality more positively than defense-retained experts. However, I predicted that when interview suggestibility was low, defense-retained experts would be less likely to testify, to provide testimony that favored the defense, and to rate the child victim and interview quality more negatively than prosecution-retained experts.
CHAPTER III

METHODOLOGY

Expert Sample

I randomly selected five states each from three geographic regions (15 states total) in the western, central, and eastern United States to identify potential expert witnesses to participate in the study. For each state, I used the terms “child sexual abuse,” “witness suggestibility,” and “forensic interviews,” to search online expert databases maintained by courts (e.g., www.lasuperiorcourt.org) or by contacting local court clerks directly. In addition, I contacted attendees of the 26th Annual San Diego Conference on Child and Family Maltreatment with expertise in witness suggestibility and forensic interviewing. The final list of potential respondents included 330 experts from 32 different states. Experts were contacted via email and snail mail (Dillman, 2000) and were offered $25 in exchange for their participation.

Forty-five percent \((n = 147)\) of the experts responded in some way. Eighty-nine provided complete responses, and 58 reported they were not qualified to serve as an expert on the topic of witness suggestibility. In part, the large number of “not qualified” respondents was due to the fact that court databases classify experts using very broad categories (e.g., “child sexual abuse”) and not by specialty (e.g., medical doctor, sexual abuse nurse examiner). As a result, our initial sample included experts qualified to testify in child sexual abuse cases, but who were not experts on child witness suggestibility specifically. Thirteen invitation emails or letters were returned as undeliverable (4%) and 170 experts did not respond in any way (51%).

The demographic information for the final expert sample appears in Table 1.
Thirty-three percent (33%) of the experts resided in California, 9% resided in Texas, and the remaining 58% were from 30 other states. The majority of respondents were male (61%) and the average age was 55 years old.

Table 1

*Demographic Information for Expert Sample*

<table>
<thead>
<tr>
<th>Demographic Variables</th>
<th>% of Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>61.0</td>
</tr>
<tr>
<td>Female</td>
<td>39.0</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Under 40</td>
<td>13.0</td>
</tr>
<tr>
<td>41-45</td>
<td>8.0</td>
</tr>
<tr>
<td>46-50</td>
<td>6.0</td>
</tr>
<tr>
<td>51-55</td>
<td>8.0</td>
</tr>
<tr>
<td>56-60</td>
<td>14.0</td>
</tr>
<tr>
<td>61-65</td>
<td>11.0</td>
</tr>
<tr>
<td>Over 65</td>
<td>16.0</td>
</tr>
<tr>
<td>Degree</td>
<td></td>
</tr>
<tr>
<td>BA- Nursing</td>
<td>8.5</td>
</tr>
<tr>
<td>MA</td>
<td>12.9</td>
</tr>
<tr>
<td>PhD/Psy D</td>
<td>52.9</td>
</tr>
<tr>
<td>MD</td>
<td>16.5</td>
</tr>
<tr>
<td>JD</td>
<td>1.2</td>
</tr>
<tr>
<td>&gt; 1 advanced degree</td>
<td>4.7</td>
</tr>
<tr>
<td>Specialization</td>
<td></td>
</tr>
<tr>
<td>Clinical Psychology</td>
<td>21.3</td>
</tr>
<tr>
<td>Child Development</td>
<td>18.0</td>
</tr>
<tr>
<td>Forensic Interviewer</td>
<td>11.2</td>
</tr>
<tr>
<td>Social Work/Law Enforcement</td>
<td>4.5</td>
</tr>
<tr>
<td>Nurse/Medical Doctor</td>
<td>22.5</td>
</tr>
<tr>
<td>Other</td>
<td>7.9</td>
</tr>
</tbody>
</table>

Note. *N* = 89. Percentages may not sum to 100 because of missing data.

Within the demographic portion of the survey, respondents were asked how many times they have been asked to testify as an expert witness, how many times they agreed to testify, and how many times they actually testified. Lastly, the respondents indicated the number of times they had testified for the prosecution or defense. Responses are
shown in Table 2.

Table 2

**Self-Reported Frequency for Experts Who Were Asked, Had Agreed, and Actually Testified**

<table>
<thead>
<tr>
<th>Number of Times</th>
<th>Experts Asked to Testify N (%)</th>
<th>Experts Agreed to Testify N (%)</th>
<th>Experts Actually Testified N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10</td>
<td>27 (30.0)</td>
<td>30 (34.0)</td>
<td>25 (28.0)</td>
</tr>
<tr>
<td>11-35</td>
<td>9 (10.0)</td>
<td>7 (8.0)</td>
<td>10 (11.0)</td>
</tr>
<tr>
<td>36-75</td>
<td>9 (10.0)</td>
<td>7 (8.0)</td>
<td>6 (7.0)</td>
</tr>
<tr>
<td>76-300</td>
<td>13 (15.0)</td>
<td>11 (12.0)</td>
<td>9 (10.0)</td>
</tr>
<tr>
<td>301-1000</td>
<td>3 (.0)</td>
<td>1 (.0)</td>
<td>1 (.0)</td>
</tr>
<tr>
<td>Total</td>
<td>61 (69.0)</td>
<td>56 (63.0)</td>
<td>51 (57.0)</td>
</tr>
</tbody>
</table>

Note. N = 89. Percentages may not sum to 100 because of missing data.

**Materials**

Respondents read that they were being sought for potential expert testimony on behalf of the prosecution or defense in a child sexual abuse case. They read one of two descriptions of a police interview of an alleged child victim. The *high suggestibility interview* condition described an interview conducted by a police officer in her full uniform using a standard interrogation room with nothing but a table and two chairs facing each other. The police officer was described as relying heavily on direct questions about the event such as how long the child had known the alleged perpetrator and whether he had ever been abusive in the past. In the *low suggestibility interview* condition, participants were informed the police officer changed out of her formal uniform and conducted the interview in a “child friendly” room with comfortable furniture and games/toys. She was described as using a series of open-ended questions followed by direct questions when necessary to clarify the central details of the alleged
abuse such as what the alleged perpetrator said and did. See Appendix for the complete interview description.

**Dependent Measures**

After reading the interview description, respondents indicated their willingness to serve as an expert on witness suggestibility in the case by answering a 7-point Likert-type scale (1 = Certainly No, 7 = Certainly Yes). An open-ended question followed, asking experts to describe what aspects of the child’s memory and the police officer’s interview they would choose to focus on if called to testify in the case. Experts then rated the child and interview by answering a series of 7-point Likert-type items (1 = Not at All, 7 = Extremely). These questions ranged from how accurate the expert believed the child witness was to how suggestive the officer’s interview was (see Table 3). Experts concluded their participation by completing a series of demographic items about their training and prior experience, reported in Table 1 and Table 2.
Table 3

*Likert-type Items for Experts’ Evaluations of Child Victim and Police Interview*

| 1. In your opinion, how accurate is Anna’s memory for the events that she described? |
| 2. How reliable was Anna’s testimony? |
| 3. How easily influenced is Anna by suggestive or misleading questions from an interviewer? ® |
| 4. How likely is it that Anna was telling the truth? |
| 5. Was Anna motivated to lie about what happened? ® |
| 6. Was Officer Olsen’s interview of Anna fair? |
| 7. Was Officer Olsen’s interview of Anna successful at obtaining the truth? |
| 8. Was Officer Olsen’s interview of Anna biased? ® |
| 9. Officer Olsen believed that Anna was sexually abused. |
| 10. Was Officer Olsen’s interview of Anna suggestive? ® |

Note. ® denotes item was reversed coded.

For the purpose of data analyses, the experts’ ratings of the child and interview were averaged into three composite variables: child victim accuracy, child victim trustworthiness, and interview quality. The composite variable for child victim accuracy consisted of three Likert-type items: perceived accuracy, reliability, and whether the child victim was influenced by the suggestive questions (Cronbach’s alpha = .76). The composite variable for child victim trustworthiness consisted of two Likert-type items: truthfulness and motivation to lie (Cronbach’s alpha = .56). The composite variable for interview quality consisted of four Likert-type items: fairness, success at obtaining the truth, suggestibility (Cronbach’s alpha = .82).

**Short Answer Coding**

Two graduate research assistants who were blind to condition and the study’s hypotheses coded the open-ended question that measured what aspects of the child’s memory and police officer’s interview the experts’ testimony would focus on. First, both research assistants read the open-ended answers to identify common trends in experts’ responses. A list of items experts mentioned most frequently was created, resulting in 36
items (12 directly pertaining to the interview suggestibility manipulation itself) listed in
Table 4. Next, the two raters read each short answer first to determine whether each
expert’s response favored the defense or prosecution overall (-1 = Pro-defense, 0 =
Neutral, 1 = Pro-prosecution) and second to code the content of each expert’s response
using the 36 pre-determined items (1 = Present, 0 = Absent). Then raters evaluated
whether each individual statement favored the defense or prosecution (-1 = Pro-defense,
0 = Neutral, 1 = Pro-prosecution). Examples of pro-prosecution statements include:
“Child friendly room,” “Introduction of Susan instead of Officer Olsen,” “Mother viewed
interview from a different room,” and “The officer took time to build report with the
child.” Examples of pro-defense statements include: “Interview took place in an
interrogation room,” “Officer appeared in full uniform,” “The report stated the child
used the word penis,” “The child was not instructed that it was okay to not answer all the
questions.” The lowest percent-agreement between the two raters was 96 for a single item
and 97 percent-agreement overall.

Procedure

I adapted several procedures from the Total Design Method of survey research
(Dillman, 2000) to increase the response rate including multiple mailings and reminder
notices at predetermined time intervals. Previous studies (Kovera & McAuliff, 2000;
McAuliff & Kovera, 2007) using similar methods with experts and legal professionals
have achieved response rates ranging from 40% to 60%.

I began by sending a contact postcard that described the study and outlined future
correspondence. One week later, I mailed and emailed the invitation letter to each expert.
If they choose to participate, they were asked to click on a web link provided in the
email/letter that took them to the survey posted on www.psychsurveys.org. A similar link was provided for those who did not wish to participate in the survey. One week after the invitation letter, I mailed/emailed a reminder postcard to experts who had not completed the survey. One month after the initial postcard was sent, I mailed/emailed a final letter that emphasized the importance of receiving completed surveys from as many experts as possible in order for the sample results to generalize to the population.
CHAPTER IV

RESULTS

All analyses were conducted using SPSS 20.0 for Windows. The first section of analyses focuses on the experts’ willingness to testify, what features of the police interview experts focused on when reviewing the case materials, and raters’ perceptions of expert bias. The second section focuses on the experts’ evaluations of the child victim’s accuracy, trustworthiness, and the quality of the police interview.

Manipulation Check

Prior to full data analyses, manipulation checks were analyzed to ensure experts were sensitive the retaining party (prosecution, defense) manipulation. Within the prosecution-retained condition, participants were more likely to correctly identify the prosecution as the retaining party (83%) compared to the defense (17%), $X^2 (1, N = 41) = 17.78, p = .001$, $\theta = .66$. Also, within the defense-retained condition, experts were more likely to correctly identify the defense as the retaining party (79%) than the prosecution (21%), $X^2 (1, N = 47) = 15.51, p = .001$, $\theta = .57$. All participants were included in the data analyses to ensure adequate statistical power. As seen in the child victim accuracy, trustworthiness, and interview quality analyses below, experts were also sensitive to the interview suggestibility manipulation.

To ensure there were no systematic difference in the frequency with which attorneys testified for the prosecution or defense across all experimental conditions, I conducted a 2 Retaining Party (Prosecution or Defense) x 2 Interview Suggestibility (Low or High) multivariate analysis of variance (MANOVA) on the self-report data from experts regarding how often they testify for each side (prosecution and defense). There
were no differences across conditions for (a) retaining party, \( F(1, 77) = .16, p = \text{ns} \), partial eta-squared = .04; (b) interview suggestibility, \( F(1, 77) = .09, p = \text{ns} \), partial eta-squared = .03; (c) or the interaction, \( F(1, 77) = 1.80, p = \text{ns} \), partial eta-squared = .05, indicating random assignment was successful.

**Section One: MANOVA Analyses**

I conducted a 2 Retaining Party (Prosecution, Defense) x 2 Interview Suggestibility (Low, High) multivariate analysis of variance (MANOVA) on the data for the following measures: willingness to testify, what features of the police interview the experts focused on, and the raters’ perceptions of expert bias. I used Pillai’s Trace criterion multivariate statistic to test the significance of all main effects and interactions. The main effect for retaining party was not significant, \( F(1, 76) = 2.16, p = \text{ns} \), partial eta-squared = .08. The main effect for interview suggestibility was statically significant, \( F(1, 76) = 9.96, p = .001 \), partial eta-squared = .30. The Retaining Party X Interview Suggestibility interaction was statistically significant also, \( F(1, 76) = 4.88, p = .004 \), partial eta-squared = .17. I followed up each significant multivariate effect using univariate \( F \)-tests for each of the three dependent measures.

**ANOVA Analyses**

**Willingness to testify.** A two-way univariate ANOVA examined the effects of retaining party and interview suggestibility on experts’ willingness to testify. The main effect of retaining party was not statistically significant, \( F(1, 85) = 7.50, p = \text{ns} \), partial eta-squared = .02. Experts were highly willing to testify in the case whether asked by the prosecution (\( M = 5.90 \)) or the defense (\( M = 5.56 \)). The main effect of interview suggestibility was not statistically significant, \( F(1, 85) = .15, p = \text{ns} \), partial eta-squared =
.02. Experts were highly willing to testify in the case whether interview suggestibility was high \( (M = 5.81) \) or low \( (M = 5.71) \).

The Retaining Party X Interview Suggestibility interaction was statistically significant, \( F(1, 85) = 4.46, p = .04 \), partial eta-squared = .05. Follow-up comparisons revealed that when interview suggestibility was high, experts were more willing to testify for the defense \( (M = 6.32) \) than the prosecution \( (M = 4.81) \), \( F(1, 41) = 5.39, p = .03 \), partial eta-squared = .12. However, when interview suggestibility was low, experts were equally willing to testify for either the prosecution \( (M = 5.90) \) or defense \( (M = 5.56) \), \( F(1, 44) = .34, p = ns \), partial eta squared = .08.

**Expert testimony focus.** A two-way univariate ANOVA examined the effects of retaining party and interview suggestibility on expert testimony focus. The main effect of retaining party was not statistically significant, \( F(1, 73) = 2.02, p = ns \), partial eta-squared = .03. \( (M = -1.29 \) prosecution, \( M = -2.41 \) defense). Both the prosecution and defense-retained experts focused more on aspects of the case that favored the defense and disfavored the prosecution (e.g., weaknesses in the police officer’s interview and the child victim’s report). The main effect of interview suggestibility was statistically significant, \( F(1, 73) = 30.36, p = .001 \), partial eta-squared = .29. Experts who read the high suggestibility version of the interview commented on more pro-defense aspects of the case \( (M = -.59) \) than experts who read the low suggestibility version \( (M = -.25) \).

This main effect was qualified by a statistically significant interaction between retaining party and interview suggestibility, \( F(1, 73) = 4.52, p = .04 \), partial eta-squared = .06. When interview suggestibility was low, experts focused on aspects of the evidence that favored the retaining party, \( F(1, 38) = 6.66, p = .01 \), partial eta-squared = .15; \( M = \)
.76 prosecution, \( M = -1.37 \) defense). When interview suggestibility was high, the difference in experts’ focus was not significantly significant, \( F(1, 35) = .24, p = ns, \) partial eta-squared = .01; \( (M = -3.82 \) prosecution, \( M = -3.40 \) defense).

Table 4  
*Frequencies for Police Interview Variables*

<table>
<thead>
<tr>
<th>Officer &amp; Interview Room</th>
<th>Pro Prosecution</th>
<th>Pro Defense</th>
<th>Neutral</th>
<th>Total</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience*</td>
<td>0 (0%)</td>
<td>1 (100%)</td>
<td>0 (0%)</td>
<td>1</td>
<td>.003</td>
</tr>
<tr>
<td>Training</td>
<td>2 (2.2%)</td>
<td>10 (11.2%)</td>
<td>1 (1.1%)</td>
<td>13</td>
<td>.043</td>
</tr>
<tr>
<td>Dress*</td>
<td>4 (4.5%)</td>
<td>23 (25.8%)</td>
<td>1 (1.1%)</td>
<td>28</td>
<td>.094</td>
</tr>
<tr>
<td>Introduction*</td>
<td>2 (2.2%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>2</td>
<td>.007</td>
</tr>
<tr>
<td>Atmosphere*</td>
<td>5 (5.6%)</td>
<td>19 (21.3%)</td>
<td>1 (1.1%)</td>
<td>25</td>
<td>.084</td>
</tr>
<tr>
<td>Mom’s presence*</td>
<td>1 (1.1%)</td>
<td>16 (18%)</td>
<td>0 (0%)</td>
<td>17</td>
<td>.057</td>
</tr>
<tr>
<td>Interview</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Don’t Know”*</td>
<td>3 (3.4%)</td>
<td>2 (2.2%)</td>
<td>0 (0%)</td>
<td>5</td>
<td>.017</td>
</tr>
<tr>
<td>“Don’t Understand”*</td>
<td>2 (2.2%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>2</td>
<td>.007</td>
</tr>
<tr>
<td>Answer all questions*</td>
<td>0 (0%)</td>
<td>4 (4.5%)</td>
<td>0 (0%)</td>
<td>4</td>
<td>.013</td>
</tr>
<tr>
<td>Initial reluctance*</td>
<td>2 (2.2%)</td>
<td>5 (5.6%)</td>
<td>0 (0%)</td>
<td>7</td>
<td>.023</td>
</tr>
<tr>
<td>Encouragement*</td>
<td>1 (1.1%)</td>
<td>3 (3.4%)</td>
<td>0 (0%)</td>
<td>4</td>
<td>.013</td>
</tr>
<tr>
<td>Use of protocol*</td>
<td>4 (4.5%)</td>
<td>4 (4.5%)</td>
<td>0 (0%)</td>
<td>8</td>
<td>.027</td>
</tr>
<tr>
<td>Question type*</td>
<td>16 (18%)</td>
<td>24 (27%)</td>
<td>3 (3.4%)</td>
<td>43</td>
<td>.144</td>
</tr>
<tr>
<td>Age-appropriate questions</td>
<td>0 (0%)</td>
<td>7 (7.9%)</td>
<td>1 (1.1%)</td>
<td>8</td>
<td>.027</td>
</tr>
<tr>
<td>Use of props</td>
<td>1 (1.1%)</td>
<td>2 (2.2%)</td>
<td>2 (2.2%)</td>
<td>5</td>
<td>.017</td>
</tr>
<tr>
<td>Video recording</td>
<td>0 (0%)</td>
<td>12 (13.5%)</td>
<td>2 (2.2%)</td>
<td>14</td>
<td>.047</td>
</tr>
<tr>
<td>Content</td>
<td>0 (0%)</td>
<td>1 (1.1%)</td>
<td>0 (0%)</td>
<td>1</td>
<td>.003</td>
</tr>
<tr>
<td>Use of the word “penis”</td>
<td>0 (0%)</td>
<td>7 (7.9%)</td>
<td>0 (0%)</td>
<td>7</td>
<td>.023</td>
</tr>
<tr>
<td>Length of interview</td>
<td>0 (0%)</td>
<td>4 (4.5%)</td>
<td>0 (0%)</td>
<td>4</td>
<td>.013</td>
</tr>
<tr>
<td>Disclosure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delay</td>
<td>0 (0%)</td>
<td>5 (5.6%)</td>
<td>3 (3.4%)</td>
<td>8</td>
<td>.027</td>
</tr>
<tr>
<td>Prior disclosure (who)</td>
<td>2 (2.2%)</td>
<td>2 (2.2%)</td>
<td>2 (2.2%)</td>
<td>6</td>
<td>.020</td>
</tr>
<tr>
<td>Spontaneous disclosure</td>
<td>0 (0%)</td>
<td>11 (12.4%)</td>
<td>0 (0%)</td>
<td>11</td>
<td>.037</td>
</tr>
<tr>
<td>Prior conversations</td>
<td>1 (1.1%)</td>
<td>1 (1.1%)</td>
<td>1 (1.1%)</td>
<td>3</td>
<td>.010</td>
</tr>
<tr>
<td>Precipitating events (family)</td>
<td>0 (0%)</td>
<td>1 (1.1%)</td>
<td>3 (3.4%)</td>
<td>4</td>
<td>.013</td>
</tr>
<tr>
<td>Precipitating events (friends)</td>
<td>0 (0%)</td>
<td>1 (1.1%)</td>
<td>0 (0%)</td>
<td>1</td>
<td>.003</td>
</tr>
<tr>
<td>Child’s demeanor</td>
<td>0 (0%)</td>
<td>3 (3.4%)</td>
<td>1 (1.1%)</td>
<td>4</td>
<td>.013</td>
</tr>
<tr>
<td>Abuse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History</td>
<td>2 (2.2%)</td>
<td>4 (4.5%)</td>
<td>3 (3.4%)</td>
<td>9</td>
<td>.030</td>
</tr>
<tr>
<td>Defendant prior record</td>
<td>1 (1.1%)</td>
<td>1 (1.1%)</td>
<td>1 (1.1%)</td>
<td>3</td>
<td>.010</td>
</tr>
<tr>
<td>Act itself</td>
<td>3 (3.4%)</td>
<td>8 (9%)</td>
<td>3 (3.4%)</td>
<td>14</td>
<td>.047</td>
</tr>
<tr>
<td>Physical evidence</td>
<td>0 (0%)</td>
<td>3 (3.4%)</td>
<td>1 (1.1%)</td>
<td>4</td>
<td>.013</td>
</tr>
<tr>
<td>Witnesses</td>
<td>1 (1.1%)</td>
<td>2 (2.2%)</td>
<td>4 (4.5%)</td>
<td>7</td>
<td>.023</td>
</tr>
<tr>
<td>Medical exam</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>2 (2.2%)</td>
<td>2</td>
<td>.007</td>
</tr>
<tr>
<td>Forensic interview</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1 (1.1%)</td>
<td>1</td>
<td>.003</td>
</tr>
<tr>
<td>Age</td>
<td>2 (2.2%)</td>
<td>9 (10.1%)</td>
<td>1 (1.1%)</td>
<td>12</td>
<td>.040</td>
</tr>
<tr>
<td>Truth/Lie</td>
<td>1 (1.1%)</td>
<td>2 (2.2%)</td>
<td>0 (0%)</td>
<td>3</td>
<td>.010</td>
</tr>
<tr>
<td>Memory</td>
<td>3 (3.4%)</td>
<td>5 (5.6%)</td>
<td>1 (1.1%)</td>
<td>9</td>
<td>.030</td>
</tr>
</tbody>
</table>

Note. * denotes variables manipulated in the interview.
**Perceived expert bias.** A two-way univariate ANOVA examined the effects of retaining party and interview suggestibility on raters’ perceptions of expert bias. The main effect of retaining party was statistically significant, $F(1, 86) = 5.62, p = .02$, partial eta-squared = .07. Raters perceived more pro-defense bias in expert testimony when experts were retained by the defense ($M = -.62$) versus prosecution ($M = -.21$). The main effect of interview suggestibility also was statistically significant, $F(1, 86) = 19.26, p = .001$, partial eta-squared = .20. Raters perceived more pro-defense bias in expert testimony for the high versus low suggestibility interview conditions ($Ms = -.78$ and -.12, respectively).

Both of these main effects were qualified by a statistically significant Retaining Party X Interview Suggestibility interaction, $F(1, 86) = 4.58, p = .04$, partial eta-squared = .06. Follow-up comparisons revealed that when interview suggestibility was low, raters perceived more expert testimony bias in favor of the retaining party, $F(1, 41) = 9.08, p = .01$, partial eta-squared = .18; ($Ms = .24$ prosecution and -.45 defense). However, when interview suggestibility was high, this difference was not statically significant, $F(1, 35) = .03, p = ns$, partial eta-squared = .00; ($Ms = -.76$ prosecution and -.80 defense).

**Section Two: MANOVA Analyses**

I conducted a 2 Retaining Party (Prosecution, Defense) x 2 Interview Suggestibility (Low, High) multivariate analysis of variance (MANOVA) on the data for the following measures: child victim accuracy, child victim trustworthiness, and interview quality. I used Pillai’s Trace criterion multivariate statistic to test the significance of all main effects and interactions. The main effect for retaining party was not significant, $F(1, 83) = .62, p = ns$, partial eta-squared = .02. The main effect for
interview suggestibility was statistically significant, $F(1, 86) = 10.15, p = .001$, partial eta-squared = .27. The Retaining Party X Interview Suggestibility interaction was not significant, $F(1, 83) = 1.22, p = ns$, partial eta-squared = .04. I followed up the significant multivariate effect using univariate $F$-tests for each of the three dependent measures.

**ANOVA Analyses**

**Child victim accuracy.** A two-way univariate ANOVA examined the effects of retaining party and interview suggestibility on experts’ evaluations of child victim accuracy. The main effect of retaining party was not statistically significant, $F(1, 83) = .84, p = ns$, partial eta-squared = .01. Experts’ perceptions of child victim accuracy were neutral; prosecution ($M = 4.00$) versus defense ($M = 3.83$). The main effect of interview suggestibility was statistically significant, $F(1, 83) = 10.20, p = .002$, partial eta-squared = .11. When interview suggestibility was high, the experts rated the child victim as less accurate ($M = 3.60$) than when interview suggestibility was low ($M = 4.20$). The Retaining Party X Interview Suggestibility interaction was not statistically significant, $F(1, 83) = 3.53, p = ns$, partial eta-squared = .04.

**Child victim trustworthiness.** A two-way univariate ANOVA examined the effects of retaining party and interview suggestibility on experts’ evaluations of child victim trustworthiness. The main effect of retaining party was not statistically significant, $F(1, 83) = .07, p = ns$, partial eta-squared = .00. Experts’ perceptions of child victim trustworthiness were relatively neutral ($M = 4.17$ prosecution, $M = 4.13$ defense). The main effect of interview suggestibility was only marginally significant, $F(1, 83) = 3.24, p = .07$, partial eta-squared = .04. However, the trend was in the expected direction; when suggestibility was high experts rated the child victim as less trustworthy ($M = 4.00$)
compared to when suggestibility was low ($M = 4.31$). The Retaining Party X Interview Suggestibility interaction was not statistically significant, $F(1, 83) = 2.50, p = \text{ns}$, partial eta-squared = .03.

**Police interview quality.** A two-way univariate ANOVA was conducted to examine the effects of retaining party and interview suggestibility on experts’ evaluations of police interview quality. The main effect of retaining party was not statistically significant, $F(1, 83) = 1.35, p = \text{ns}$, partial eta-squared = .07. Both means fell just below the scale midpoint ($M = 3.88$ prosecution, $M = 3.63$ defense). The main effect of interview suggestibility was statistically significant, $F(1, 83) = 24.79, p = .001$, partial eta-squared = .23. Experts in the high suggestibility condition evaluated the interview quality less favorably ($M = 3.20$) than experts in the low suggestibility condition ($M = 4.30$). The Retaining Party X Interview Suggestibility interaction was not statistically significant, $F(1, 83) = 1.91, p = \text{ns}$, partial eta-squared = .02.
CHAPTER V
DISCUSSION

This thesis was designed to examine the potential role of adversarial allegiance on expert testimony. In other words, do experts’ evaluations of evidence vary as a function of being asked to testify for one side versus the other in court? Previous research on adversarial allegiance is limited to only four studies over the past two decades. The results of my thesis are consistent with the findings of Murrie et al. (2009). They observed low levels of interrater agreement between evaluators in sexual violent predator proceedings who were retained by opposing sides. My data revealed similar bias based on adversarial allegiance; however, the suggestibility of the police interview used to question the alleged child sexual abuse victim moderated this effect. In general, experts focused on aspects of the evidence that favored the retaining party when suggestibility was low, and blind raters’ perceptions of expert bias mirrored these findings. Despite experts’ biased review of the evidence, the effects of adversarial allegiance did not carry over into experts’ evaluations of the child victim’s accuracy, trustworthiness, or police interview quality. Experts’ ratings on these measures did not vary as a function of retaining party and fell near the mid-point on these scales. In other words, the retaining party may influence what aspects of the case an expert focuses on, but it does not appear to detrimentally affect the conclusions they draw.

Discussion of the Findings

Willingness to Testify

I predicted that willingness to testify would vary as a function of retaining party and police interview suggestibility. Specifically, I hypothesized that when interview
suggestibility was high, prosecution-retained experts would be less willing to testify than defense-retained experts, and that when interview suggestibility was low, defense-retained experts would be less willing to testify than prosecution-retained experts. This hypothesis was partially supported. Results indicated that experts were more willing to testify for the defense versus prosecution when interview suggestibility was high, but equally willing to testify for either party when interview suggestibility was low.

This finding raises the question of why defense (and not prosecution) experts were more willing to testify when interview suggestibility was high (and not low). One possibility is that the flaws or negative aspects of the police interview are simply easier for experts to detect than the strengths or positive aspects. This explanation is supported by a closer examination of Table 4, which reveals that defense experts provided a larger number of responses overall when asked what aspects of the child’s memory and the police interview they would focus on if called to testify than did prosecution experts. If flaws are easier to detect than strengths, this provides more information for the expert to discuss when testifying for the defense compared to the prosecution. The obvious nature of the interview flaws may also appeal to defense experts because their testimony should be less contentious and less subject to scrutiny by the opposing side (Mnookin, 2008).

A second explanation may stem from the “helpfulness” standard governing the admissibility of expert testimony set forth in the Federal Rules of Evidence. Rule 702 states that expert testimony is admissible only if it is relevant and assists the trier of fact to understand the evidence. It is possible that when interview suggestibility was high (and consequently the child’s report was less accurate), defense experts believed their testimony would be the most helpful and were more willing to testify than prosecution-
retained experts. This may especially be the case given the defendant’s presumption of innocence in a criminal trial. Not only would the expert testimony be helpful, but also it would be necessary to prevent the wrongful conviction of an innocent person based on the child’s inaccurate report. Consistent with the explanation, Otto (1989) explained that adversarial bias tends to vary along with the outcome of the situation: the more detrimental the consequences, the stronger the bias.

Research has shown that expert testimony on witness suggestibility is needed in court. McAuliff and Kovera (2007) surveyed potential jurors and found that laypeople did not understand factors that moderate witness suggestibility, such as interviewer authority, which was manipulated in the present study. Additional research has demonstrated that expert testimony can improve jurors’ decisions with regard to forensic interviewing also. Buck and colleagues (2011) found when expert testimony was present and interview quality was poor, participants acquitted the defendant more often than when expert testimony was absent.

**Actual and Perceived Expert Bias**

The hypotheses for actual and perceived expert bias were partially supported. I predicted an interaction between retaining party and interview suggestibility such that defense-retained experts would focus on more evidence that favored the defense than prosecution experts, but the opposite pattern of effect would emerge when interview suggestibility was low: prosecution-retained experts would focus more on evidence that favored the prosecution than defense-retained experts. The interaction was significant; however, adversarial allegiance only emerged when interview suggestibility was low.

Although there has been no previous research on the potential role of adversarial
allegiance and interview suggestibility specifically, these findings are consistent with prior research involving other types of psychological testimony by experts (Otto, 1989; Zusman & Simon, 1983). Zusman and Simon found that plaintiff-retained mental health practitioners who evaluated the plaintiffs’ psychological distress saw less improvement than did defense-retained practitioners. Similarly, clinical graduate students in an experiment by Otto were more likely to believe the defendant was guilty when they reviewed case materials that were ostensibly provided by the defense versus the prosecution. The results from the present study add to this growing body of literature documenting the influence of adversarial allegiance on experts’ decisions. In my study, experts focused on the substandard evidence when interview suggestibility was high; however, when suggestibility was low, experts sought out information that favored the retaining party. In addition, raters who were blind to the study’s hypotheses and experimental conditions perceived bias that mirrored these findings.

That said, it is important to point out that evidence of adversarial allegiance was only present when interview suggestibility was low. This could be a result of confirmation bias. Individuals utilize confirmation bias as a way to reduce cognitive load (Klayman, 1995). If, as previously discussed, flaws in the high suggestibility interview were easier to detect than strengths in the low suggestibility interview, then experts in the latter condition may have relied on confirmation bias to help reduce their cognitive loads. The possibility of confirmation bias is particularly troubling when we consider that many experts consider themselves “neutral” parties in the adversarial system. Commons, Miller, and Gutheil (2004) asked experts if they believed that certain aspects of expert testimony were considered biased (e.g., payment, recurring testimony for one side). Most
experts considered themselves unbiased and able to correct any biases should they exist. If experts fail to realize they are biased, it is unlikely they will correct for adversarial bias by actively trying to remain objective when reviewing evidence and rendering opinions in court.

The presence of adversarial allegiance when interview suggestibility was low is even more interesting when we consider that experts were more willing to testify when interview suggestibility was high. As stated above, perhaps the willingness to testify when interview suggestibility was high may have been due to the experts’ obligation to ensure a fair trial. However, when interview suggestibility was low, experts may have relied on confirmation bias to reduce their cognitive loads when faced with the challenge of evaluating a sound police interview of the child victim.

It is also worthwhile to note the differences in the area of focus between prosecution-retained experts and defense-retained experts, as reported in Table 4. Prosecution-retained experts focused on the interview and instructions given to the child, (“Don’t know, don’t understand” instruction, and question type) whereas defense-retained experts focused on the police officer and location of the interview (police officer’s dress and interview room).

**Evaluations of Child Victim and Police Interview**

The hypotheses for experts’ evaluations of the child victim and police interview were not supported. I predicted that when interview suggestibility was high, prosecution-retained experts would rate the child victim and police interview more favorably compared to defense-retained experts, but when interview suggestibility was low, defense-retained experts would rate the child victim and police interview more negatively.
compared to prosecution-retained experts. Unlike the experts’ willingness to testify and evaluations, their perceptions of the child victim and the police interview did not vary as a function of retaining party and police interview suggestibility. However, the expected trend for suggestibility was present. When police interview suggestibility was high, all experts, irrespective of which side retained them, rendered less favorable evaluations of the child victim’s accuracy and police interview quality compared to when the interview suggestibility was low.

These findings are the most encouraging of the results. They suggest that despite the presence of adversarial allegiance towards retaining party in how experts focus on evidence, experts appear to possess the ability to control their bias. According to Wilson and Brekke (1994) there are two different classes of mental errors: (a) the failure to know or apply the information and (b) mental contamination (unwanted judgment or emotion). The latter applies to experts and the presence of adversarial allegiance. When experts first identify what aspects of the evidence to focus on for trial, perhaps their unwanted judgments or behaviors contaminate their focus. This outcome is especially disconcerting because studies have shown that individuals are often unable to avoid mental contamination (Wilson & Brekke, 1994). In other words, individuals are often unable to ignore outside influences when making decisions. Wegener and Petty (1995) found that individuals are able to possess opposing theories about the same situation. It is possible for an individual to perceive the same situation as having a positive effect on one judgment and a negative effect on another. Further, individuals who possess opposing theories actively try to remove bias by adjusting their judgments in the opposite direction (Wegener & Petty, 1995). These results could help explain the data from my thesis in
terms of child victim and interview quality evaluations. With the presence of adversarial allegiance during the initial examination of evidence, it is possible that experts attempt to remove bias from their subsequent evaluations of the child victim and interview quality.

Perhaps experts are able to control biases when rendering an evaluation for a hired party. At the same time, this interpretation should be evaluated with extreme caution because experts may simply want to appear unbiased despite actual bias. According to Leary (2004), when placed in unfamiliar situations individuals are acutely aware of the impression they are creating. Thus, experts may attempt to correct any biases during testimony. It is also possible that experts’ evaluations were influenced by a socially desirable response bias. Experts may have disguised their bias for the retaining-party by deliberately evaluating the child victim and interview quality in a neutral manner. One reason for this response bias is that individuals, as social creatures, seek to create and maintain favorable impressions of themselves in other people. Because experts are expected to be neutral parties, they may fear negative evaluation if they appear biased. Given the variability in experts’ responses and that several items were reverse coded to detect a response bias (if present), the more plausible explanation seems to be that experts controlled their bias when evaluating the child and interview.

**Limitations**

Certain limitations must be addressed before considering the implications of the results. First, the data were collected using self-report surveys online. Personal email addresses were used to reduce the chance that someone other than the expert would complete the survey, but there is no way to ensure that each participant who completed the survey was the intended participant. Second, the original sample of experts solicited
was too broad. There were a large number of respondents who indicated they were not qualified to complete the survey (mainly doctors and sexual assault nurse examiners). On the positive side, my efforts resulted in a final sample of experts from across the United States. Third, the description of the police interview in the stimulus materials lacked ecological validity. The survey was short in order to prevent respondent fatigue and an excessive dropout rate. Experts in actual cases would receive complete interviews and not be allowed to testify about ultimate opinion issues, such as a particular child witness’s accuracy or trustworthiness. Even though experts are not allowed to testify about ultimate issues, they still form personal opinions about the child’s accuracy, trustworthiness, and interview quality and may even be asked to express these opinions through the use of hypotheticals in their testimony. The fourth limitation to the present study is the low reliability of the child victim trustworthiness composite variable; the results should be interpreted with caution. There are a few possible reasons that could explain the low Cronbach’s alpha = .56: (a) the composite variable was comprised of only two Likert-type items. Perhaps looking at each of those items separately would provide a more accurate measure of the child victim trustworthiness; (b) there could have been conflict between the two Likert-type items chosen for the child victim trustworthiness composite variable. I used how truthful the experts saw the child and how motivated the experts believed she was to lie. Although unlikely, it is possible that a person could be motivated to lie and still appear truthful; (c) The child’s age was not manipulated within the study. It is possible that experts were not comfortable indicating a 5-year-old child is untrustworthy, which would explain why a majority of the experts rated her at the midpoint of the scale (4 out 7).
Implications and Future Research

There are two main implications in the present research. First, experts were able to distinguish between low versus high interview suggestibility and previous research has shown that jurors lack a sufficient understanding of factors, such as interview suggestibility, that affect witness suggestibility (McAuliff & Kovera, 2007; Quas et al., 2005). Therefore, expert testimony should satisfy the helpfulness standard Rule 702 of the Federal Rules of Evidence and be admissible in court. This is important because some states allow expert testimony on witness suggestibility in court whereas others do not (McAuliff & Kovera, 2007).

Second, previous research has shown that expert testimony can help improve jurors’ decision-making (Buck et al., 2011), but the question is whether adversarial allegiance might bias the information experts’ disseminate to jurors.

Even though adversarial bias did not affect experts’ evaluations of the child victim’s accuracy, trustworthiness, or police interview quality it did affect their willingness to testify and focus on the case evidence. These findings are consistent with previous research.

Murrie et al. (2009) observed a strong correlation for evaluators of sexually violent predator evaluations and retaining party. Otto (1989) found that prosecution-retained participants in a simulated criminal case were more likely to convict compared to defense-retained participants.

Even though research suggests that adversarial allegiance exists among experts, it is impossible and undesirable to eliminate expert testimony from jury trails altogether. According to Kraftka et al. (2002) seventy-three percent of trials conducted during 1998
had expert witnesses for both the defense and the prosecution. The average number of expert witnesses the prosecution called was 2.47, while the defense called 1.87 on average. Experts possess specialized knowledge that can help jurors better understand evidence and decide cases (Kassin, Tubb, Hosch, & Memon, 2001; McAuliff & Kovera, 2007). Yet something must be done to prevent or correct adversarial allegiance in expert testimony.

One potential remedy is the use of court-appointed experts (Mnookin, 2008). Court-appointed experts testify on behalf of the court as opposed to one side or the other. Typically, the judge is responsible for soliciting the expert’s involvement and serves as the conduit of information from attorneys for both sides to the expert. The use of court-appointed experts should reduce, if not eliminate, the influence of the psychological mechanisms (anchoring effect, confirmation bias, mere exposure, and norm of reciprocity) previously discussed. Receiving case-related information from the judge on behalf of both parties should eliminate anchoring and confirmation effects because the information provided will be more balanced and representative of the entire case as opposed to just one side. The mere exposure effect may still influence experts’ evaluations, but this effect should be spread more equally across both parties compared to receiving information from only one party. Finally, the norm of reciprocity should shift away from either party to the court because the expert is testifying on behalf of (and paid by) the court. By reducing or possibility eliminating adversarial allegiance the judicial system would prove to ensure individuals receive a fair trial within the field of expert testimony.

Further research should improve upon the limitations of the current study. First,
when asking participants to complete an online survey (www.psychsurveys.org), a check box indicating that the person completing the survey was the intended recipient. Second, access to specific databases identifying experts’ specialties would eliminate contacting individuals who consider themselves unqualified to render opinions on forensic interviews. Third, providing a complete transcript of the police interview rather than a summary would enhance ecological validity.

**Conclusion**

This thesis examined the effects of adversarial allegiance on expert testimony in a simulated child sexual abuse case. The data suggest that adversarial allegiance influences what aspects of the evidence experts choose to focus on. However, experts were able to remain unbiased when rendering evaluations of key evidence for the retaining party. These findings are critical not only in the field of psychology, but also in the legal field as well. Lawyers and judges can be more confident that the majority of information rendered during a jury trial is unbiased and honest to the best of the expert’s ability. Further, these results support the continued use of experts in trials, making difficult concepts easier to understand for laypeople serving as jurors. Future research including a court-appointed expert condition will assist in further understanding why bias exists and how bias can be reduced, or preferably, eliminated.
REFERENCES


doi:10.1177/0146167205282152


Wilson, T. D., & Brekke N. (1994). Mental contamination and mental correction:


APPENDIX

Police Interview

Susan Olsen is a police officer at the Sherman Oaks Police Department. Officer Olsen has been a member of the homicide squad for nearly ten years until August 2010 when she accepted a transfer to the sexual victimization unit. She is responsible for responding to and investigating allegations of sexual crimes and that a large part of her responsibility is interviewing alleged victims and witnesses.

Officer Olsen reported that around 5:30 p.m. on August 20, 2010, she received a radio dispatch stating that a woman was bringing her daughter to the station for questioning about a sexually abusive incident involving the woman’s husband. Officer Olsen returned to the station and met with Sonja Richardson and her daughter Anna when they arrived.

*High Suggestibility*

Officer Olsen was dressed in her full uniform, complete with her badge, gun, and ammunition belt. She conducted the interview in a standard interrogation room that was empty except for a table with two chairs facing each other and one extra seat for an observer. Mrs. Richardson was seated next to Anna during the entire interview. Officer Olsen began the interview by introducing herself as “Officer Olsen” and then explained to Anna what they were going to do. She told Anna it was very important to answer all of the interview questions. The interview consisted mainly of direct questions about information such as what her stepfather did and said during the incident. According to Officer Olsen, Anna was reluctant to talk at first but after she and Sonja encouraged Anna a little bit she began to open up. Anna said she had gone to the living room to watch television with her stepfather and that he made her touch his penis and “move it around.” When asked, Anna said the defendant had never done anything like that to her before. Officer Olsen estimated that the investigative
interview of Anna lasted about 30 minutes.

*Low Suggestibility*

Shortly before their arrival, Officer Olsen changed out of her uniform into her street clothes and removed her gun and ammunition belt. She conducted the interview in a special interrogation room with comfortable chairs and some games and puzzles for children. Mrs. Richardson observed the interview through a one-way mirror in an adjacent room. Officer Olsen testified that she began the interview by introducing herself as “Susan” and the explaining to Anna what they were going to do. She told Anna it was very important to say “I don’t know” or “I don’t understand” if she did not know the answer or did not understand any of the interview questions. The interview consisted mainly of open-ended questions followed by direct questions about information mentioned by Anna, such as what her stepfather did and said during the incident. According to Officer Olsen, Anna was reluctant to talk at first but after a little bit she began to open up. Anna said she had gone to the living room to watch television with her stepfather and that he made her touch his penis and “move it around.” When asked, Anna said the defendant had never done anything like that to her before. Officer Olsen estimated that the investigative interview of Anna lasted about 30 minutes.