# Consistency and Credibility of Intimate Partner Abuse Reports: An Examination of Pathological or Generalisable Phenomena

by

Sarah L. Desmarais

B.A. (Hons), University of Guelph, 2003
M.A., Simon Fraser University, 2005

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

In the

Department of Psychology



SIMON FRASER UNIVERSITY

Spring 2008

# APPROVAL

| | |
|---|---|
| **Name:** | Sarah Desmarais |
| **Degree:** | Doctor of Philosophy (Department of Psychology) |
| **Title of Thesis:** | Consistency and Credibility of Intimate Partner Abuse Reports: An Examination of Pathological or Generalisable Phenomena |
| **Chair:** | Dr. Cathy McFarland<br>Professor |
| **Examining Committee:** | Dr. J. Donald Read<br>Senior Supervisor<br>Professor |
| | Dr. Marlene Moretti<br>Supervisor<br>Professor |
| | Dr. Kevin Douglas<br>Supervisor<br>Assistant Professor |
| **Internal Examiner:** | Dr. Margaret Jackson<br>Professor<br>School of Criminology |
| **External Examiner:** | Dr. Neil Brewer<br>Professor<br>Flinders University of South Australia |
| **Date Approved :** | April 1, 2008 |


SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

# STATEMENT OF ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Bennett Library
Simon Fraser University
Burnaby, BC, Canada

# ABSTRACT

Despite mixed research findings, there remains a pervasive belief in the legal community that testimonial inconsistencies are detrimental to eyewitness, complainant, and defendant credibility generally, and to domestic violence complainants in particular. Studied extensively in other contexts, little research has examined consistency of reports of intimate partner abuse (IPA) victimization over time and its role in perceived credibility. The first study of this dissertation compared consistency of reports of IPA victimization with consistency of everyday autobiographical memory event prevalence. Study 1 participants ($n$ = 276) completed two calendar-based online surveys approximately six weeks apart. Participants who self-identified as experiencing psychological, physical, or sexual abuse in a romantic relationship ($n$ = 138) completed questions assessing IPA victimization and participation in leisure activities (LA). A matched sample of 138 comparison participants completed only the LA questions. Few differences between report consistency of abusive experiences and everyday memory events were found. When significant effects were observed, results demonstrated differences between categories of autobiographical memory events within but not between IPA and LA reports. The second study investigated whether IPA allegations are received with scepticism, and if so, why. Study 2 participants ($n$ = 374) evaluated the effectiveness of a 'complainant' reporting on IPA victimization or LA participation during two interviews. Complainant gender and consistency of reports across repeated interviews were manipulated. Results demonstrated that consistent complainants were

evaluated more favourably than were inconsistent complainants, as were LA compared with IPA complainants. Further analyses supported the role of social categorization in evaluations of complainant effectiveness, demonstrating that when a complainant is seen as a member of one's own group, credibility is enhanced: Compared to complainants reporting on IPA victimization, LA 'complainants' were judged to be more similar and more likely to belong to the same group as participants which was associated with more positive evaluations of LA than IPA complainant effectiveness. Overall, findings suggest that although actual differences in consistency of event prevalence are few, reports of IPA victimization are received with greater scepticism than reports of everyday events. Such prejudice may contribute to disbelief of IPA allegations, potentially precluding appropriate legal intervention.

# DEDICATION

To my father.

# ACKNOWLEDGEMENTS

I would first like to thank my supervisory committee. To Don Read - Your mentorship and guidance throughout my graduate career have been invaluable. It has been a wonderful five years. To Marlene Moretti and Kevin Douglas – Thank you for your feedback, expertise, and contributions to this project. Second, I would like to acknowledge the funding that made this research possible: a Social Sciences and Humanities Research Council of Canada Doctoral Fellowship, an American Psychology-Law Society (Division 41 of the American Psychological Association) Student Grant-In-Aid, and a Natural Science and Engineering Research Council of Canada Discovery Grant (No. 7920-05). I would also like to thank Colleen Pillar, as well as Carmelina Barone, Vanessa Leung, and Kaylie Maughan for their assistance with data collection and coding. I am extremely grateful to all of the people who participated in this research, especially the women and men who experienced intimate partner abuse and generously gave of themselves. Many thanks to my friends in the Mental Health, Law and Policy Institute, colleagues at the Forensic Psychiatric Hospital, and Tonia Nicholls, in particular, for being a mentor and good friend. Finally, and most importantly, thank you to Adam for sharing this life with me and my parents for showing me how to live it.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# INTRODUCTION

## *Intimate Partner Abuse*

Comprising a pattern of physical, psychological, and sexual abuse (Tjaden & Thoennes, 2000), intimate partner abuse (IPA) is widely acknowledged as a serious social problem than can have significant, deleterious impacts on the physical and mental health of victims and costs to society (e.g., public assistance, child welfare, and other health, education, legal, and social services) (APA, 2002; Danis, 2003; Desmarais, Gibas, & Nicholls, in press; Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002). Reviewing 48 population-based studies, the World Health Organization (Krug et al., 2002) found that between 10% and 69% of women surveyed reported having experienced IPA in their lifetime. A national survey suggests that the Canadian lifetime prevalence estimate is around 25% (Statistics Canada, 1993), with approximately 4% of Canadian women and men physically assaulted by their partners each year (Laroche, 2005).

Formerly thought of as a private matter for a couple or family to resolve behind closed doors, outside of the criminal justice system (e.g., Robinson & Chandek, 2000; Russell & Light, 2006), IPA has undergone a process of criminalization over recent decades through which efforts have been made to address the issue by means of the enactment and enforcement of criminal and civil laws (Danis, 2003; Salazar, Baker, Price, & Carlin, 2003). As a result, criminal justice systems today must frequently deal with IPA allegations (Blackwell & Vaughn, 2003), as the main legal issue or in conjunction with other legal issues (APA, 2002). However, there remain numerous

obstacles which contribute to a reluctance to pursue prosecution (e.g., Bennett, Goodman, & Dutton, 1999; Blackwell & Vaughn, 2003; Rebovich, 1996; Scheppele, 1992; Stanko, 1982). Reasons cited for this reluctance include lack of physical evidence or injury, victim noncooperation or unwillingness to testify, gender discrimination, relationship status between the alleged perpetrator and victim, and *complainant credibility* (Bennett et al., 1999; Bachman & Coker, 1995; Ferraro, 1989; Lewin, Strand, & Belfrage, 2007; Scheppele, 1992; Stanko, 1982).

This dissertation reports on the findings of two studies that explored factors that may affect the perceived credibility of IPA complainants. The first study examined consistency over time of reports of IPA victimization prevalence and the second investigated the effects of report consistency, as well as social categorization (i.e., ingroup-outgroup identification), on perceived credibility. Comparing reports of IPA victimization and everyday events within each study, the overall purpose of the research was to determine whether the observed phenomena are unique to IPA event prevalence reports or instead may be characteristic of autobiographical memory reports more generally. To provide a context for the research, I begin with a discussion of perceived credibility and report consistency.

## *Perceived Credibility*

Progressing through the criminal justice system, evidence in a particular case is evaluated and, inevitably, the issue of credibility arises (Whobrey, Sales, & Elwork, 1981). Credibility may be defined as the perceiver's evaluation of the communicator's believability or worthiness of belief (O'Keefe, 2003). In cases where there is little corroborating evidence, decision-making across various stages of the criminal justice

system (e.g., whether to proceed with the investigation, to proceed to trial, and verdict at trial) hinges on complainant credibility. Complainant credibility may be particularly relevant to allegations of IPA because they often involve one party's word against the other's (e.g., Hartley, 2001; see also Connolly & Read, 2003, 2006). Consequently, attaining credibility in the criminal justice system may be one of the greatest challenges facing IPA complainants, with their credibility being questioned for a variety of reasons.

Many factors may influence these credibility evaluations, including characteristics of the statements and of the witnesses or complainants themselves. For example, Desmarais and Yarmey (2004) found that mock jurors' judgements of perceived credibility differed significantly as a function of eyewitness performance (accurate vs. inaccurate) and honesty (truthful vs. deceptive), with truthful eyewitnesses perceived as more credible when they were accurate than when they were inaccurate in their identification of a target. Investigating perceptions of children's reports of a unique, non-violent event and an instance of a repeated event, Connolly, Price, Lavoie, and Gordon (2008) demonstrated that witness age and event frequency also can affect credibility evaluations. Abshire and Bornstein (2003) found that eyewitness credibility in a simulated murder trial varied as a function of eyewitness ethnicity, regardless of mock juror ethnicity. Ruva and Bryant (2004) established that age, speech style, and question form had significant effects on university students' evaluations of witness credibility in simulated robbery and murder trial transcripts.

The factors examined in the above studies are estimator variables; that is, these are not under the control of the criminal justice system and their effects on accuracy can only ever be estimated (see Wells & Olson, 2003). They also generally are unverifiable

outside of the research context; for example, eyewitness performance as ground truth may not be confirmable. In contrast, *consistency* of information reported from police interview to courtroom testimony is one estimator variable for which we can make objective evaluations and which may be influenced by criminal justice system procedures.

## *Report Consistency*

When objective verification of facts is not possible, consistency may be used as a proxy for accuracy when evaluating the credibility or veracity of statements (e.g., McNally, 2003). Whereas report accuracy refers to the "agreement between the individual's recall and either an objective record of the event or social consensus from other participants of the event as to what occurred" (Fivush, 1993, p. 22), report consistency describes whether the "same information... [is] reported at different points in time" (van Giezen, Arensman, Spinhoven, & Wolters, 2005, pp. 936-937). Report consistency is particularly relevant to legal decision-making because key witnesses, such as the IPA complainant, may be interviewed multiple times should a case proceed to trial; for example, during the police investigation, at the preliminary inquiry, and at trial by the prosecution and defence attorneys (Haber & Haber, 2000; Memon, Vrij, & Bull, 2003).

## *Consistency and Credibility*

Research findings are equivocal regarding the statistical relationship between consistency and accuracy, as well as perceptions of the inferences that can be drawn about accuracy from levels of consistency. For example, Brewer, Potter, Fisher, Bond, and Luszcz (1999) examined the relationship between consistency and accuracy of

eyewitness testimony. Results of their first study demonstrated that potential jurors considered inconsistencies in witness statements to be the strongest marker of inaccuracy, as compared to other witness behaviours. In contrast, their Study 2 results suggested that the actual relationship between testimony consistency and accuracy is weak, if present at all. Fisher and Cutler (1995) and Penrod and Cutler (1995) found similar results. Examining beliefs regarding the relationships between 12 witness behaviours (including consistency) and testimonial accuracy, Potter and Brewer's (1999) findings echoed those results: Across samples including police detectives, prosecution and defence lawyers, and mock jurors, results emphasize the belief that inconsistencies are a strong indicator of inaccuracy. Lindsay, Lim, Marando, and Cully (1986), however, failed to find a relationship between testimonial inconsistencies and mock juror decision-making. Specifically, Lindsay et al. found that mock jurors' verdicts (guilty vs. not guilty) were not systematically affected by testimonial contradictions, despite observing lower global ratings of the prosecution witness's testimonial consistency for not guilty compared to guilty verdicts.

Considering results across studies, it is unclear whether beliefs about consistency translate into reduced credibility ratings and, perhaps more importantly, influence decision-making. In contrast to the findings reviewed above, studies of witness credibility conducted by Berman and Cutler (1996), Berman, Narby, and Cutler (1995), and Brewer and Hupfeld (2004) generally demonstrated a negative impact of inconsistency. Specifically, across various inconsistent testimony conditions (i.e., information given on the stand but not during the pre-trial investigation; contradictions between on-the-stand and pre-trial statements; contradictions made on the stand), Berman and Cutler (1996)

found that mock jurors gave lower ratings of eyewitness credibility or effectiveness, thought the defendant was less culpable, and were less likely to convict compared to the consistent testimony condition. Manipulating the centrality of the information about which the witness was inconsistent, Berman et al. (1995) similarly found that mock jurors perceived the eyewitness as less credible and the defendant as less culpable when testimony was inconsistent on both central and peripheral details. Further, mock jurors were less likely to convict when exposed to inconsistency on central details. Brewer and Hupfeld (2004) also identified main effects of testimonial consistency on both ratings of witness effectiveness and the probability that the defendant committed the crime. Brewer and Burke (2002) found that although consistency was considered by participants to be an important indicator of accuracy, its impact on judgements of guilt (rendered based on the totality of evidence presented) was negligible. In fact, witness confidence had much more pronounced effects on judgements in that project than did consistency.

### *Credibility and Gender*

In addition to characteristics of the testimony, such as consistency across repeated questioning, juror preconceptions or expectations regarding the witness (e.g., appropriate or acceptable behaviours) may influence credibility evaluations (see Leippe & Romanczyk, 1989). In particular, in any examination of credibility, the role of complainant gender, another estimator variable, cannot be ignored. Gender plays a fundamental role in social cognition and person perception (e.g., Bussey & Bandura, 1999; Deaux & Lafrance, 1998; Maccoby, 1988; Macrae & Bodenhausen, 2000; Stangor, Lynch, Duan, & Glass, 1992). Some legal literature suggests that female complainants may be disadvantaged by virtue of being female; that is, women may be seen as less

credible than men in the courtroom (e.g., Schafran, 1997; Scheppele, 1992; Stanko, 1982) and perhaps even as communicators more generally (e.g., Eagly, Makhijani, & Klonsky, 1992).

Stereotypes regarding IPA victims may further influence credibility assessments. For example, in their gender analysis of the social construction of trials, Jenkins and Davidson (1990) described how myths about women, and abused women in particular, can be used by legal counsel to manipulate perceptions of guilt in an IPA case (e.g., failure to live up to the standard of being a 'good woman' vs. being portrayed as a passive, helpless victim). Moreover, there has been a focus on IPA perpetrated by men against their female partners, even though research demonstrates IPA is not clearly divided by gender lines (Dutton & Nicholls, 2005). In a seminal study, Stets and Straus (1992) found that approximately half of the incidents of violence reported by 526 heterosexual dating couples were cases of reciprocal violence, compared to approximately one quarter being cases of male-only violence and the other quarter being female-only incidents of violence. Similarly, Desmarais, Nicholls, and Koch (2005) found that 93% of women who participated in a study on health and decision-making in intimate relationships had perpetrated some form of psychological IPA and 54%, some form of physical IPA. Research with adolescent samples provides corroborative data, demonstrating female youths' potential for aggressive behaviour against peers, as well as in dating relationships (see, for example, Moretti, Catchpole, & Odgers, 2005; Odgers, Moretti, & Reppucci, 2005; Whitaker, Halleyesus, Swahn, & Saltzman, 2007).

Despite this research, the pervasive "feminist paradigm supports the notion that domestic violence is primarily a culturally supported male enterprise and that female

violence is always defensive and reactive" (Dutton & Nicholls, 2005, p. 683). Thus, expectations regarding an IPA complainant may be dependant upon the complainant being female. Allegations of IPA made by male complainants, consequently, may be less likely to be believed than those made by female complainants for several reasons: female-perpetrated IPA is generally believed to be less injurious than male-perpetrated IPA (e.g., Follingstad, DeHart, & Green, 2004; Tjaden & Thoennes, 2000), fewer official IPA complaints are made by male victims (e.g., Straus, 1993; Tjaden & Thoennes, 2000), and claims made by male victims may be seen as gender-incongruent (e.g., McKimmie, Newton, Terry, & Schuller, 2004; Eagly et al., 1992). Whatever the underlying mechanism, such gender bias has been demonstrated in several studies. For example, Feather (1996) found that female perpetrators are believed to be less responsible for their IPA behaviours than are male perpetrators. Across identical scenarios, Hamel, Desmarais, and Nicholls (2007) found that male-perpetrated IPA was seen as more coercive than female-perpetrated IPA. Two independent surveys demonstrated that police officers rated male IPA victims as more responsible for precipitating the abusive incidents than female victims (Finn & Stalans, 1997; Stewart & Maddren, 1997).

## *The Present Research*

Little research has examined the consistency of reports of IPA victimization experiences across repeated questioning, although it has been studied extensively in other contexts, such as child sexual abuse (e.g., Peterson, Moores, & White, 2001; Saywitz & Camparo, 1998), traumatic events (e.g., Flin, Boon, Knox, & Bull, 1992; Peace & Porter, 2004; Porter & Peace, 2007), and more generally in the literature on survey methodologies (e.g., Wight & West, 1999). Desmarais, Klein, Nicholls, Read, and Koch

(2006) conducted one study that expressly examined inconsistencies in event prevalence across repeated reports of IPA victimization. Eighty-one women self-identified as IPA victims participated in two interviews conducted on average one year apart ($M_{elapsed\ time}$ = 13.57 months, $SD$ = 4.68) and completed measures including the *Revised Conflict Tactics Scale* (CTS2; Straus, Hamby, Boney-McCoy, & Sugarman, 1996). Comparisons of responses to the CTS2 items across interviews demonstrated inconsistencies in reports which, in many instances, reflected remarkably high rates of apparent recantation (e.g., 83% reported at the first interview that their partner pushed/shoved them vs. 54% at the second interview; 73% reported at the first interview that their partner grabbed them vs. 51% at the second interview). However, as report consistency was not the original focus of the study, conclusions were limited due to methodological issues including considerable variation in the delay between interviews (Range = 7 – 22 months) and the time elapsed since the abusive relationship.[1] Additionally, because autobiographical memory for events decline over time (Rubin & Wenzel, 1996) and there were no comparison data for other types of events, these decreased reports may reflect normative forgetting rather than IPA-specific losses. Nonetheless, the high rates of recantations of abusive experiences observed have obvious implications for the prosecution of IPA cases and supports the need for continued research on the topic.

Further, there remains a pervasive belief in the legal community that inconsistencies are detrimental to eyewitness, complainant, and defendant credibility generally (e.g., Glissan, 1991; Salhany, 1991; Stuesser, 1993) and IPA complainants in particular (e.g., Hartley, 2001; Scheppele, 1992). In fact, Canadian case law upholds that

[1] Only women who were no longer in the abusive relationship discussed at T1 and who completed the CTS2 with regard to the same relationship at both T1 and T2 were included in the study. Time elapsed since the abusive relationship and the first interview ranged between 0 and 34 years.

a witness may be disbelieved based on the content of the testimony, as contrasted with earlier statements. In *R. v. Burke* (1996), the Supreme Court of Canada held that "obvious inconsistencies ... [in the complainant's] testimony render the trial judge's finding of credibility unreasonable." More recently in *R. v. Gagnon* (2006), consistency of statements over time was found to be an appropriate basis upon which to judge complainant credibility. The Canada Evidence Act holds that a party can give proof that a witness's current statement is inconsistent with a prior statement (R.S., 1985, c. C-5, s. 10; 1994, c. 44, s. 86; R.S., c. E-10, s. 11), also specifying that a statement "made at other times... [may be considered] in determining whether in the opinion of the court the witness is adverse" (R.S., 1985, c. C-5, s. 9; 1994, c. 44, s. 85). Similarly, jury instructions across North American jurisdictions direct jurors to consider inconsistencies between prior statements and courtroom testimony, typically with a focus on contradictions, when evaluating witness credibility (e.g., Canadian Judicial Council Model Jury Instructions in Criminal Matters, § 11.10.2; Judicial Council of California Criminal Jury Instructions, 2007-2008, CALCRIM No. 318; Florida Standard Jury Instructions in Criminal Cases, 1987, #2.04, sub-para8; New York Criminal Jury Instructions 2d, Credibility of Witnesses-Inconsistent Statements, 2007; Sixth Circuit Criminal Pattern Jury Instructions, No. 107, 2005).[2]

For these reasons, this dissertation examined factors that may influence evaluations of IPA complainant credibility, focusing primarily on the role of report consistency.

[2] The Canadian Model Jury Instructions do, however, specify that the importance of inconsistencies may vary and that jurors should "consider the fact, nature and extent of any differences" when evaluating credibility.

# STUDY 1: CONSISTENCY OF REPORTS OF ABUSIVE AND EVERYDAY AUTOBIOGRAPHICAL MEMORY EVENT PREVALENCE

In a recent review of the traumatic memory literature, Brewin (2007) noted, "the failure of all but a few studies to compare stability of recall for a traumatic and a non-traumatic event has limited the conclusions that can be drawn. Recall of trauma clearly shows some inconsistency, but we cannot say with any confidence whether this is more or less than for everyday events" (p. 237; see also Read, 2001). Further, no research of which I am aware has compared consistency in reports of event prevalence (i.e., occurrence and frequency of repeated events as compared to narratives of a single event) for traumatic and everyday events generally, nor abuse experiences specifically. Thus, the first study of this dissertation compared consistency of reports of abusive experiences and everyday autobiographical memory events to understand what level of consistency can reasonably be expected over time and across repeated questioning. Such a comparison may elucidate whether inconsistencies in reported event prevalence reflect normative rather than abuse-specific memory impairments.

There is no doubt that experiences of IPA victimization generally represent highly negative experiences. The American Psychiatric Association (2000) defines trauma as comprising "direct personal experience of an event that involves actual or threatened death or serious injury, or other threat to one's physical integrity" and requiring a response involving "intense fear, helplessness, or horror" (p. 463). Indeed, IPA is commonly referenced in clinical guides as a traumatic stressor contributing to PTSD

(e.g., Briere & Scott, 2006; Herman, 1997), and the association between IPA victimization and posttraumatic symptomatology is well-documented for physical, sexual, and even psychological abuse (e.g., Arias & Pape, 1999; Coker et al., 2002; Follingstad, Rutledge, Berg, Hause, & Polek, 1990). Although comparison of reports of IPA victimization (of which some proportion may be truly traumatic) and reports of everyday autobiographical memory events may be distinct from comparisons of reports of traumatic to non-traumatic memory events, the present study may inform the traumatic memory debate by increasing our understanding of how consistency differs as a function of autobiographical memory type.

## *Introduction*

### *Background*

Often framed in terms of (test-retest) reliability or report accuracy, the issue of report consistency has been a topic of considerable interest in survey research (e.g., Catania, Gibson, Chitwood, & Coates, 1990; DiFranceisco, McAuliffe, & Sikkema, 1998; Loftus, Smith, Klinger, & Fiedler, 1992; Wight & West, 1999). Research demonstrates that the reliability and validity of estimates of event prevalence, and behavioural frequency in particular, may be compromised for various reasons, including the fallibility of personal recall (e.g., Loftus et al., 1992; Pearson, Ross, & Dawes, 1992). Retrospective or memory-based reports rely on respondents' recollections of past behaviours or attitudes, for example, and doing so often involves a reconstructive process (e.g., Pearson et al., 1992). In this reconstruction, the present serves as an anchor or starting point upon which estimates of the past are made. However, what constitutes the 'present' will change across questioning times, potentially affecting respondents' recall

and thus, reducing consistency over repeated questioning. Additionally, retrieval cues that normally are effective in eliciting recall of IPA events while still in a dyadic relationship may no longer be available once the relationship has ended and, for this reason alone, we may anticipate apparent losses in recollection.

*Semi-structured calendar-based* techniques are the emerging state-of-the-art in surveying and interviewing (e.g., Belli, Shay, & Stafford, 2001; Belli, 1998; Room, 1990; Schwarz & Oyserman, 2001; Sobell & Sobell, 1992). These methods require respondents to estimate daily frequency of target behaviour(s), using a calendar covering the reference period as a memory aid. Personally significant and easy-to-remember landmark dates are marked on the calendar to serve as temporal anchors (Belli et al., 2001). One such procedure is the *Timeline FollowBack* (TLFB), an interview technique originally developed to assist research participants and treatment clients in recalling past drinking behaviour (Sobell & Sobell, 1996). The approach has been adopted across various research areas examining behavioural frequency estimates, including drug abuse (e.g., Fals-Stewart, O'Farrell, Freitas, McFarlin, & Rutigliano, 2000), cigarette smoking (e.g., Brown et al., 1998), risky sexual behaviours (e.g., Carey, Carey, Masito, Gordon, & Weinhardt, 2001), and IPA (e.g., Fals-Stewart, Birchler, & Kelley, 2003; Vendetta, Stappenbeck, & Fals-Stewart, 2004).

Reports obtained through calendar-based techniques have greater validity than do basic questions about usual quantity and frequency of the target behaviour(s) (e.g., O'Hare, 1991; Sobell & Sobell, 1992). Specifically, such techniques capitalize on the hierarchical structure of autobiographical memory, using top-down, sequential, and parallel strategies to guide the retrieval process (*cf.* Belli, 1998). In top-down retrieval,

personally significant or thematic knowledge (i.e., regarding persons and types of events) and temporal (i.e., regarding place in time and sequences of events) information cue the recall of more specific event information (Anderson & Conway, 1993). In contrast, sequential and parallel retrieval involve the chronological recall of autobiographical events (Conway, 1996). Whereas traditional methods typically obtain retrospective reports using either top-down, sequential, or parallel retrieval, calendar-based methods capitalize on these varied and complementary pathways by first cueing respondents to recall easy-to-remember, personally significant (and thus, higher-order) events. Respondents are subsequently asked to recall the chronological sequencing of the target events, using the higher-order events to cue time and place, as well as associations between them, contributing to improvements in the quality of retrospective reports over reports elicited through traditional methods (*cf.* Belli, 1998; Sobell & Sobell, 1992).

*Study 1*

Using TLFB-derived methodology, this study compared the consistency of reports of autobiographical memories of abusive experiences with autobiographical memories of everyday events, both within and between participants. Briefly, all participants completed a calendar-method survey once during the baseline session and again following a six-week delay. On the one hand, participants in the IPA sample, who self-identified as experiencing IPA in a romantic relationship in the past year (defined as psychological, physical, or sexual abuse), completed a survey assessing abusive experiences as well as a survey assessing participation in leisure activities (LA) in the past year. Participants in the comparison sample, on the other hand, only completed the survey assessing LA participation. By comparing reports of IPA victimization to reports of everyday

autobiographical event prevalence, the goal was to examine whether the consistency of IPA reports differed quantitatively (i.e., with regard to prevalence) or qualitatively (i.e., with regard to type of inconsistency, occurrence or frequency; direction of inconsistency, increased or decreased disclosure; and confidence) from the consistency of reports of everyday events, *within* a sample self-identified as victims of IPA and *between* victim and comparison samples.

*Hypotheses*

There is a long-standing debate in the memory field regarding whether memories of highly emotional or 'traumatic' events are characteristically different than other types of memories (e.g., APA, 1998; Brewin, 2007; Read, 2001; Read & Lindsay, 1997). The *traumatic-memory theory* posits that traumatic memories are encoded differently from non-traumatic ones and that memory impairments increase with the level of the 'trauma' (e.g., Freyd, 1996; Herman, 1997; Herman & Schatzow, 1987; Terr, 1994; van der Kolk & Fisler, 1995). In contrast, the *trauma-superiority theory* suggests that trauma may actually improve memory accuracy and consistency over time (e.g., Peace & Porter, 2004; Porter & Birt, 2001; Porter & Peace, 2007; Shobe & Kihlstrom, 1997). A third perspective holds that, considered together, there is no definitive evidence that memories of traumatic experiences are unique, or more or less accurate than other memories, and that many of the same cognitive processes and mechanisms may apply (e.g., Greaves, 2005; Geraerts et al., 2007, Kihlstrom, 1996; Lana & Loftus, 2005; McNally, 2003; Pezdek & Taylor, 2002; Read, 2001; Talarico & Rubin, 2003, 2007).

This debate leads to competing hypotheses. First, following the latter position that there is no special traumatic memory mechanism, one might anticipate that report

consistency over time would not differ significantly between the IPA and comparison conditions and samples. Alternatively, we might expect greater consistency over time in reports of IPA victimization, arguably traumatic in nature, not because of specialized memory mechanisms, but as a function of normal memory processes such as rehearsal. For example, research suggests that the ruminations and intrusions associated with psychological sequelae of negative, highly emotional experiences (such as PTSD or depression) may increase rehearsal of the related memories, thereby increasing their strength and availability in recall (e.g., Greaves, 2005; Hertel, 2004; Krinsley, Gallagher, Weathers, Kutter, & Kaloupek, 2003; Read & Lindsay, 2000; Southwick, Morgan, Nicolaou, & Charney, 1997). Others may anticipate differences in consistency over time as a consequence of fundamental differences in the original encoding of abuse memories compared to memories of everyday events (for reviews, see Brewin, 2001; Ehlers & Clark, 2000).

Evidently, the impact of trauma on consistency is a contentious issue. Depending largely on the research methodology and type of data gathered, there is some support for each of the varying positions. Further, there is no research speaking to the issue for event prevalence. As a result, I am not predicting one outcome over another, but rather am proposing a unique and objective empirical manner (i.e., comparing within and between samples) through which the issue may be approached.

## *Method*

### *Participants*

*IPA Sample.* Participants in the IPA sample were 27 men and 111 women who volunteered to participate in an online study examining relationship conflicts, self-

identified as experiencing IPA, and completed the online surveys at two points in time.[3] IPA was defined as behaviours or "means, in the sense of overt actions, used to resolve conflicts of interest [differences in personal desires] by intimate partners" (Straus, Hamby, & Warren, 2003, p. 6), including verbal (e.g., yelling, insulting, calling names), physical (e.g., pushing, shoving, hitting), or sexual (e.g., forcing sexual contact) abuse.

Community men and women who had experienced IPA in a relationship (greater than three months in length) in the 12 months prior to signing up for the study were recruited via posters distributed to domestic violence shelters and organizations throughout North America. The posters described the nature of the project, what would be involved in participation, and the potential benefits of the research. Posters also were placed in public venues and university campuses throughout the Lower Mainland. Advertisements were placed in local newspapers. Undergraduate men and women were recruited through the Simon Fraser University Introductory Psychology Research Participation system. The final IPA sample included 138 participants who were recruited through newspaper advertisements (4%, $n = 6$), posters (20%, $n = 28$), the undergraduate subject pool (65%, $n = 90$), and other methods (10%, $n = 14$) (e.g., snowball sampling).

Most IPA sample participants were Caucasian (38%, $n = 52$) or Asian (49%, $n = 67$), with the remainder of Aboriginal (1%, $n = 2$), African (1%, $n = 2$), Hispanic (1%, $n = 2$), and other (9%, $n = 14$) ethnicities. The majority (60%, $n = 83$) indicated they spoke English as their primary language at home. Participants ranged in age from 17 to 65 years, with a mean age of 23.43 ($SD = 7.49$). The IPA sample was well-educated: The vast majority of participants had completed their high school education (96%, $n = 132$)

[3] In total, 210 respondents completed the baseline session and 170 completed the follow-up (81%). Of those who completed the follow-up, 32 respondents did not successfully submit all study components at each time point, for a final IPA sample of 138 participants.

and 61% ($n$ = 84) reported at least some post-secondary education. Nearly three-quarters of IPA sample participants (74%, $n$ = 102) reported an ongoing relationship with their abusive partner at the baseline session (Time 1: T1), which decreased to about two-thirds (65%, $n$ = 89) at the follow-up session (Time 2: T2). The abusive intimate relationships ranged in length from one to 24 years, with a mean length of 2.97 ($SD$ = 4.31).[4] The average length of delay between T1 and T2 for IPA sample participants was 45.11 days ($SD$ = 9.10; Range = 28 - 84).

*Comparison Sample.* A comparison sample was created by oversampling 164 men ($n$ = 35) and women ($n$ = 129) who volunteered to participate in an online study examining LA participation in the 12 months prior to signing up for the study and completed the online surveys at both time points.[5] Comparison sample participants were recruited through the same means as IPA sample participants (with the exception of posters distributed to domestic violence shelters and organizations). The samples then were matched at the group level on gender, participant setting (i.e., undergraduate vs. community), age at T1, and length of delay between T1 and T2. The final comparison sample comprised 27 men and 111 women who were recruited through posters (20%, $n$ = 27), the undergraduate subject pool (65%, $n$ = 90), and other methods (15%, $n$ = 21).

As with the IPA sample, most comparison sample participants were Caucasian (39%, $n$ = 54) or Asian (49%, $n$ = 67), with the remainder of Hispanic (4%, $n$ = 5) and other (9%, $n$ = 12) ethnicities. Again, the majority (65%, $n$ = 90) indicated that they spoke English as their primary language at home. Comparison sample participants ranged

[4] Relationship length was calculated as a function of T1 date for participants who reported ongoing relationships.

[5] In total, 238 respondents completed the baseline and 183 completed the follow-up (77%). Of those who completed the follow-up, 19 respondents did not successfully submit all study components at each time point, for a comparison sample of 164 participants.

in age from 17 to 61 years, with a mean age of 23.09 (*SD* = 7.10). Participants in the comparison sample also were well-educated, with the vast majority having completed their high school education (99%, *n* = 137) and 72% (*n* = 99) reporting at least some post-secondary education. The average length of delay between T1 and T2 for the comparison sample was 44.04 days (*SD* = 7.18; Range = 32 - 77).

*Design*

Study 1 was a 2 (Participant Sample: IPA, Comparison) x 2 (Survey Content: IPA, LA) x 2 (Questioning Time: T1, T2) incomplete mixed factorial design. Participant Sample was a between-subjects variable. Survey Content was the within-subjects variable for participants in the IPA sample, whereas participants in the comparison sample only completed surveys assessing LA participation. In practice, Participant Sample was determined according to whether participants volunteered for a study assessing relationship conflicts (IPA sample) or whether they volunteered for a study assessing LA participation (comparison sample). Participation in one sample precluded participation in the other sample; that is, volunteers who signed up for the study on relationship conflicts were unable to later sign up for the LA study and vice versa.

*Materials*

*Relationship Behaviours Questionnaire* (RBQ; see Appendix 1.1). The RBQ was developed by modifying the *Timeline FollowBack Spousal Violence Interview* (TLFB-SV; Vendetta et al., 2004), a calendar-based method designed to assess daily patterns and frequency of spousal violence. Interviewers present respondents with a daily calendar dating back from the interview date for the number of days in the target interval. Standard

U.S. holidays noted on the calendar[6] and respondents mark other personally significant days (e.g., anniversaries, birthdays, school holidays, paydays, relationship break-ups, and major community events). To assess spousal violence, respondents then are asked to identify, on the calendar, occurrences of physical aggression during the specified time period (12 months in the case of the present study), indicating the specific type based on the eight behaviours included in the Physical Aggression subscale on the original *Conflict Tactics Scale* (Straus, 1990: i.e., 1 – pushed, grabbed, or shoved; 2 – slapped; 3 – threw something; 4 – kicked, bit, or hit with a fist; 5 – hit or tried to hit with something; 6 – beat up; 7 – threatened with a knife or gun; 8 – used a knife or gun; 9 – other). A relatively new measure, the extant research demonstrates the reliability and validity of the TLFB-SV. For example, among a sample of couples ($n$ = 104) in which the male partners were entering a spousal violence outpatient treatment program, Fals-Stewart and colleagues (2003) found the interview to have high test-retest reliability (*ICC*'s > .91, $p$'s < .01) and significant moderate correlations ($r$'s > .32, $p$'s < .05) across subscales with measures of partner violence and dyadic adjustment (see also Fals-Stewart & Golden, 2002; Fals-Stewart, 2003; Parks & Fals-Stewart, 2004; Vendetta et al., 2004).

To create the RBQ for the present study, the TLFB-SV was expanded to include three categories of relationship behaviours included in the CTS2 (Psychological Aggression, Physical Assault, and Sexual Coercion), whereas in its original form, the TLFB-SV included only Physical Assault. Additionally, questions about criminal justice system involvement, hospitalization, and substance use were excluded. Minor modifications also were made to the wording of instructions as relevant to the online

[6] Canadian holidays also were noted in the present study.

administration (e.g., providing guidance regarding how to complete the online calendar, submitting responses, etc.).

*Leisure Activities Questionnaire* (LAQ; see Appendix 1.2). The LAQ was constructed for the present study based on the *Minnesota Leisure Time Physical Activity Questionnaire* (MLTPA; Taylor, Jacobs, Shucker, Knudsen, Leon, & DeBacker, 1978). The MLTPA assesses the type, frequency, and duration of participation in a range of physical activities during the past year. Respondents are asked their level of participation in 62 activities of varying intensity in eight categories: (1) Walking, (2) Conditioning Exercise, (3) Water Activities, (4) Winter Activities, (5) Sports, (6) Lawn and Garden Activities, (7) Home Repair Activities, and (8) Fishing and Hunting. The MLTPA has been used extensively in research assessing cardiovascular health and has demonstrated good test-retest reliability ($r$'s > .69, $p$'s < .05) (Folsom, Jacobs Jr., Caspersen, Gomez-Martin, & Knudsen, 1986) and moderate to large correlations with other measures of physical activity ($r$'s > .27, $p$'s < .05) (Richardson, Leon, Jacobs Jr., Ainsworth, & Serfass, 1994).

I selected the MLTPA as the model for the comparison measure in this study for a couple of reasons. First, it was designed following a calendar-based approach. Second, like the CTS2, the MLTPA assesses specific behaviours that fall under more general categories. For the present study, the LAQ included four categories of activities: (1) Walking (and related activities), (2) Conditioning Exercise, (3) Sports, and (4) Outdoor Activities (the MLTPA Water Activities and Winter Activities combined). These four were selected based on pilot testing of behaviours that could reasonably be expected within our target sample (e.g., Fishing and Hunting behaviours were not once endorsed

during pilot testing and thus, the category was excluded) and inclusion of approximately the same number of behaviours as included in the RBQ. The administration procedure and format of the MLTPA responses were modified such that the LAQ survey was conducted following the same procedure as for the RBQ. Duration of participation in each activity, although included in the MLTPA, was not assessed in the present study.

*Procedure*

At the baseline session, participants completed the RBQ and/or LAQ (as appropriate for their study conditions). In light of research demonstrating the effects of witness confidence on perceived credibility (e.g., Brewer & Burke, 2002; Penrod & Cutler, 1995) and the association between confidence and traumatic memory (e.g., Talarico & Rubin, 2003, 2007), participants were additionally asked to indicate their confidence (1 = Not at all confident; 6 = Very confident) regarding: (1) the occurrence of the category of behaviours (e.g., physical assault), (2) the occurrence and frequency of specific behaviours if the broader category was endorsed (e.g., being hit), and (3) their report overall. Participants also completed brief demographic questions (such as age, ethnicity, educational attainment, and language spoken at home).

During the follow-up session approximately six weeks later, participants again completed the online surveys as detailed above, with the specification to recall behaviours which occurred in the 12 months *before the baseline session.* Any behaviours occurring during the six-week delay between sessions were not reported and, thus, did not contribute to consistency scores (calculations described in more detail below). Participants were reminded via email of their participation timeline and specific dates. Community IPA sample participants were entered into two $100 draws for participating

in T1 and two $150 draws for participating in T2. Community LA sample participants were entered into two $75 draws for participating in T1 and two $100 draws for participating in T2. Undergraduate participants received extra course credit for their participation. The order in which the RBQ and the LAQ were completed was fully counterbalanced between and within IPA sample participants.

*Consistency Calculations*

Inconsistency was operationalized to include instances in which participants endorsed the occurrence of a behaviour at one time point, but not the other (e.g., in response to the closed-ended question, participant reported at T1 that s/he experienced physical abuse, but reported at T2 that s/he did not experience physical abuse) and also instances in which the frequency of the behaviours reported differed across time points (e.g., participant reported experiencing 14 unique instances of physical abuse at T1, but only two instances of physical abuse at T2). Inconsistency scores were computed as a function of the proportion of responses (see Koriat & Goldsmith, 1994) with respect to (1) the occurrence of the category (e.g., did the participant report physical abuse during T1 and T2); (2) the occurrence of specific behaviours (e.g., did the participant report being hit during T1 and T2); and (3) the frequency of behaviours.

With respect to the first two coding methods, inconsistencies were coded dichotomously: *yes* (i.e., inconsistent) or *no* (i.e., consistent). Proportion Inconsistent scores were computed by summing the number of yeses (1) for all categories and dividing by the number of categories, (2) for behaviours within each specific category and dividing by the number of behaviours within the category, and (3) for behaviours overall (i.e., across categories) and dividing by the number of behaviours. For the RBQ,

this resulted in five Proportion Inconsistent scores: (1) All IPA Categories (i.e., number of IPA categories endorsed); (2) Psychological Aggression (3) Physical Assault; (4) Sexual Coercion; and (5) All IPA Instances. For the LAQ, this resulted in six Proportion Inconsistent scores: (1) All LA Categories (i.e., number of LA categories endorsed); (2) Walking and Related Activities; (3) Conditioning Exercise; (4) Sports; (5) Outdoor Activities; and (6) All LA Instances. The possible range for these scores was from 0 to 1.00, where a score of zero represents no inconsistency. For example, inconsistent endorsement of three of the 12 behaviours within the category of physical assault would result in a Proportion Inconsistent score of .25 (i.e., 3/12 = .25).

With respect to the frequency of behaviours, Degree Inconsistent scores were computed as a function of the frequency of behaviours reported across categories: Degree Inconsistent = Absolute Value [T1 frequency – T2 frequency / (T1 frequency + T2 frequency)]. For the RBQ, this resulted in four Degree Inconsistent scores: (1) Psychological Aggression; (2) Physical Assault; (3) Sexual Coercion; and (4) All IPA Instances. For the LAQ, this resulted in five Degree Inconsistent scores: (1) Walking and Related Activities; (2) Conditioning Exercise; (3) Sports; (4) Outdoor Activities; and (5) All LA Instances. The possible range of scores was from 0 to 1.00, where a score of zero represents no inconsistency. For example, if a participant reported 24 unique instances of IPA at T1 but only 11 instances of IPA at T2, this would result in a Degree Inconsistent score of .37 (i.e., 24 – 11 / [24 + 11] = .37) for IPA overall.

## *Results*

### *Autobiographical Events Reported*

Descriptive statistics for the occurrence and frequency of IPA victimization and LA participation are presented in Table 1.1 as a function of Participant Sample.

*IPA Victimization.* Overall, Psychological Aggression was the most commonly endorsed category of IPA: The vast majority of IPA sample participants reported experiencing psychological aggression at both T1 and T2 (see Table 1.1). In comparison, experiences of physical assault and sexual coercion were considerably less common, reported by approximately one-quarter of the sample. This pattern of responding was reflected in the number of instances of IPA reported: Instances of psychological aggression were most frequent overall, followed by sexual coercion then physical assault. On average, the number of instances of psychological aggression participants reported was approximately twice as great as the number of instances of sexual coercion and approximately three times greater than for physical assault (see Table 1.1).

Examination of reports of IPA victimization at the aggregate level demonstrated a trend for occurrence to decrease over time, although the percentage of participants reporting IPA victimization did not differ significantly from T1 to T2 category ($p$'s > .05). Repeated measures analyses of variance (ANOVAs) revealed that the mean number of instances reported decreased significantly from T1 to T2 for IPA overall (across categories), $F(1, 104) = 5.66, p < .05, \eta_p^2 = .05$, as well as for psychological aggression, $F(1, 102) = 4.74, p < .05, \eta_p^2 = .04$ (see Table 1.1). Differences in the frequency of physical assault and sexual coercion reported did not reach significance ($p$'s > .05).

*LA Participation.* Across samples, Walking and Related Activities and Conditioning Exercise were the most commonly endorsed LA categories with more than three-quarters of both the IPA and comparison samples reporting participation at T1 and T2 (see Table 1.1). In comparison, reports of participation in sports and outdoor activities were somewhat less common, but still were reported by more than half of all participants. As for the IPA victimization results, this pattern of responding was reflected in the frequency of instances reported: Instances of conditioning exercise were most frequent overall, followed by walking and related activities, then sports, and lastly, outdoor activities. On average, the frequency of walking and related activities and of conditioning exercise reported were approximately twice that of sports and outdoor activities.

Comparison sample participants were significantly more likely than IPA sample participants to report engaging in walking and related activities, conditioning exercise, and outdoor activities at T1, $\chi^2$'s $(1, N = 276) \geq 3.95$, $p$'s $\leq .05$, $\Phi \geq .12$, and conditioning exercise at T2, $\chi^2(1, N = 276) = 12.89$, $p < .001$, $\Phi = .22$. No differences were observed, however, in the percentage of participants reporting any LA participation at either T1 or T2 ($p$'s $> .05$). In terms of frequency, reports differed significantly between samples only for overall LA participation at T2: Comparison sample participants reported approximately 25 more instances than IPA sample participants, $t(252) = 2.20$, $p < .05$, $d = .28$.

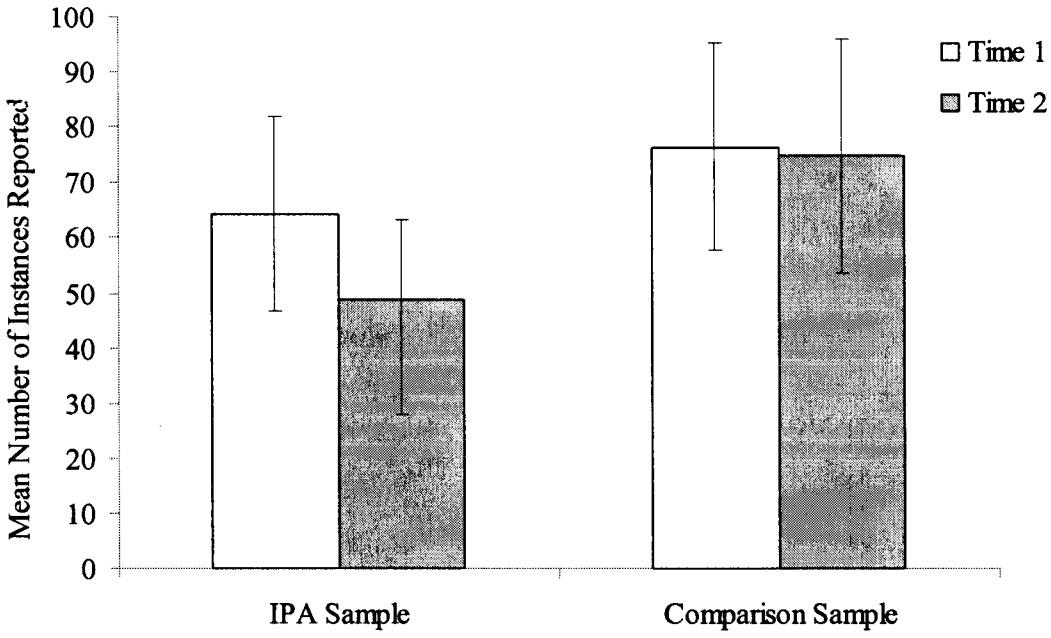
***Table 1.1. Occurrence and Frequency of Intimate Partner Abuse and Leisure Activities Reported***

| Behaviours | Participants who Reported Behaviour Occurred | | Number of Instances Reported | |
|---|---|---|---|---|
| | T1<br>% (*n*) | T2<br>% (*n*) | T1<br>*M* (*SD*) | T2<br>*M* (*SD*) |
| *IPA Sample* | | | | |
| IPA Overall | 96 (133) | 89 (123) | 16.94 (32.05) | 10.79 (17.17) |
| Psychological Aggression | 91 (125) | 86 (119) | 14.10 (27.98) | 8.83 (14.93) |
| Physical Assault | 20 (28) | 25 (34) | 2.96 (3.64) | 2.04 (2.34) |
| Sexual Coercion | 32 (44) | 25 (34) | 7.00 (8.88) | 4.94 (8.97) |
| LA Overall | 98 (135) | 96 (132) | 60.90 (89.51) | 42.55 (67.82) |
| Walking and Related Activities | 86 (118) | 83 (114) | 28.21 (57.08) | 21.14 (43.69) |
| Conditioning Exercise | 76 (105) | 68 (94) | 31.14 (56.15) | 19.07 (44.54) |
| Sports | 51 (71) | 46 (64) | 16.57 (28.95) | 13.78 (26.29) |
| Outdoor Activities | 57 (78) | 56 (77) | 8.37 (17.63) | 6.72 (15.67) |
| *Comparison Sample* | | | | |
| LA Overall | 99 (137) | 97 (134) | 69.20 (96.25) | 67.51 (107.95) |
| Walking and Related Activities | 94 (129) | 86 (119) | 26.71 (38.97) | 24.19 (39.84) |
| Conditioning Exercise | 86 (118) | 86 (119) | 35.78 (71.77) | 38.91 (94.27) |
| Sports | 62 (86) | 58 (80) | 13.87 (19.99) | 12.04 (19.51) |
| Outdoor Activities | 68 (94) | 64 (88) | 8.78 (14.62) | 7.07 (10.30) |

*Notes.* IPA sample $n = 138$; comparison sample $n = 138$.

Examination of LA reports at the aggregate level demonstrated a non-significant trend for occurrence to decrease over time ($p$'s > .05). A repeated measures ANOVA revealed that the mean overall number of LA instances reported across samples decreased significantly from T1 to T2 ($M_{T1}$ = 65.08, $SD$ = 92.88; $M_{T2}$ = 55.13, $SD$ = 90.99), $F(1, 252) = 7.94$, $p < .01$, $\eta_p^2 = .03$, and also revealed a significant Time x Participant Sample interaction, $F(1, 252) = 5.48$, $p < .01$, $\eta_p^2 = .02$. *Post hoc* comparisons revealed that the mean number of instances reported decreased significantly over time for IPA but not comparison sample participants (see Figure 1.1), $t(125) = 3.47$, $p < .001$, $d = .62$, a pattern that proved to be relatively stable across individual categories.

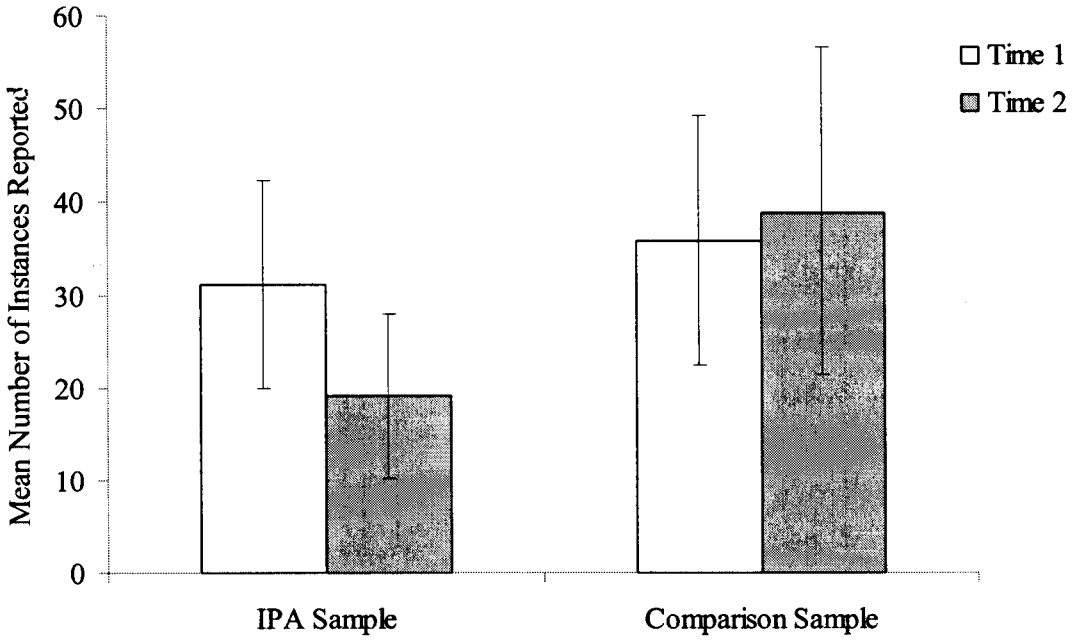
***Figure 1.1. Mean Number of Instances of Leisure Activities Reported Overall at Time 1 and Time 2 as a Function Participant Sample***




*Notes.* Bars represent 95% confidence intervals around the means.

Subsequent repeated measures ANOVAs at the category level revealed a modest but significant decrease in the mean number of instances of walking reported over time ($M_{T1}$ = 27.31, $SD$ = 48.46; $M_{T2}$ = 22.83, $SD$ = 41.74), $F(1, 212) = 5.05$, $p < .05$, $\eta_p^2 = .02$, as well as a Time x Participant Sample interaction for conditioning exercises, $F(1, 211) = 5.87$, $p < .05$, $\eta_p^2 = .03$. As before, *post hoc* comparisons revealed that the mean number of instances reported decreased significantly over time for IPA but not comparison sample participants, $t(99) = 3.09$, $p < .01$, $d = .62$ (see Figure 1.2). Differences in the mean number of sports and outdoor activities reported did not reach significance, nor were there significant differences over time as a function of Participant Sample ($p$'s > .05).

***Figure 1.2. Mean Number of Instances of Conditioning Exercise Reported at Time 1 and Time 2 as a Function Participant Sample***




*Notes.* Bars represent 95% confidence intervals around the means.

*Number of Categories.* The number of categories of IPA victimization and LA participation reported also was explored. On average, IPA sample participants endorsed 1.43 ($SD$ = 0.71) of the three IPA categories at T1 and 1.36 ($SD$ = 0.79) at T2, representing a non-significant decrease over time ($p > .05$). With regard to LA categories, IPA sample participants reported participation in 2.70 ($SD$ = 1.10) of the four at T1 and 2.53 ($SD$ = 1.12) at T2, whereas LA sample participants reported participation in 3.09 ($SD$ = 0.94) of the four at T1 and 2.94 ($SD$ = 1.03) at T2. Echoing earlier comparisons across Participant Sample, pairwise comparisons revealed that IPA sample participants reported participating in significantly fewer LA categories on average than did LA sample participants at both T1 and T2, $t$'s (274) $\geq$ 3.20, $p$'s < .01, $d$'s $\geq$ .39. A repeated measures ANOVA failed to reveal a significant Time x Participant Sample interaction ($p > .05$). This analysis, however, did reveal a modest but significant decrease in the mean number of LA categories endorsed over time, $F(1, 274) = 7.98$, $p < .01$, $\eta_p^2 = .03$.

*Participant Setting.* To determine whether undergraduate and community participants could reasonably be combined into one sample, analyses were conducted to examine the effects, if any, of Participant Setting on the occurrence and frequency of IPA victimization and LA participation reported. Descriptive statistics for the occurrence and frequency of IPA victimization and LA participation as a function of Participant Setting are presented in Tables 1.2 and 1.3.

***Table 1.2. Occurrence and Frequency of Intimate Partner Abuse Reported by Undergraduate and Community Participants***

| Behaviours | Participants who Reported Behaviour Occurred | | Number of Instances Reported | |
|---|---|---|---|---|
| | T1<br>% (*n*) | T2<br>% (*n*) | T1<br>*M* (*SD*) | T2<br>*M* (*SD*) |
| *Undergraduate Participants* | | | | |
| IPA Overall | 98 (88) | 89 (80) | 13.53 (18.84) | 10.84 (14.62) |
| Psychological Aggression | 93 (84) | 86 (77) | 11.38 (19.60) | 8.88 (13.16) |
| Physical Assault | 19 (17) | 21 (19) | 3.13 (4.41) | 1.80 (2.31) |
| Sexual Coercion | 27 (24) | 21 (19) | 5.55 (8.22) | 5.77 (9.34) |
| *Community Participants* | | | | |
| IPA Overall | 94 (45) | 90 (43) | 24.06 (49.03) | 10.69 (22.26) |
| Psychological Aggression | 85 (41) | 88 (42) | 19.88 (44.51) | 8.71 (18.71) |
| Physical Assault | 23 (11) | 31 (15) | 2.77 (2.65) | 2.36 (2.46) |
| Sexual Coercion | 42 (20) | 31 (15) | 9.29 (9.71) | 3.54 (8.47) |

*Notes.* IPA sample $n = 138$ (90 undergraduate participants; 48 community participants).

***Table 1.3. Occurrence and Frequency of Leisure Activities Reported by Undergraduate and Community Participants***

| Behaviours | Participants who Reported Behaviour Occurred | | Number of Instances Reported | |
|---|---|---|---|---|
| | T1<br>% (*n*) | T2<br>% (*n*) | T1<br>*M* (*SD*) | T2<br>*M* (*SD*) |
| *Undergraduate Participants* | | | | |
| LA Overall | 99 (178) | 99 (178) | 66.01 (86.93) | 59.99 (94.97) |
| Walking and Related Activities | 88 (159) | 83 (150) | 26.78 (48.39) | 24.20 (43.40) |
| Conditioning Exercise | 84 (151) | 79 (143) | 32.79 (59.73) | 31.53 (83.27) |
| Sports | 64 (116) | 59 (107) | 15.82 (25.76) | 13.27 (22.74) |
| Outdoor Activities | 62 (112) | 59 (106) | 9.27 (18.29) | 7.18 (14.39) |
| *Community Participants* | | | | |
| LA Overall | 98 (94) | 92 (88) | 63.30 (103.86) | 45.79 (82.54) |
| Walking and Related Activities | 92 (88) | 87 (83) | 28.66 (48.92) | 19.85 (38.19) |
| Conditioning Exercise | 75 (72) | 73 (70) | 35.26 (74.53) | 25.64 (57.26) |
| Sports | 43 (41) | 39 (37) | 12.92 (19.76) | 11.43 (22.58) |
| Outdoor Activities | 63 (60) | 62 (59) | 7.33 (10.39) | 6.42 (9.59) |

*Notes.* $n = 276$ (180 undergraduate participants; 96 community participants).

Overall, no significant differences were found in reports of IPA victimization. Specifically, chi-square analyses demonstrated no significant differences between community and undergraduate participants' reports regarding whether they experienced IPA overall or by category ($p$'s > .05), although slightly higher percentages were generally observed among community participants (see Table 1.2). The number of IPA instances reported by community participants was greater on average than the frequency reported by undergraduate participants overall and with respect to psychological aggression and sexual coercion at T1, as may be seen in Table 1.2. In contrast, undergraduate participants actually reported more instances of physical assault and sexual coercion on average than community participants at T2. Differences again, however, failed to reach statistical significance ($p$'s > .05). Pairwise comparisons also demonstrated that undergraduate participants did not differ significantly from community participants in the IPA sample ($p$'s > .05) with respect to the number of categories of IPA endorsed at T1 ($M_{undergrad}$ = 1.39, $SD$ = 0.63; $M_{community}$ = 1.50, $SD$ = 0.85) and T2 ($M_{undergrad}$ = 1.28, $SD$ = 0.73; $M_{community}$ = 1.50, $SD$ = 0.88).

A few significant differences in the occurrence and frequency of LA participation as a function of Participant Setting were found. Chi-square analyses revealed that undergraduate participants were significantly more likely to report participating in sports at T1 and T2 than were community participants, $\chi^2$'s (1, $N$ = 276) $\geq$ 10.96, $p$'s < .001, $\Phi$ = .20, and also significantly more likely to report LA participation overall at T2, $\chi^2$(1, $N$ = 276) = 9.35, $p$ < .01, $\Phi$ = .18 (see Table 1.3). Although the number of LA instances reported by undergraduate participants was greater on average than the frequency reported by community participants (see Table 1.3), pairwise comparisons revealed that

differences did not reach statistical significance ($p$'s > .05). Nor did undergraduate participants differ significantly from community participants ($p$'s > .05) on the number of LA categories endorsed at T1 ($M_{undergrad}$ = 1.39, $SD$ = 0.63; $M_{community}$ = 1.50, $SD$ = 0.85) and T2 ($M_{undergrad}$ = 1.39, $SD$ = 0.63; $M_{community}$ = 1.50, $SD$ = 0.85).

In summary, very few significant differences were found as a function of Participant Setting, affording the opportunity to collapse across undergraduate and community participants within the IPA and comparison samples. In contrast, however, length of the abusive intimate relationships reported by undergraduate participants in the IPA sample, not surprisingly, was significantly shorter on average ($M$ = 1.92 years, $SD$ = 1.78) compared with community participants ($M$ = 4.94 years, $SD$ = 6.49), $t(136) = 4.15$, $p < .001$, $d = .71$. As such, relationship length will be tested as a covariate in subsequent analyses within the IPA sample.

*Question Order.* Within the IPA sample, Question Order (1: IPA questions followed by LA questions; 2: LA questions followed by IPA questions) was counterbalanced at T1 and T2, such that IPA participants were randomly assigned to one of four Question Order conditions (Order 1 at T1, Order 1 at T2; Order 1 at T1, Order 2 at T2; Order 2 at T2, Order 1 at T2; Order 2 at T2, Order 2 at T2). Descriptive statistics for the occurrence and frequency of IPA victimization and LA participation as a function of Question Order are presented in Table 1.4. Chi-square analyses identified only one significant effect of Question Order on the occurrence of IPA victimization and LA participation reported: Participants who received Question Order 2 at T1 were significantly more likely to report engaging in walking and related activities at T1

compared with participants who received Question Order 1, $\chi^2(1, N = 138) = 6.55, p < .01, \Phi = .22$ (see Table 1.4).

Further analyses showed no significant differences in the frequency of IPA instances reported by participants at T1 or T2 as a function of Question Order ($p$'s > .05). Nor did the number of IPA and LA categories endorsed at T1 and T2 differ significantly between participants who received Question Order 1 versus Question Order 2 ($p$'s > .05). The frequency of LA instances reported at T2, however, did vary significantly as a function of Question Order at T2. As may be seen in Table 1.4, on average, participants who received Question Order 1 reported significantly more instances of LA participation overall at T2 than participants who received Question Order 2, $t(124) = 2.20, p < .05, d = .40$. Although there was a trend in the opposite direction at T1, the difference between instances of LA participation overall reported at T1 was not significant ($p > .05$).

In general, Question Order was not found to be systematically associated with responding, affording the examination consistency of reports of autobiographical memory events over time as independent from this possible procedural effect. However, Question Order will be tested as a covariate in subsequent analyses of the occurrence of Walking and Related Activities and frequency of LA instances overall.

***Table 1.4. Occurrence and Frequency of Intimate Partner Abuse and Leisure Activities Reported as a Function of Question Order***

| Behaviours | Participants who Reported Behaviour Occurred | | Number of Instances Reported | |
|---|---|---|---|---|
| | T1<br>% (*n*) | T2<br>% (*n*) | T1<br>*M* (*SD*) | T2<br>*M* (*SD*) |
| *Order 1 (IPA - LA)* | | | | |
| IPA Overall | 94 (63) | 91 (62) | 18.30 (25.46) | 13.64 (20.88) |
| Psychological Aggression | 97 (58) | 88 (60) | 14.53 (20.11) | 11.06 (17.74) |
| Physical Assault | 22 (15) | 29 (20) | 3.13 (3.74) | 2.29 (2.37) |
| Sexual Coercion | 31 (21) | 31 (21) | 10.20 (10.09) | 6.05 (9.95) |
| LA Overall | 97 (65) | 94 (64) | 48.80 (58.27) | 55.65 (81.54) |
| Walking and Related Activities | 78 (52) | 79 (54) | 21.51 (32.63) | 28.96 (53.65) |
| Conditioning Exercise | 76 (51) | 75 (51) | 22.30 (36.33) | 25.45 (56.15) |
| Sports | 46 (31) | 43 (29) | 20.00 (31.61) | 15.52 (27.20) |
| Outdoor Activities | 58 (39) | 59 (40) | 8.80 (23.20) | 6.41 (10.84) |
| *Order 2 (LA - IPA)* | | | | |
| IPA Overall | 99 (70) | 87 (61) | 15.64 (37.52) | 7.94 (11.95) |
| Psychological Aggression | 94 (67) | 84 (59) | 13.68 (34.28) | 6.68 (11.37) |
| Physical Assault | 18 (13) | 20 (14) | 2.77 (3.65) | 1.75 (2.38) |
| Sexual Coercion | 32 (23) | 19 (13) | 4.71 (7.32) | 3.29 (7.30) |
| LA Overall | 99 (70) | 97 (86) | 72.25 (110.44) | 29.44 (47.72) |
| Walking and Related Activities | 93 (66) | 86 (60) | 34.07 (71.80) | 14.04 (30.97) |
| Conditioning Exercise | 76 (54) | 61 (43) | 39.98 (69.94) | 11.27 (21.97) |
| Sports | 56 (40) | 50 (35) | 14.19 (27.15) | 12.10 (25.73) |
| Outdoor Activities | 55 (39) | 53 (37) | 7.94 (9.60) | 7.09 (20.05) |

*Notes*. $n = 138$. Order 1 = IPA questions followed by LA questions; Order 2 = LA questions followed by IPA questions.

*Consistency of Reports over Time*

Descriptive statistics for the Proportion and Degree Inconsistent scores for IPA victimization and LA participation are presented in Table 1.5. Only two participants were completely consistent in their reports of both the occurrence and frequency of experiences (one IPA sample participant and one comparison sample participant). Generally, participants were fairly consistent in their endorsement of categories of experiences, as well as the occurrence of specific behaviours within those categories, but were quite inconsistent in frequency reports. As may be seen in Table 1.5, among IPA participants means for the Proportion Inconsistent scores ranged from .04 to .19 (*SD*'s = .07 - .22) for IPA victimization and .05 to .19 (*SD*'s = .07 - .22) for LA participation. Also presented in Table 1.5, means for the Degree Inconsistent scores were higher (indicating greater inconsistency), ranging from.48 to .55 (*SD*'s = .36 - .42) for IPA victimization and .34 to .54 (*SD*'s = .32 - .46) for LA participation. Among LA sample participants, Proportion Inconsistent scores ranged from .06 to .17 (*SD*'s = .08 - .19) and Degree Inconsistent scores range from .34 to .41 (*SD*'s = .32 - .41).

Data analysis of report consistency involved two components: (1) comparisons of the consistency of reports of IPA victimization with the consistency of reports of LA participation *within* IPA sample participants, and (2) comparisons of the consistency of LA reports made by IPA sample participants with the consistency of those made by comparison sample participants. First, findings with respect to the consistency of the occurrence of behaviours are described (Proportion Inconsistent scores), followed by a description of findings with respect to the consistency of the frequency of behaviours reported (Degree Inconsistent scores).

***Table 1.5. Proportion and Degree Inconsistent Scores for Intimate Partner Abuse and Leisure Activities as a Function of Participant Sample***

| Behaviours | Proportion Inconsistent | | Degree Inconsistent | |
|---|---|---|---|---|
| | *M* (*SD*) | Range | *M* (*SD*) | Range |
| *IPA Sample* | | | | |
| IPA Reports | | | | |
| All IPA Categories | .19 (.22) | 0 – 1.00 | -- | -- |
| Psychological Aggression | .17 (.15) | 0 – .75 | .55 (.41) | .33 – 1.00 |
| Physical Assault | .04 (.08) | 0 – .33 | .55 (.42) | 0 – 1.00 |
| Sexual Coercion | .06 (.12) | 0 – .57 | .52 (.41) | .67 – 1.00 |
| All IPA Instances | .09 (.07) | 0 – .37 | .48 (.36) | .39 – 1.00 |
| LA Reports | | | | |
| All LA Categories | .19 (.22) | 0 – .75 | -- | -- |
| Walking and Related Activities | .15 (.15) | 0 – .75 | .49 (.36) | .17 – 1.00 |
| Conditioning Exercise | .16 (.17) | 0 – .63 | .54 (.40) | 0 – 1.00 |
| Sports | .06 (.09) | 0 – .42 | .41 (.39) | 0 – 1.00 |
| Outdoor Activities | .05 (.07) | 0 – .25 | .51 (.38) | .33 – 1.00 |
| All LA Instances | .09 (.07) | 0 – .30 | .44 (.46) | .18 – .38 |
| *Comparison Sample* | | | | |
| LA Reports | | | | |
| All LA Categories | .14 (.19) | 0 – 1.00 | -- | -- |
| Walking and Related Activities | .15 (.14) | 0 – .88 | .39 (.32) | 0 – 1.00 |
| Conditioning Exercise | .17 (.17) | 0 – 1.00 | .34 (.35) | 0 – 1.00 |
| Sports | .06 (.11) | 0 – .92 | .40 (.39) | 0 – 1.00 |
| Outdoor Activities | .06 (.10) | 0 – .75 | .41 (.41) | 0 – 1.00 |
| All LA Instances | .10 (.08) | 0 – .73 | .34 (.33) | 0 – 1.00 |

*Notes.* IPA sample $n$ = 138; comparison sample $n$ = 138. Proportion Inconsistent scores refer to inconsistencies in reporting whether or not the behaviour occurred (e.g., did the participant report experiencing physical assault during T1 and T2). Degree Inconsistent scores refer to inconsistencies in reporting the number of instances of specific behaviours (e.g., participant reported 24 instances of IPA at T1 but only 11 instances of IPA at T2). Range of scores = 0 to 1.00, where 0 represents no inconsistency. -- = statistic could not be computed.

*Consistency in the Occurrence of Behaviours - Proportion Inconsistent Scores.*

Analyses of the occurrence of behaviours demonstrated differences in consistency over time within Survey Content (IPA vs. LA), but neither between Survey Content within IPA participants nor between IPA and comparison sample participants. Specifically, a repeated measures ANOVA demonstrated no differences in Proportion Inconsistent scores for IPA and LA reports among IPA sample participants for either the number of categories endorsed (compare All IPA Categories and All LA Categories in Table 1.5) or for the occurrence of instances overall (compare All IPA Instances and All LA Instances) ($p$'s > .05). However, a repeated measures ANOVA did reveal that Proportion Inconsistent scores differed significantly as a function of IPA category, $F(2, 274) = 54.96$, $p < .001$, $\eta_p^2 = .29$. *Post hoc* comparisons revealed that participants were significantly less consistent over time in their reports of whether they experienced psychological aggression (.17) compared with reports of both physical assault (.04) and sexual coercion (.06) as demonstrated by higher Proportion Inconsistent scores, $t$'s (137) $\geq 7.35$, $p$'s < .001, $d$'s $\geq 1.26$. The difference between reports of physical assault and sexual coercion approached significance ($p = .06$).[7] Although no longer statistically significant, the pattern remained after controlling for the mean frequency of instances reported within each category of IPA ($p = .08$), with only a slight reduction in effect size, $\eta_p^2 = .24$.

Comparisons of Proportion Inconsistent scores for LA reports as a function of Participant Sample echoed the within sample results. One-way ANOVAs with Proportion Inconsistent scores as the dependent variable showed no significant differences between

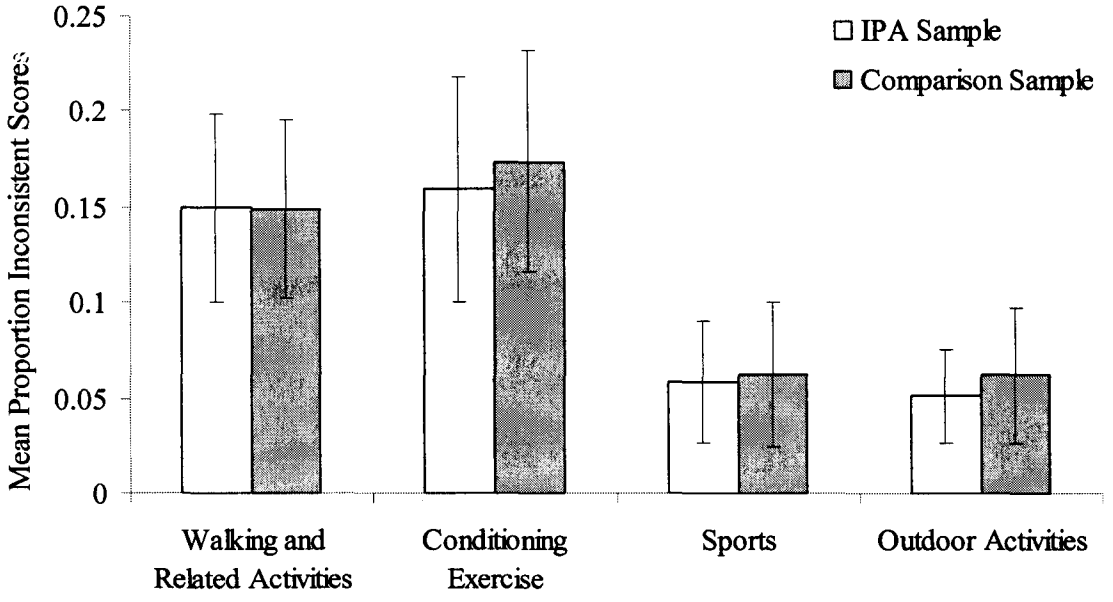
[7] Inclusion of relationship length as a covariate in analyses of Proportion Inconsistent scores for IPA reports did not produce significantly different results, nor was relationship length found to be a significant predictor of the Proportion Inconsistent scores.

IPA and comparison sample participants' report consistency with regard to either the number of categories endorsed (compare IPA and comparison sample Proportion Inconsistent scores for All LA Categories in Table 1.5) or the occurrence of instances overall (compare IPA and comparison sample Proportion Inconsistent scores for All LA Instances) ($p$'s > .05). Although there were no differences between IPA and comparison samples ($p > .05$), a repeated measures ANOVA revealed that Proportion Inconsistent scores differed significantly as a function of LA category, $F(3, 825) = 63.55, p < .001, \eta_p^2 = .19$.

As may be seen in Figure 1.3, *post hoc* comparisons revealed that participants were significantly less consistent over time in their reports of whether they participated in walking and related exercises than both sports and outdoor activities, as demonstrated by higher Proportion Inconsistent scores, $t$'s $(275) \geq 9.11$, $p$'s < .001, $d$'s $\geq 1.10$. Reports of participation in conditioning exercise also were less consistent than reports of participation in sports and outdoor activities, $t$'s $(275) \geq 9.27$, $p$'s < .001, $d$'s $\geq 1.12$. The Proportion Inconsistent scores for walking and related exercises versus conditioning exercise and sports versus outdoor activities were not significantly different ($p$'s > .05).[8] The differences remained statistically significant after controlling for the mean frequency of instances reported within each category of LA, $F(3, 234) = 18.18, p < .001, \eta_p^2 = .19$.

[8] Inclusion of Question Order as a covariate in analyses of Proportion Inconsistent scores for LA reports did not produce significantly different results, nor was Question Order found to be a significant predictor of Proportion Inconsistent scores.

***Figure 1.3. Mean Proportion Inconsistent Scores for Walking and Related Activities, Conditioning Exercise, Sports, and Outdoor Activities as a Function of Participant Sample***




*Notes.* Bars represent 95% confidence intervals around the means. The possible range of scores was from 0 to 1.00, where a score of zero represents no inconsistency.

*Consistency in the Frequency of Instances - Degree Inconsistent Scores.* Analyses of consistency in the number of instances reported revealed even fewer significant differences. In fact, no significant differences were observed: Repeated measures ANOVAs demonstrated no differences in Degree Inconsistent scores for IPA reports and LA reports within IPA sample participants for either the frequency of instances reported overall (compare All IPA Instances and All LA Instances in Table 1.5) or the frequency of instances within IPA categories (compare Psychological Aggression, Physical Assault, and Sexual Coercion Degree Inconsistent scores) ($p$'s > .05).[9]

With respect to consistency in the frequency of LA instances reported over time, a one-way ANOVA demonstrated that IPA and comparison sample participants were similar in their inconsistency for the number of LA instances reported overall (compare Degree Inconsistent scores for LA Instances Overall in Table 1.5) and a repeated measures ANOVA showed the same result within LA categories (compare Degree Inconsistent scores for Walking and Related Activities, Conditioning Exercise, Sports, and Outdoor Activities) ($p$'s > .05).[10]

In sum, comparisons of consistency in reports of IPA victimization with the consistency in reports of LA participation *within* IPA sample participants and consistency in reports of LA participation *between* IPA and comparison sample participants demonstrated few significant differences. No differences in report consistency were observed between IPA and comparison sample participants. Results instead highlighted

[9] Inclusion of relationship length as a covariate in analyses of Degree Inconsistent scores for IPA reports did not produce significantly different results, nor was relationship length found to be a significant predictor of the Degree Inconsistent scores.

[10] Inclusion of Question Order as a covariate in analyses of Degree Inconsistent scores for LA reports did not produce significantly different results, nor was Question Order found to be a significant predictor of any of the Degree Inconsistent scores.

differences in consistency over time *between* categories of behaviours *within* Survey Content. That is, participants were more consistent over time in reporting certain categories of IPA victimization and also were more consistent over time in their endorsement of certain categories of LA participation. Controlling for the number of instances reported within each category of behaviour did not affect these differences.

*Direction of Inconsistency.* To further understand the nature of report inconsistencies, the direction of observed inconsistencies was explored. First, inconsistencies in the occurrence of behaviours were identified as either being in the direction of decreased disclosure (i.e., participant responded yes at T1, but no at T2) or in the direction of increased disclosure (i.e., participant responded no at T1, but yes at T2). Table 1.6 presents the percentage (and number) of participants evidencing decreased and increased disclosures for the occurrence of behaviours and number of instances reported over time, calculated as a function of the subsample of participants who demonstrated inconsistencies within that category of behaviour. As may be seen in Table 1.6, in general, inconsistency in the direction of decreased disclosure was more common than increased disclosure. With the exception of physical assault (where the pattern was reversed), IPA sample participants generally decreased disclosure more often than they increased disclosure of victimization experiences; that is, they were more likely to endorse experiencing IPA at T1, but then fail to endorse the experience at T2. Chi-square analyses revealed that this directional difference was statistically significant for reports of sexual coercion, $\chi^2(1, N = 26) = 3.85, p < .05, \Phi = .38$, as well as for IPA overall, $\chi^2(1, N = 20) = 5.00, p < .05, \Phi = .50$.

***Table 1.6. Decreased vs. Increased Disclosures for Occurrence of Behaviours and Number of Instances Reported over Time***

| Behaviours | Occurrence | | Number of Instances | |
|---|---|---|---|---|
| | Decreased Disclosure % (*n*) | Increased Disclosure % (*n*) | Decreased Disclosure % (*n*) | Increased Disclosure % (*n*) |
| *IPA Sample* | | | | |
| IPA Overall | 75 (15) | 25 (5) | 63 (60) | 37 (36) |
| Psychological Aggression | 60 (18) | 40 (12) | 65 (59) | 35 (32) |
| Physical Assault | 36 (8) | 64 (14) | 70 (16) | 30 (7) |
| Sexual Coercion | 69 (18) | 31 (8) | 73 (22) | 27 (8) |
| All IPA Categories | 54 (34) | 46 (29) | -- | -- |
| LA Overall | 67 (6) | 33 (3) | 66 (77) | 34 (39) |
| Walking and Related Activities | 57 (16) | 43 (12) | 65 (58) | 35 (31) |
| Conditioning Exercise | 76 (16) | 24 (5) | 67 (62) | 33 (30) |
| Sports | 63 (17) | 37 (10) | 65 (32) | 35 (17) |
| Outdoor Activities | 52 (15) | 48 (14) | 56 (33) | 44 (26) |
| All LA Categories | 62 (40) | 38 (25) | -- | -- |
| *Comparison Sample* | | | | |
| LA Overall | 80 (4) | 20 (1) | 55 (67) | 45 (55) |
| Walking and Related Activities | 75 (15) | 25 (5) | 61 (60) | 39 (38) |
| Conditioning Exercise | 46 (5) | 54 (6) | 53 (55) | 47 (48) |
| Sports | 64 (14) | 36 (8) | 70 (46) | 30 (20) |
| Outdoor Activities | 64 (14) | 36 (8) | 54 (36) | 46 (31) |
| All LA Categories | 63 (34) | 37 (20) | -- | -- |

*Notes.* Occurrence refers to whether or not the participant reported that the behaviour occurred (e.g., Decreased Disclosure = participant reported experiencing physical assault at T1 but denied experiencing physical assault at T2; Increased Disclosure = participant denied experiencing physical assault at T1 but reported experiencing physical assault at T2). Percentages are calculated as a function of the subsample of participants who demonstrated inconsistencies within that category of behaviour. -- = statistic could not be computed.

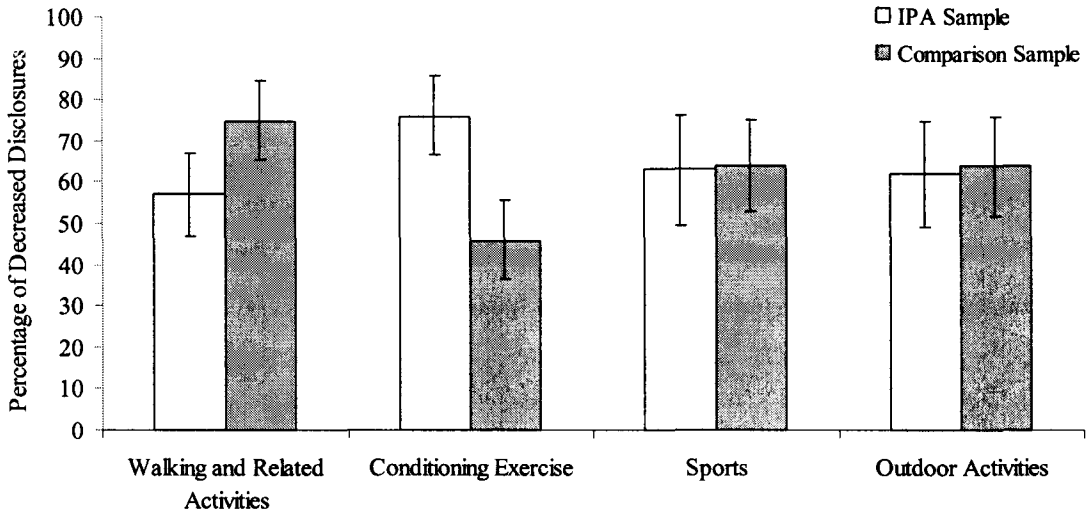
Participants also were significantly more likely to demonstrate inconsistencies in the direction of decreased disclosure for endorsement of walking and related activities $\chi^2(1, N = 48) = 4.08, p < .05, \Phi = .29$, and the number of LA categories endorsed, $\chi^2(1, N = 119) = 9.15, p < .05, \Phi = .28$. No significant differences in the direction of inconsistencies were observed for endorsement of LA participation overall, sports, or outdoor activities, although the patterns were in the same direction (see Table 1.6) and approached significance for Sports ($p = .06$). No significant effects of Participant Sample (main or interactive) where observed. However, comparison sample participants demonstrated a non-significant trend towards increased disclosure of conditioning exercise (see Table 1.6).

Second, inconsistencies in the frequencies of instances reported over time were coded as either decreased or increased disclosure by a simple subtraction of the frequency reported at T2 from the frequency reported at T1. Scores above zero represented inconsistencies in the direction of decreased disclosure (e.g., 20 instances reported at T1 – 10 instances reported at T2 = 10), whereas scores below zero represented inconsistencies in the direction of increased disclosure (e.g., 10 instances reported at T1 – 20 instances reported at T2 = -10). Chi-square analyses revealed that IPA sample participants were significantly more likely to decrease the reported frequency of psychological aggression, $\chi^2(1, N = 91) = 8.01, p < .01, \Phi = .30$, sexual coercion, $\chi^2(1, N = 30) = 6.54, p < .01, \Phi = .47$, and IPA overall, $\chi^2(1, N = 96) = 6.00, p < .05, \Phi = .25$, than to disclose more instances of victimization over time. A similar trend was observed for experiences of physical assault (see Table 1.6), but did not reach significance ($p = .06$).

As for the occurrence of behaviours, participants also were more likely to demonstrate inconsistencies in the direction of decreased disclosure for frequency of walking and related activities, $\chi^2(1, N = 187) = 12.84, p < .001, \Phi = .26$, sports, $\chi^2(1, N = 115) = 14.62, p < .001, \Phi = .36$, and LA participation overall, $\chi^2(1, N = 238) = 10.50, p < .001, \Phi = .21$. No significant differences in the direction of inconsistencies over time were observed for reported frequencies of outdoor activities and LA participation overall ($p$'s > .05). Again, analyses did not reveal significant effects (main or interactive) of Participant Sample ($p$'s > .05).

Given the observed differences in consistency over time between categories of behaviours, the direction of inconsistencies within Survey Content (IPA vs. LA) also was examined. IPA sample participants were significantly more likely to decrease disclosure of sexual coercion (69% of participants who demonstrated inconsistencies in reports of sexual coercion) compared with physical assault (36% of participants who demonstrated inconsistencies in reports of physical assault), $t(44) = 2.40, p < .05, d = .72$. As may be seen in Figure 1.4, likelihood of decreased disclosure did not differ significantly across LA categories, whereas comparison sample participants were significantly more likely to decrease disclosure regarding the frequency of participation in conditioning exercise compared with sports, $t(167) = 2.20, p < .05, d = .34$. No other significant differences in the direction of inconsistencies were observed between categories of behaviours.

***Figure 1.4. Percentage of Participants Demonstrating Inconsistencies in the Direction of Decreased Disclosure for Walking and Related Activities, Conditioning Exercise, Sports, and Outdoor Activities***



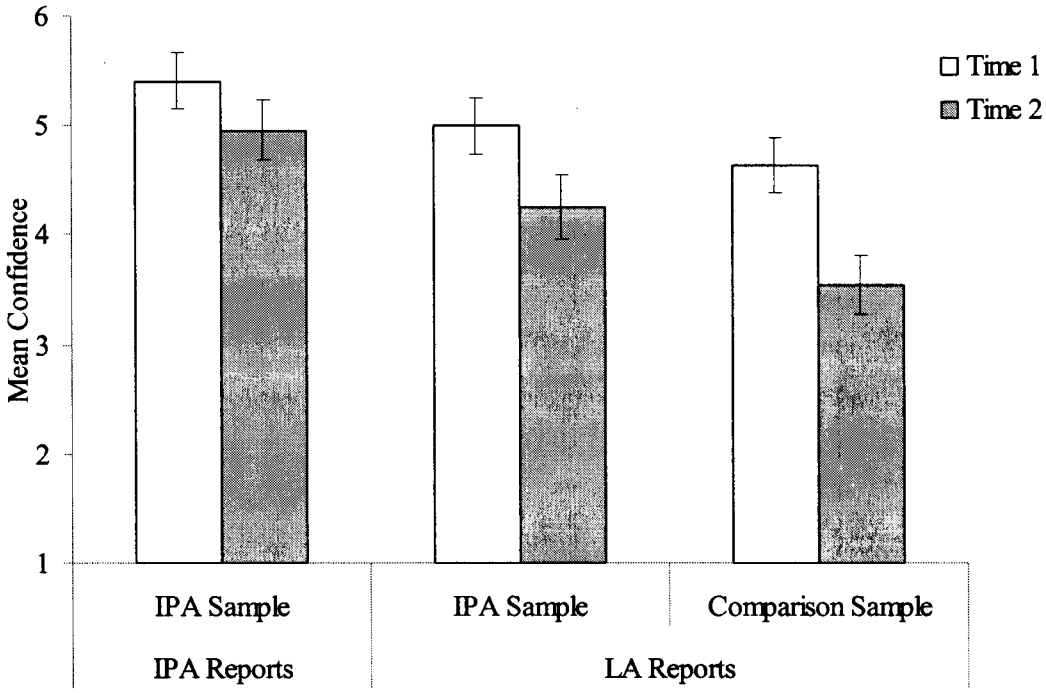

*Notes.* Walking and Related Activities $n = 98$; Conditioning Exercise $n = 103$; Sports $n = 66$; Outdoor Activities $n = 67$. Decreased Disclosure = Yes at Time 1, No at Time 2. Bars represent 95% confidence intervals around the percentage. Percentage calculated as a function of the subsample of participants who demonstrated inconsistencies within that category of behaviour.

*Confidence*

For reasons described earlier, additional analyses were conducted to examine whether report consistency differed systematically with confidence in the accuracy of reports. Overall, participants were quite confident in the accuracy of their reports. On a scale from 1 (Not at all confident) to 6 (Very confident), average confidence in IPA reports was 5.14 ($SD$ = 1.55) at T1 and 4.96 at T2 ($SD$ = 1.64), with a mean of 5.19 ($SD$ = 1.32) across time. Confidence for LA reports was slightly lower with an average rating of 5.00 ($SD$ = 1.55) at T1, 4.28 ($SD$ = 1.78) at T2, and 4.64 ($SD$ = 1.41) across time for IPA sample participants, and 4.64 ($SD$ = 1.43) at T1, 3.54 ($SD$ = 1.54) at T2, and 4.08 ($SD$ = 1.27) across time for comparison sample participants.

Repeated measures ANOVAs revealed significant main effects of Time and Participant Sample on confidence ratings (see Figure 1.5). IPA sample participants were significantly more confident in the overall accuracy of their IPA reports at T1 than at T2, $F(1, 136) = 19.71, p < .001, \eta_p^2 = .06$. Confidence in the accuracy of LA reports also decreased significantly over time across both samples ($M_{T1} = 4.82$, $SD = 1.50$; $M_{T2} = 3.90$, $SD = 1.70$), $F(1, 272) = 80.36, p < .001, \eta_p^2 = .23$. IPA sample participants were significantly more confident in the accuracy of their LA reports overall ($M = 4.64$, $SD = 1.41$) than were comparison sample participants ($M = 4.08$, $SD = 1.27$), $F(1, 272) = 11.30, p < .001, \eta_p^2 = .04$. No significant differences between IPA sample participants' confidence in the overall accuracy of IPA and LA reports were observed ($p > .05$).

***Figure 1.5. Mean Confidence Ratings in the Overall Accuracy of Reports as a Function of Participant Sample and Survey Content***




*Notes.* Bars represent 95% confidence intervals around the means. 1 = Not at all confident; 6 = Very confident.

***Table 1.7. Descriptive Statistics for Confidence Ratings in the Accuracy of Intimate Partner Abuse and Leisure Activities Reported***

| Behaviours | Occurrence | | | Number of Instances | | |
|---|---|---|---|---|---|---|
| | Overall *M* (*SD*) | T1 *M* (*SD*) | T2 *M* (*SD*) | Overall *M* (*SD*) | T1 *M* (*SD*) | T2 *M* (*SD*) |
| *IPA Sample* | | | | | | |
| IPA Reports | | | | | | |
| Psychological Aggression | 5.22 (0.86) | 5.26 (1.10) | 5.19 (1.13) | 3.83 (1.40) | 4.11 (1.53) | 3.61 (1.60) |
| Physical Assault | 5.52 (0.66) | 5.75 (0.56) | 5.28 (1.13) | 4.31 (1.55) | 4.86 (1.18) | 3.85 (1.64) |
| Sexual Coercion | 5.50 (0.68) | 5.60 (0.81) | 5.40 (0.87) | 4.24 (1.28) | 4.33 (1.32) | 3.94 (1.63) |
| LA Reports | | | | | | |
| Walking and Related Activities | 5.24 (0.88) | 5.30 (1.19) | 5.18 (1.19) | 4.15 (1.28) | 4.52 (1.34) | 3.77 (1.64) |
| Conditioning Exercise | 5.22 (0.84) | 5.36 (1.06) | 5.09 (1.14) | 4.03 (1.27) | 4.29 (1.38) | 3.71 (1.64) |
| Sports | 5.22 (0.88) | 5.47 (0.96) | 4.98 (1.25) | 4.20 (1.29) | 4.49 1.26) | 3.94 (1.42) |
| Outdoor Activities | 5.21 (0.80) | 5.47 (0.85) | 4.98 (1.10) | 4.17 (1.39) | 4.51 (1.53) | 3.79 (1.54) |
| *Comparison Sample* | | | | | | |
| LA Reports | | | | | | |
| Walking and Related Activities | 4.93 (1.17) | 5.23 (1.21) | 4.62 (1.63) | 3.74 (1.25) | 4.07 (1.31) | 3.34 (1.49) |
| Conditioning Exercise | 5.14 (0.95) | 5.30 (1.08) | 4.99 (1.29) | 3.45 (1.23) | 3.91 (1.35) | 3.02 (1.50) |
| Sports | 5.01 (1.04) | 5.26 (1.10) | 4.77 (1.38) | 3.60 (1.26) | 3.93 (1.45) | 3.23 (1.56) |
| Outdoor Activities | 5.28 (0.91) | 5.44 (0.97) | 5.12 (1.24) | 3.94 (1.23) | 4.24 (1.29) | 3.50 (1.49) |

*Notes.* Occurrence refers to whether or not the participant reported that the behaviour occurred (yes vs. no). IPA sample $n = 138$; comparison sample $n = 138$. 1 = Not at all confident; 6 = Very confident.

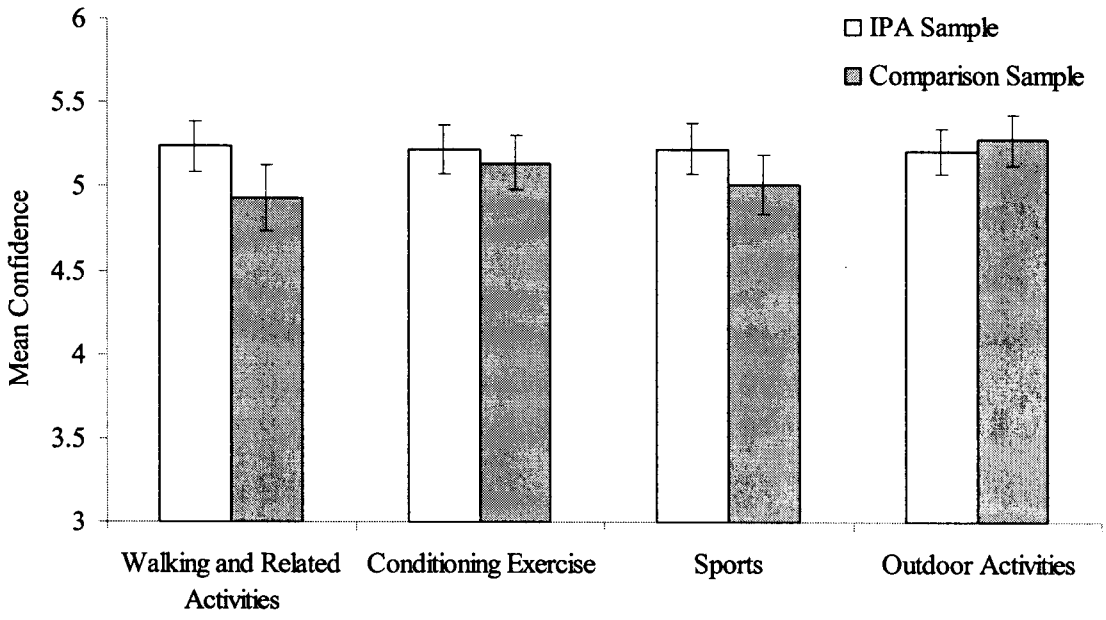
Table 1.7 presents descriptive statistics for mean confidence ratings in reports of IPA victimization and LA participation by category as a function of Participant Sample. As may be seen in Table 1.7, mean confidence in the accuracy of reports within categories of behaviours also was high, and, as for ratings in confidence of reports overall, there were no instances in which mean confidence fell below the midpoint of the scale. Repeated measures ANOVAs revealed that confidence in the accuracy of the occurrence of behaviours differed significantly as a function of IPA category, $F(2, 272) = 9.60, p < .001, \eta_p^2 = .07$, but not for the frequency of IPA reported ($p > .05$). *Post hoc* comparisons revealed that participants were significantly less confident in the accuracy of whether they experienced psychological aggression ($M = 5.22$, $SD = 0.86$) compared with both physical assault ($M = 5.52$, $SD = 0.66$), $t(137) = 3.51, p < .001, d = .60$, and sexual coercion ($M = 5.50$, $SD = 0.68$), $t(137) = 3.32, p < .001, d = .57$. The difference between confidence ratings for physical assault and sexual coercion was not significant ($p > .05$). A repeated measures ANCOVA revealed that differences in confidence as a function of IPA category were no longer significant after controlling for the mean frequency of instances reported within each category ($p > .05$).

With respect to reports of LA participation, both within and across samples, analyses revealed no significant differences in confidence by LA category for occurrence and frequency of behaviours ($p$'s $> .05$). A repeated measures ANOVA, however, revealed a significant main effect of category on confidence in reports of the occurrence of LA behaviours, $F(3, 822) = 2.88, p < .05, \eta_p^2 = .01$, as well as a significant Category x Participant Sample interaction, $F(3, 822) = 4.04, p < .01, \eta_p^2 = .07$. *Post hoc* comparisons demonstrated that comparison sample participants' confidence ratings for endorsement of

outdoor activities was significantly higher than for the other three categories, $t$'s (137) $\geq$ 2.03, $p$'s $\leq .05$, $d$'s $\geq .35$, as were confidence ratings for conditioning exercise compared with walking and related activities (see Figure 1.6), $t(137) = 2.32$, $p < .05$, $d = .40$. In contrast, confidence ratings for IPA sample participants did not differ significantly across LA categories ($p > .05$). The differences in confidence as a function of LA category remained significant after controlling for the mean frequency of instances reported within each category, $F(3, 231) = 3.36$, $p < .05$, $\eta_p^2 = .05$; however, the Category x Participant Sample interaction was no longer significant ($p > .05$).

Table 1.8 presents correlations between mean confidence ratings and inconsistency scores. Very few significant associations were observed; correlations were, however, in the expected direction when observed (i.e., confidence decreased as inconsistency increased). No significant associations were observed between confidence ratings and the frequency of IPA experiences reported; however, mean confidence across time and at T1 was significantly correlated with consistency of reports of the number of IPA categories endorsed (see Table 1.8). Similarly, mean confidence ratings were significantly correlated with consistency of the occurrence of sexual coercion across time. With respect to LA reports, as for IPA reports, comparison sample participants' ratings of confidence across time and at T1 decreased as inconsistency in the endorsement of LA participation increased. Confidence at T1 also decreased as inconsistency in the frequency of conditioning exercise increased. No significant associations were observed between confidence ratings and either the occurrence or frequency of LA participation in the IPA sample (see Table 1.8).

***Figure 1.6. Category x Participant Sample Interaction Effect on Mean Confidence in the Accuracy of Leisure Activities Reported***




*Notes.* Bars represent 95% confidence intervals around the means. 1 = Not at all confident; 6 = Very confident.

***Table 1.8. Correlations between Mean Confidence Ratings and Proportion and Degree Inconsistent Scores as a Function of Participant Sample***

| Behaviours | Proportion Inconsistent | | | Degree Inconsistent | | |
|---|---|---|---|---|---|---|
| | Overall | T1 | T2 | Overall | T1 | T2 |
| *IPA Sample* | | | | | | |
| IPA Reports | | | | | | |
| All IPA Categories | -.21* | -.32*** | -.04 | -- | -- | -- |
| Psychological Aggression | -.07 | -.03 | .08 | -.07 | -.18 | .05 |
| Physical Assault | .00 | -.08 | .05 | .11 | .08 | .07 |
| Sexual Coercion | -.29*** | -.20* | -.26** | .29 | .17 | .31 |
| All IPA Instances | -.07 | -.14 | .02 | .08 | .09 | .10 |
| LA Reports | | | | | | |
| All LA Categories | .13 | .03 | .17 | -- | -- | -- |
| Walking and Related Activities | -.01 | -.02 | .00 | -.14 | -.12 | -.18 |
| Conditioning Exercise | -.05 | .04 | -.12 | -.17 | -.16 | -.18 |
| Sports | .09 | .12 | .03 | -.04 | -.21 | .08 |
| Outdoor Activities | -.03 | -.04 | .03 | .03 | .01 | -.04 |
| All LA Instances | -.07 | -.12 | -.02 | -.06 | -.08 | -.03 |
| *Comparison Sample* | | | | | | |
| LA Reports | | | | | | |
| All LA Categories | -.02 | -.05 | .02 | -- | -- | -- |
| Walking and Related Activities | -.10 | -.04 | -.11 | .03 | .12 | -.18 |
| Conditioning Exercise | .06 | .14 | -.03 | -.15 | -.19* | -.10 |
| Sports | .03 | .06 | -.01 | .05 | .07 | -.05 |
| Outdoor Activities | -.11 | -.04 | -.13 | -.03 | -.07 | -.09 |
| All LA Instances | -.21* | -.22** | -.14 | .00 | -.06 | .06 |

*Notes.* * $p < .05$. ** $p < .01$. *** $p < .001$. -- = statistic could not be computed.

Further analyses were conducted using repeated measures ANOVAs with direction as a grouping variable (decreased disclosure vs. increased disclosure) to determine whether confidence ratings were associated with the direction of inconsistencies observed. Results were significant for only one comparison: Across samples, mean confidence ratings were significantly lower for participants who evidenced decreased disclosure inconsistencies (i.e., yes at T1 and no at T2) in their endorsement of participation in sports across interviews ($M = 5.28$, $SD = 0.83$) compared to participants who increased disclosure ($M = 5.48$, $SD = 0.80$), $F(1, 41) = 6.57$, $p < .01$, $\eta_p^2 = .14$.

On the whole, few significant associations between confidence and report consistency were found and there were no differences between confidence in reports of abusive and everyday events among IPA participants. Results again highlighted differences *between* categories of behaviours *within* Survey Content. Specifically, IPA sample participants evidenced lower confidence in the accuracy of their reports of whether they experienced psychological aggression compared with both physical assault and sexual coercion, it would seem as a function of the frequency of instances within each category. Comparison sample participants demonstrated greater confidence in reports of whether they participated in outdoor activities, as well as conditioning exercise compared with walking and related activities. Analyses also revealed the general tendency for confidence to decrease over time with decreasing consistency.

*Gender*

As reviewed in the Introduction, female-perpetrated IPA against men has received considerably less attention despite evidence that perpetration prevalence rates are fairly

comparable across gender (Dutton & Nicholls, 2005). There is also little research investigating whether men and women differ in consistency of reporting IPA victimization (or traumatic experiences more generally). For reasons reviewed earlier, male victims may be less likely than female victims to disclose their experiences (e.g., Felson, Messner, & Hoskin, 1999); therefore, for male participants who *do* disclose abusive experiences, their reluctance to discuss such experiences may affect the consistency of their reports. Thus, additional analyses were conducted to explore possible gender differences.

*Occurrence of Behaviours.* Descriptive statistics for the occurrence and frequency of IPA victimization and LA participation reported are presented in Tables 1.9 and 1.10 as a function of Participant Gender. Very few significant differences were observed in the occurrence of either IPA or LA behaviours. With the exception of physical assault, the frequency of IPA reported by female participants was greater on average than the frequency reported by male participants (see Tables 1.9 and 1.10). Differences, however, failed to reach statistical significance ($p$'s > .05). In fact, the only significant differences observed were with regard to reports of LA participation. Chi-square analyses demonstrated that male participants were significantly more likely to report participating in sports at T1 and T2 than were female participants, $\chi^2$'s $(1, N = 276) \geq 8.91$, $p$'s $\leq .01$, $\Phi$'s $\geq .18$ (see Table 1.10). Male participants did not differ significantly from female participants ($p$'s > .05) on the number of categories of LA endorsed at T1 ($M_{female} = 2.86$, $SD = 1.03$; $M_{male} = 3.04$, $SD = 1.06$) or T2 ($M_{female} = 2.72$, $SD = 1.10$; $M_{male} = 2.80$, $SD = 1.07$).

***Table 1.9. Occurrence and Frequency of Intimate Partner Abuse Reported by Female and Male Participants***

| Behaviours | Participants who Reported Behaviour Occurred | | Number of Instances Reported | |
|---|---|---|---|---|
| | T1 | T2 | T1 | T2 |
| | % (*n*) | % (*n*) | *M* (*SD*) | *M* (*SD*) |
| *Female Participants* | | | | |
| IPA Overall | 96 (106) | 91 (101) | 17.86 (34.94) | 11.54 (17.46) |
| Psychological Aggression | 90 (100) | 88 (98) | 14.72 (30.44) | 9.22 (14.93) |
| Physical Assault | 21 (23) | 27 (29) | 2.77 (3.37) | 2.24 (2.39) |
| Sexual Coercion | 34 (38) | 26 (29) | 7.55 (9.44) | 5.43 (9.56) |
| *Male Participants* | | | | |
| IPA Overall | 100 (27) | 82 (22) | 12.63 (10.64) | 7.37 (15.75) |
| Psychological Aggression | 93 (25) | 78 (21) | 11.11 (9.02) | 6.89 (15.22) |
| Physical Assault | 19 (5) | 19 (5) | 3.67 (4.80) | 1.20 (2.17) |
| Sexual Coercion | 22 (6) | 19 (5) | 3.60 (2.30) | 2.00 (2.92) |

*Notes.* IPA sample $n = 138$ (111 female participants; 27 male participants).

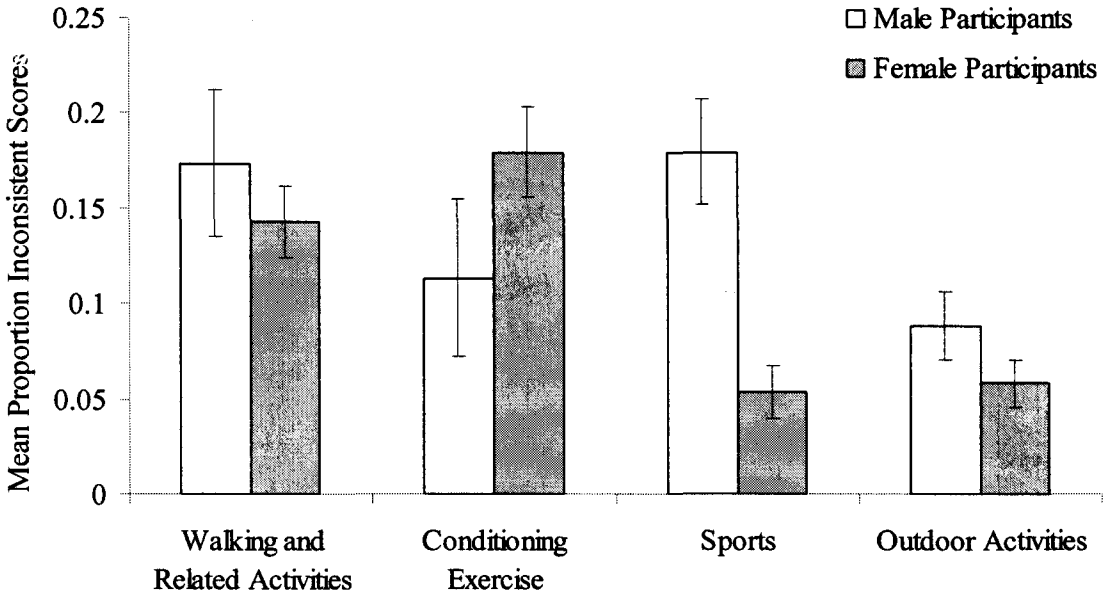
***Table 1.10. Occurrence and Frequency of Leisure Activities Reported by Female and Male Participants***

| Behaviours | Participants who Reported Behaviour Occurred | | Number of Instances Reported | |
|---|---|---|---|---|
| | T1<br>% (*n*) | T2<br>% (*n*) | T1<br>*M* (*SD*) | T2<br>*M* (*SD*) |
| *Female Participants* | | | | |
| LA Overall | 99 (220) | 96 (214) | 65.04 (89.69) | 55.37 (88.21) |
| Walking and Related Activities | 89 (198) | 86 (191) | 26.67 (38.31) | 23.44 (39.19) |
| Conditioning Exercise | 82 (182) | 78 (172) | 34.44 (67.86) | 29.06 (71.21) |
| Sports | 51 (114) | 48 (106) | 14.88 (24.97) | 12.79 (23.21) |
| Outdoor Activities | 64 (141) | 61 (135) | 7.81 (12.56) | 6.26 (9.95) |
| *Male Participants* | | | | |
| LA Overall | 96 (52) | 96 (52) | 65.24 (106.23) | 54.12 (102.81) |
| Walking and Related Activities | 91 (49) | 78 (42) | 30.95 (80.63) | 19.42 (52.17) |
| Conditioning Exercise | 76 (41) | 76 (41) | 29.85 (49.44) | 32.23 (93.88) |
| Sports | 80 (43) | 71 (38) | 15.64 (22.48) | 12.82 (21.06) |
| Outdoor Activities | 57 (31) | 57 (30) | 13.36 (29.34) | 11.14 (24.64) |

*Notes.* $n = 276$ (222 female participants; 54 male participants).

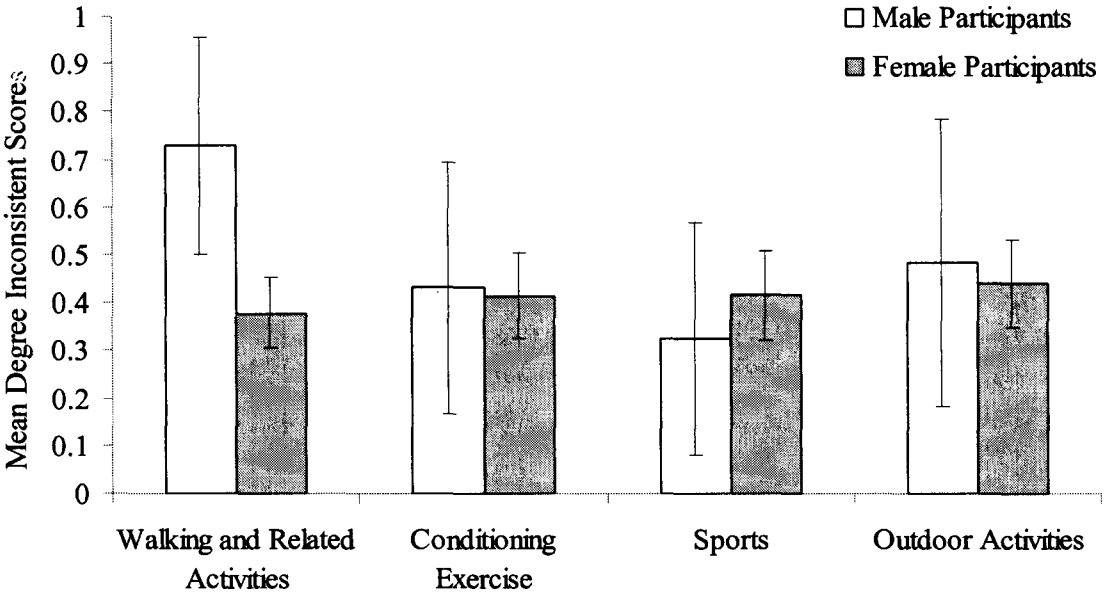
*Consistency of Reports over Time.* All comparisons between male and female participants regarding the consistency of IPA victimization reports failed to produce significant differences ($p$'s > .05). Male and female participants did not differ significantly in consistency of IPA victimization experiences reported at T1 and T2, nor were there any interactive effects of gender and IPA category. Again, when differences in consistency were observed, they were found for reports of LA participation. Specifically, repeated measures ANOVAs revealed a modest but significant interaction between participant gender and LA category for both Proportion Inconsistent scores, $F(3, 822) = 6.65$, $p < .001$, $\eta_p^2 = .02$, and Degree Inconsistent scores, $F(3, 816) = 6.61$, $p < .001$, $\eta_p^2 = .02$. As may be seen in Figure 1.7, male participants were more consistent in their endorsement of conditioning exercises, $t(274) = 2.54$, $p < .05$, $d = .31$, and less consistent in their endorsement of sports than female participants, $t(274) = 2.20$, $p < .05$, $d = .27$. As may be seen in Figure 1.8, female participants also were significantly more consistent in the frequency of walking and related activities reported, $t(274) = 2.03$, $p < .05$, $d = .25$. Further, whereas male participants differed significantly across LA categories in the consistency of frequency reported, $F(3, 207) = 3.85$, $p < .05$, $\eta_p^2 = .26$, female participants showed no such differences ($p > .05$).

***Figure 1.7. Mean Proportion Inconsistent Scores for Walking and Related Activities, Conditioning Exercise, Sports, and Outdoor Activities as a Function of Participant Gender***




*Notes.* Bars represent 95% confidence intervals around the means. The possible range of scores was from 0 to 1.00, where a score of zero represents no inconsistency.

***Figure 1.8. Mean Degree Inconsistent Scores for Walking and Related Activities, Conditioning Exercise, Sports, and Outdoor Activities as a Function of Participant Gender***




*Notes.* Bars represent 95% confidence intervals around the means. The possible range of scores was from 0 to 1.00, where a score of zero represents no inconsistency.

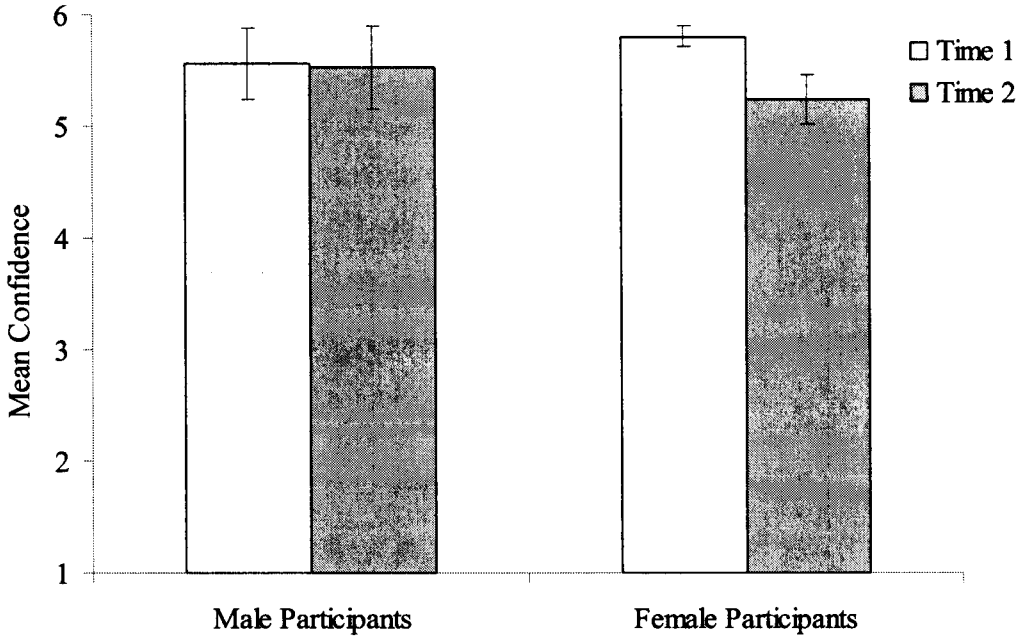
The direction of report inconsistencies over time also was explored as a function of participant gender. Chi-square analyses revealed that male participants were significantly more likely to evidence decreased disclosure with regard to the frequency of psychological aggression (88%, $n = 14$) compared with female participants (60%, $n = 30$), $\chi^2(1, N = 91) = 4.37, p < .05, \Phi = .22$. This pattern was found after controlling for the mean frequency of psychological aggression reported, but was no longer significant ($p = .06$). A similar pattern was observed for the frequency of IPA reported overall (male: 82%, $n = 14$; female: 42%, $n = 33$), but did not reach significance ($p = .06$). In the comparison sample male participants were more likely to recant their endorsement of LA categories (92%, $n = 11$) than female participants (60%, $n = 25$), $\chi^2(1, N = 54) = 4.34, p < .05, \Phi = .28$. Controlling for the number of LA categories endorsed did not alter the pattern; however, the gender difference was no longer significant ($p = .09$).

*Confidence.* Few significant differences in confidence between male and female participants were observed. In fact, there were no significant gender differences for confidence in the accuracy of LA reports at T1 or T2, between or within IPA and comparison sample participants ($p$'s > .05) However, a repeated measures ANOVA demonstrated a significant main effect of Participant Gender on confidence in the accuracy of reports of the frequency of psychological aggression: Male participants reported significantly greater confidence on average than female participants ($M_{male} = 4.28$, $SD = 1.53$; $M_{female} = 3.72$, $SD = 1.35$), $F(1, 105) = 5.66, p < .05, \eta_p^2 = .05$. This difference remained statistically significant after controlling for the frequency of psychological aggression reported, $F(2, 104) = 5.78, p < .05, \eta_p^2 = .05$.

Analyses also revealed a significant Time x Participant Gender interaction effect on confidence ratings for reports regarding whether participants experienced physical assault, $F(1, 136) = 4.57, p < .05, \eta_p^2 = .03$. Specifically, as may be seen in Figure 1.9, *post hoc* comparisons demonstrated that female participants' confidence in the accuracy of their reports decreased significantly over time ($M_{T1} = 5.80$, $SD = 0.48$; $M_{T2} = 5.23$, $SD = 1.17$), $t(110) = 5.29, p < .001, d = 1.01$, whereas male participants' confidence ($M_{T1} = 5.56$, $SD = 0.80$; $M_{T2} = 5.52$, $SD = 0.94$) remained fairly constant ($p > .05$). This difference was no longer significant after controlling for frequency of physical assault reported ($p > .05$).

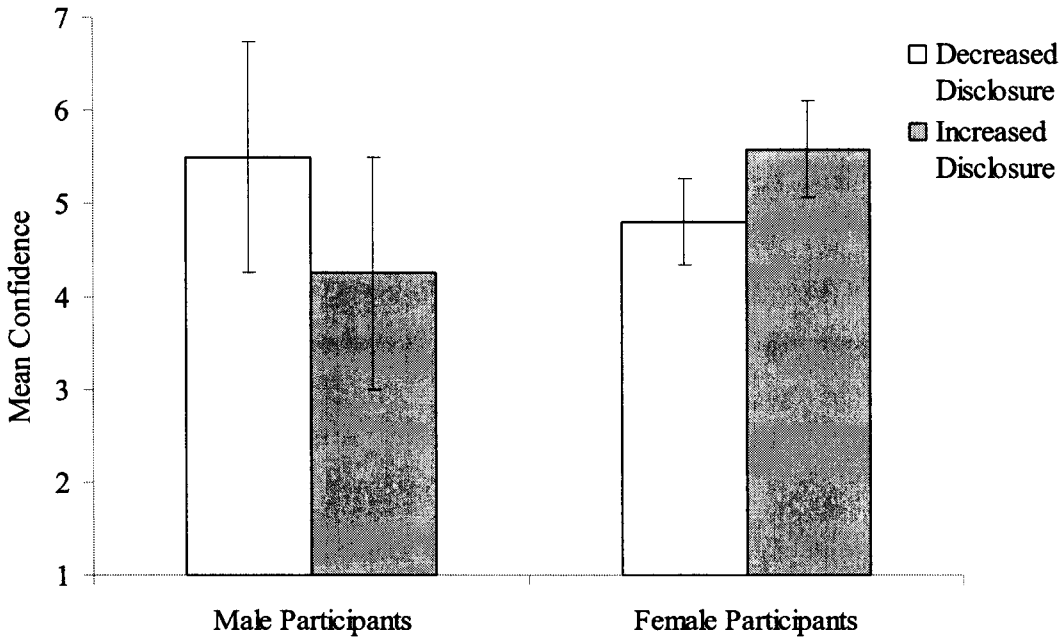
Finally, depicted in Figure 1.10, a repeated measures ANOVA demonstrated a significant Gender x Direction of Inconsistency effect on confidence ratings in reports regarding whether participants experienced sexual coercion, $F(1, 22) = 5.79, p < .03, \eta_p^2 = .21$. *Post hoc* comparisons revealed that female participants who decreased disclosure were significantly less confident in their reports ($M = 4.80$, $SD = 0.84$) than were female participants who disclosed sexual coercion experiences at T2, but not at T1 ($M = 5.58$, $SD = 0.49$), $t(19) = 2.12, p < .05, d = .97$. As may be seen in Figure 1.10, although the difference did not reach significance ($p > .05$), the trend was reversed for male participants (decreased disclosure: $M_{confidence} = 5.50$, $SD = 0.50$; increased disclosure: $M_{confidence} = 4.25$, $SD = 1.77$). This difference remained statistically significant even after controlling for frequency of sexual coercion reported, $F(2, 33) = 7.02, p < .05, \eta_p^2 = .18$.

***Figure 1.9. Time x Participant Gender Interaction Effect on Mean Confidence in Reports of Physical Assault Experiences***




*Notes.* Bars represent 95% confidence intervals around the means. 1 = Not at all confident; 6 = Very confident.

***Figure 1.10. Direction of Inconsistency x Participant Gender Interaction Effect on Mean Confidence in Reports of Sexual Coercion Experiences***




*Notes.* Male participants $n = 5$; female participants $n = 21$. Decreased Disclosure = Yes at Time 1, No at Time 2; Increased Disclosure = No at Time 1; Yes at Time 2. Bars represent 95% confidence intervals around the means. 1 = Not at all confident; 6 = Very confident.

Generally, male and female participants were quite similar in report consistency. Analyses revealed no gender differences in the frequency, occurrence, or consistency of IPA victimization reports. Differences for LA reports showed increased consistency for reports made by female participants. The directionality of inconsistencies and confidence in the accuracy of reports also were very similar. When observed, differences demonstrate that male participants were more likely than female participants to be evidence decreased disclosure, but also generally more confident in the accuracy of their reports.

## *Discussion*

Using methodology reflective of the state-of-the-art and designed to maximize report consistency, the purpose of this study was to compare the consistency of reports of abusive experiences and everyday event prevalence over time. Specifically, this study examined whether the consistency of reports of IPA victimization differed quantitatively or qualitatively (i.e., with regard to type of inconsistency, occurrence or frequency, and direction of inconsistency, increased or decreased disclosure) from the consistency of reports of everyday experiences (i.e., LA participation), *within* a sample of IPA victims and *between* victim and comparison samples.

In contrast with recent findings demonstrating superiority of recall for traumatic memory events (e.g., Krinsley et al. 2003; Peace & Porter, 2004; Porter & Birt, 2001; Porter & Peace, 2007), comparisons between abuse and everyday autobiographical memory event prevalence in the present study revealed few significant differences. Results instead are similar to findings of previous research showing few differences between memory for traumatic and non-traumatic events in terms of quality or

consistency (e.g., Berntsen, Willert, & Rubin, 2003; Bohanek, Fivush & Walker, 2005; Geraerts et al., 2007; Geraerts et al., 2007; Herlihy, Scragg, & Turner, 2002; Koss, Figueredo, Bell, Tharan, & Tromp, 1996; Peace, Porter, & ten Brinke, 2008[11]). Across comparisons, results failed to provide evidence that memories of abuse are unique, or that they are more or less consistent (and arguably, more or less accurate) in terms of event prevalence (i.e., occurrence and frequency), than other memories. Specifically, analyses failed to identify qualitative or quantitative differences in the consistency of reports of event prevalence between abusive and everyday autobiographical experiences.

There are important differences between the present study's analyses of consistency of reports of the prevalence of abusive and everyday autobiographical experiences and comparisons drawn between reports of traumatic and non-traumatic experiences in past research. Most importantly, the data discussed here represent reports of the occurrence and frequency of events rather than of the details of events. There may very well exist differences in the latter, as suggested by research demonstrating inferiority and superiority of traumatic memory, without the former varying significantly. In particular, the work referenced above compared characteristics of narrative reports of a single memory event across memory type (i.e., traumatic and everyday) whereas the present study considered occurrence and frequency of abusive experiences which may have been distributed over multiple events. Although some participants in the present study may have reported multiple abusive experiences that occurred within one violent transaction, more analogous to the single events elicited in the above mentioned work, other participants may have reported abusive experiences that occurred across repeated

[11] Although analyses identified some differences in memory characteristics, no main effect of the nature of the memory event (i.e., sexual trauma, non-sexual trauma, positive experience) on coherence was found.

single events. Additionally, volunteers for the IPA sample were recruited to participate in a study examining abusive experiences, and not traumatic experiences *per se*. Although any experience of IPA may be traumatic by definition, IPA sample participants were not asked to rate the level of emotional 'trauma' associated with these experiences and thus, distinction between the abusive and everyday experiences as being traumatic versus non-traumatic is not possible.

Nonetheless, findings of the present study suggest that consistency of behavioural reports over time may depend not on who but what you ask (i.e., within categories of autobiographical memory events), regardless of whether the memories are of abusive or non-abusive experiences. When observed, significant differences in consistency were found *between* categories of behaviours *within* event type (abusive vs. everyday). That is, participants were more consistent in reporting the occurrence and frequency certain categories of IPA victimization and were similarly more consistent in their reports for certain categories of LA participation. Not surprisingly, comparison of the occurrence and frequency of behaviours reported with inconsistencies observed in those categories demonstrates that the more common or frequent the behaviour, the more likely participants were to be inconsistent. For example, overall, Psychological Aggression was the most commonly endorsed category of IPA; it was also the IPA category for which the greatest inconsistency in reports over time was observed. Similarly, Walking and Related Activities was the most commonly endorsed LA category and it was also the LA category for which the greatest inconsistency in reports over time was observed. Indeed, controlling for the frequency of events reported eliminated some of these differences. Such findings add to a well-established literature demonstrating that the more often an

event is experienced, the more difficult it is to estimate frequency and to recall specific instances of the event (e.g., Bogart et al., 2007; Conrad, Brown, & Cashman, 1998; Means & Loftus, 1991; Smith, Jobe, & Mingay, 1991; *cf.* Jobe, Tourageau, & Smith, 1993; Pearson et al., 1992; Read & Connolly, 2007; Wright & Loftus, 1998). Such findings also may reflect an increased range for detecting change with more frequently occurring events.

The finding of differential consistency as a function of category of behaviours (even after controlling for frequency of behaviours) is in line with past research comparing consistency across categories of abusive experiences. For example, in their study of the consistency of 222 children's reports of sexual and physical abuse, Ghetti, Goodman, Eisen, Qin, and Davis (2002) found that reports of the prevalence of sexual abuse were more consistent over repeated interviews than were reports of physical abuse. Similarly, Dill, Chu, Grob, and Eisen (1991) found that female psychiatric patients' reports of prevalence of sexual abuse were more consistent across repeated questioning than were reports of physical abuse. Interestingly, however, differences observed in the present study were in the opposite direction of those found in the above cited research: There was a trend for greater consistency to be found in reports of physical assault compared to reports of sexual coercion. Differences between the patterns of results of these studies and the present research may be attributable to characteristics of the participants and experiences. Specifically, in the research of Ghetti et al., children reported on childhood abuse and respondents in the Dill et al. study were female psychiatric patients. No research, to my knowledge, has included comparisons with reports of psychological abuse.

With regard to the direction of inconsistencies observed, participants across samples were overwhelmingly more likely to decrease disclosure of their experiences over time (i.e., at T1 indicating that the behaviour had occurred and at T2 indicating that the behaviour had not occurred). However, increased disclosure over time was observed for one category of IPA victimization: physical assault. Instead of decreasing disclosure of physical assault experiences, participant reports were indicative of a phenomenon known as *reminiscence* (i.e., increased availability of new information over repeated recall; Herlihy et al., 2002). Such reminiscence may have resulted from a continued memory search for physical assault experiences after the baseline session, contributing to increased remembering over time. However, participants may have consciously withheld endorsement of physical assault during the baseline session, but decided to report their experiences during the follow-up session. The data do not allow for a distinction between these two possibilities. Similarly, there are many reasons for observing inconsistencies in the direction of decreased disclosure, from increased memory impairment, such as *retroactive memory interference* (i.e., the encoding in memory of new information that interferes with the retrieval of previously learned information), simple memory decay or forgetting, and decreased participant motivation over time. With the present design, it is not possible to identify the reasons for inconsistencies. I return to this limitation later.

The overall pattern of decreased disclosure observed for IPA victimization, and for psychological aggression and sexual coercion in particular, is consistent with past research demonstrating reduced disclosure of traumatic experiences over time. For example, Spinhoven, Bean, and Eurelings-Bontekoe's (2006) found that, overall, refugees often reported traumatic experiences during the baseline interview that were

subsequently omitted during a follow-up interview and with respect to reports of sexual abuse in particular. Similarly, Mollica, Caridad, and Massagli (2007) found that when inconsistencies in refugees' reports of experiences of wartime trauma and torture were observed, they were most likely to be in the direction of decreased endorsement at follow-up; that is, event prevalence was greater at baseline compared to follow-up. Lee and Brown (2003) also found that participants reported significantly less information over time in their study of recollections for the events of September 11$^{th}$, 2001.

However, with the exception of the increased disclosure of physical assault, the direction of inconsistency findings in the present study are in contrast with other studies demonstrating increased disclosure of traumatic experiences, and autobiographical events more generally, over time (e.g., Rubin & Wenzel, 1996). For instance, Krinsley et al. (2003) investigation of exposure to traumatic events reported by 76 male military veterans demonstrated significant increases in the number of traumatic events reported over time. In fact, 51% of respondents reported more traumatic experiences over time compared with 38% who reported fewer experiences over time. Similarly, 70% of Southwick et al.'s (1997) sample of 59 Operation Desert Storm veterans reported at least one more combat-related traumatic event at follow-up that they had not reported during the baseline session, compared with 46% who retracted at follow-up one or more experiences they had reported at baseline.

Again, there exist substantial differences between the present study and the above-reviewed research preventing direct comparison of results. Direction of inconsistencies observed in the present study were examined as a function of category of abusive experience, and category of IPA victimization specifically, whereas past research

generally has considered a variety of traumatic experiences in aggregate. It is possible, then, that a more fine-grained analysis of the direction of inconsistencies found in past research would reveal patterns more in line with those observed in the present study. Information regarding what participants were doing over the interval between questioning sessions also would be informative; undergoing therapy for traumatic experiences would potentially contribute to increased disclosure over time, for example. Further, many of the participants sampled in the cited studies (e.g., wartime veterans and refugees) arguably experienced more traumatic events than those sampled in the present study and thus, might have experienced PTSD at greater rates than in the present sample.

Analyses of the association between confidence and consistency in reports of event prevalence demonstrated few differences overall, and no differences as a function of recall for abusive experiences versus everyday events. Results, however, did highlight the general tendency for confidence to decrease over time with decreasing consistency. This finding is in accordance with some research demonstrating general decreases in confidence over repeated questioning (e.g., Coluccia, Bianco, & Bradimonte, 2006). In contrast, other research demonstrates that confidence in statements for which a public commitment was made, such as eyewitness identification, can, and often does, increase over time despite decreased accuracy (Shaw, McClure, & Dystrak, 2007). With regard to traumatic and everyday memory events, research comparing characteristics of single events generally has demonstrated increased confidence in the accuracy of reports of traumatic experiences that is persistent over time. Weaver (1993), for instance, found that confidence in the accuracy of memory for traumatic events was higher than for memories of a non-traumatic event and did not decrease over time, as did Paradis, Solomon, Florer

and Thompson (2004). Talarico and Rubin (2003, 2007) similarly found belief in memory accuracy declined over time for everyday but not traumatic memories, leading the authors to suggest that it is such high levels of confidence that characterise so-called *flashbulb memories* (i.e., "extremely vivid, long-lasting memories for unexpected, emotionally laden, and consequential events", Talarico & Rubin, 2007, p. 455).

Given the high frequency of psychological aggression reported compared to the other categories of IPA, the observed lower levels of confidence in the accuracy of psychological aggression reports supports past findings demonstrating that metamemory judgements regarding the accuracy of frequency estimates are more likely to be erroneous as the frequency of the event increases (e.g., Thompson & Mingay, 1991). Conversely, the significant associations observed between confidence and consistency for reports of sexual coercion, the least common IPA category, were well-calibrated (i.e., confidence increased as inconsistencies decreased). Again, earlier literature focused on a single event, not a class of events, precluding direct comparisons.

A minority of studies have compared reports of men to women, with many exclusively sampling male (e.g., Krinsley et al., 2003; Southwick et al., 1997) or female (e.g., Bohanek et al., 2005; Dill et al., 1991; Peace, Porter, & ten Brinke, 2008) participants, or simply not examining report consistency as a function of gender (e.g., Berntsen et al., 2000; Geraerts et al., 2007; Herlihy et al., 2002; Porter & Peace, 2007). Comparisons of reports as a function of participant gender in the present study produced few significant results. This finding supports the results of Spinhoven et al. (2006) who found no gender differences in memory consistency for adolescents' reports of the prevalence of traumatic life events. In contrast, Ghetti et al. (2002) identified gender

differences in the consistency of children's reports of abuse: Girls were more consistent than boys in their reports of the prevalence sexual abuse. Ghetti et al. speculated that male victims might be less comfortable than female victims providing information regarding IPA experiences, resulting in decreased consistency scores (where scores of 0 reflected no inconsistent information; 1 reflected inconsistent, but not contradictory information; 2 reflected inconsistent and contradictory information).[12] Thus, differences between the results of the present study and that of Ghetti et al. may reflect the increased anonymity afforded by the online data collection procedure of the present study compared with the interview format used by Ghetti et al. Again, Ghetti et al.'s sample of children discussing child abuse experiences is very different from the present study's sample of adults discussing IPA; not only is the focus of the questioning different, but there exist significant differences in the cognitive abilities of children and adults that may affect reporting (Saywitz & Camparo, 1998).

Goodman et al. (1999) and Mollica et al. (2007) also identified gender differences in consistency of adults' reports of victimization experiences over time. Specifically, Goodman et al. interviewed 50 men and women with a serious mental illness (e.g., schizophrenia, bipolar disorder) about adult physical and sexual abuse, as well as childhood sexual abuse. Findings demonstrated that reports of adult sexual abuse made by men were significantly less consistent than those made by women. Across participants, the direction of inconsistency in event prevalence also differed: Whereas men reported fewer experiences of sexual abuse over time, women tended to show increases in the prevalence reported. In contrast, Mollica et al.'s examination of

[12] Direction of inconsistencies not reported.

consistency in reports of wartime traumatic experiences mentioned earlier showed that women demonstrated significant decreases in the number of traumatic events reported over time whereas men's reports evidenced no significant change. Regardless of the direction observed, these differences contradict the present study's finding that men and women did not differ significantly in their consistency of the frequency or occurrence of IPA victimization experiences reported over time. Again, such discrepancies may reflect the methods of data collection used and the samples surveyed (e.g., general population vs. persons suffering from severe mental illness or wartime refugees).

The gender differences identified by both Ghetti et al. and Goodman et al., however, were within the same category of IPA for which the present study revealed a significant Gender by Direction of Inconsistency effect on mean confidence ratings. Specifically, recall that even after adjusting for frequency, men who increased disclosure of sexually coercive experiences over time were less confident than men who decreased disclosure. Women's confidence in the accuracy of their reports of sexual coercion, in contrast, decreased as the likelihood of decreased disclosure increased. Thus, it is possible that the newly disclosed events were accompanied by lower confidence or that the decreased confidence is reflective of a reluctance and increased discomfort in discussing such experiences. Research, for example, has identified significant stigma associated with experiences of sexual abuse among male victims (e.g., Lisak, 1994).

*Limitations*

There are limitations to the present study. IPA sample participants were not asked to identify whether they perceived the IPA victimization to be 'traumatic' memory events. The level of 'trauma' associated with these events may differ significantly across

participants. The same may be true for the level of emotion associated with the reported LA participation; that is, although non-traumatic for all participants, it is possible that LA participation actually reflects very positive memory events for some and neutral memory events for others. Informal feedback from a representative of a domestic violence shelter where the study was advertised suggested that individuals who have experienced severe IPA in the last 12 months may be unlikely to come forward to disclose their experiences. As such, the followback timeframe may have prevented individuals experiencing more severe IPA from volunteering insofar that they were not yet ready to openly discuss their experiences. Those who were willing to discuss the IPA victimization and volunteered to participate in the present study may represent those individuals whose IPA victimization would fall on the lower end of the severity continuum and, thus, related memory events may be lower in level of 'trauma'. Additionally, individual differences of potential relevance to the traumatic memory debate (e.g., prevalence of psychological sequelae resulting from the IPA victimization, autobiographical memory capacity) were not included due to time constraints.

Gender comparisons are limited insofar that men comprised a considerably smaller percentage of participants compared with women (approximately 20% vs. 80%). However, this ratio is quite consistent with that observed in official reports of domestic violence. For instance, in a study of domestic violence protective orders filed in the Sacramento Family Court between 2002 and 2003, Muller, Desmarais, and Hamel (under review) found that 80% were filed by female plaintiffs against their male intimate partners and 20% were filed by male plaintiffs against their female intimate partners. Brown (2004) found that female victims were four times more likely to report partner

violence to the police than were male victims (81% vs. 19%). In 1999, women comprised between 35% of domestic violence arrests in Concord, New Hampshire and 23% in Vermont (Goldberg, 1999). Of domestic violence offenders in Connecticut in 1999, 20% were women (Connecticut Department of Public Safety, 1999). As a last example, among arrests for domestic violence in California in 1998, approximately 17% of the cases were identified as female-perpetrated (State of California, 1999). Thus, it would appear that the gender ratio observed in the present study is, in fact, very representative of cases seen in the criminal justice system.

Perhaps the most significant limitation, the reasons for the inconsistencies observed in the present study could not be explored. Whereas the decreased disclosure over time may represent forgetting of autobiographical memory events, it is also possible that this pattern of results reflects participants' decreased motivation over time. Indeed, a distinction between report consistency and memory consistency over time is not possible. Although report consistency is largely a function of memory accuracy and consistency, there are several other factors which may be involved. In the context of IPA victimization, for example, failure to follow through with initial reports, and decreased disclosure of the allegations in particular, may arise not due to memory deficits, but for many other reasons, such as pressure from family or promises to change made by the abusive partner (see Scheppele, 1992; Stanko, 1982). Further, whereas the decreased disclosure over time may represent forgetting of autobiographical memory events, it is also possible that this pattern of results reflects participants' decreased motivation over time. Consider, for example, the data patterns for consistency for Psychological Aggression and Walking and Related Activities. As well as being the most prevalent

events, they also were the categories for which participants demonstrated the most inconsistencies. As suggested earlier, such findings may reflect increased difficulty in recalling higher frequency events or an increased range for detecting change with more frequent event. Alternatively, even though instructions emphasized the importance of complete recall, participants' motivation to fill out the calendar portion of the questionnaire may have decreased with the increased frequency. Finally, the reports elicited in the present study were retrospective in nature and no corroborating information was obtained. Although calendar-based methods reduce reporting error, the validity of reports cannot be ensured due to reconstructive nature of autobiographical memory. Thus, it is not possible to distinguish between report and memory consistency.

*Conclusions and Future Directions*

This study contributes to the literature on report or testimonial consistency and the traumatic memory debate, suggesting that consistency of event prevalence reports may vary categorically, as a function of frequency, rather than as a function of the nature of the autobiographical memory (i.e., memories of abusive experiences vs. everyday events). Comparing consistency of IPA victimization reports to consistency of LA participation reports, I asked whether it is realistic to expect victims of IPA to consistently report the occurrence and frequency of victimization over time and across repeated questioning. Though we generally recognize the fallibility of our own memory for autobiographical events, there is an expectation that the same is not true for a complainant's memory for victimization experiences and that, as such, inconsistencies are indicative of inaccuracies or falsehoods. However, "these errors in memory [and inconsistencies in reporting]... are especially likely to occur for the kinds of events that

are reported in courtroom testimony" (Haber & Haber, 2000, p. 1057). As evidenced by the pervasiveness of inconsistencies in reports of the occurrence and frequency of both abusive and everyday autobiographical memory events, absolute consistency of reports over time is simply not normative, suggesting that interpretation of inconsistency as indicating low credibility is not necessarily warranted.

The strengths of the study lie in the relatively limited and controlled variation in the delay between questionings, time elapsed since the abusive relationship (12 months or less), and timeframe in which the behaviours of interest could occur (12 month period preceding T1). Such design features represent improvements upon past research examining consistency of reports of abusive experiences and traumatic memory events (e.g., Desmarais et al., 2006; Porter & Peace, 2007). Given the dearth of research comparing consistency of traumatic and everyday memory events (*cf.* Brewin, 2007), and with respect to event prevalence in particular, an additional strength of this study is its comparison of the consistency of reports of the occurrence and frequency of abusive experiences and everyday autobiographical events *within* a victim sample and *between* victim and matched comparison samples. The prevalence of inconsistencies in the direction of decreased disclosure over increased disclosure in the present study, as reported by Desmarais et al. (2006) as well, is of particular significance and concern for the prosecution of IPA cases. Whether this pattern of results reflects memory processes associated with varying types of autobiographical memory events or, at least in part, was influenced by the task requirements is unclear and there is a need for continued research.

Building upon the present design, future research examining quantitative and qualitative differences in consistency between categories of autobiographical memory

events could contribute to our understanding of traumatic memory by comparing, for example, the valence of categories of reported events (e.g., eliciting ratings of level of trauma or emotional impact) and the psychological sequelae associated with specific categories of experiences. Increased disclosure of physical assault experiences may be associated with decreased psychological symptomatology, such as PTSD and depression. For example, Foa, Molnar, and Cashman (1995) found that as rape narratives increased in length over time and with repeated questioning, scores on psychopathology measures improved. In contrast, other research (e.g., Hertel, 2004; Krinsley et al., 2003; Southwick et al., 1997) demonstrates that increased disclosure may be associated with worsening symptoms (ruminations and intrusions in particular). Additionally, the relationship between report consistency and memory consistency merits further investigation, as interpretation of the present results are limited by the inability to make such a distinction. Studies comparing the consistency of reports of traumatic and non-traumatic *true* autobiographical memory events with mistaken or false reports would contribute to the traumatic memory debate and further elucidate the relationship between report consistency and memory accuracy.

A distinctive feature of this study was its use of online calendar-based methods to examine the consistency of reports of autobiographical memory events. Traditionally administered in interview format, the TLFB-SV was expanded in the present study to capture the CTS2 categories of psychological aggression and sexual coercion, in addition to physical assault, and was modified to allow for online data collection, the first such adaptation to my knowledge. In general, informal feedback from participants during debriefing endorsed the use of web-based materials, noting an increased comfort in

answering sensitive questions due to the anonymity afforded by this methodology as well as the convenience of being able to complete the surveys on their own time, from a location of their choice. Other comments, however, suggested that completing the online surveys in two separate internet windows (i.e., the questionnaire in one window, the calendar in the other) was somewhat cumbersome and the calendar itself tedious. Although participants included in analyses successfully submitted both the survey and calendar at each time point, not all respondents were equally successful. Specifically, 32 respondents completed the study but did not successfully submit both components at each time point and thus were excluded from the final study sample. The development of technology that enabled the automatic submission of the completed survey and completed calendar would facilitate data collection and enhance the integrity of results.

Collecting data through online methods raises questions regarding the characteristics of individuals most likely to volunteer for participation. A recent Canadian survey suggests that the likelihood of using the internet for personal reasons is higher for women compared to men, for someone residing in an urban area compared with someone from a rural or small town, and for someone with at least some post-secondary education compared with no post-secondary (see McKeown, Noce, & Czerny, 2007). Results of this survey also indicated that internet use decreased as age increased and household income decreased. Indeed, these odds are reflected in the present study's sample of predominantly well-educated women in their twenties. Future research could address these sample limitations and improve upon the present design by using stratified sampling, for example stratifying on race, age, gender, income, and education.

# STUDY 2: PERCEIVED CREDIBILITY OF 'COMPLAINANTS' REPORTING ON ABUSIVE EXPERIENCES AND EVERYDAY EVENTS

Having established comparable levels of consistency for reports of abusive experiences and everyday autobiographical event prevalence in Study 1, the second study of this dissertation examined how inconsistencies affect evaluations of complainant credibility and whether these perceptions are differentially affected by inconsistencies as a function of report content.

## *Introduction*

### *Background*

As reviewed earlier, prior research examining the effects of inconsistencies on mock jurors' perceptions of witness credibility, decisions regarding defendant guilt, and trial verdict have produced contrasting results. Even when associations are observed, consistency does not fully or reliably explain the variability in judgments. For example, although Leippe and Romanczyk (1989) found that inconsistency detrimentally affected evaluations of child witness testimony, no effects of inconsistency were found on evaluations of adult witness testimony. Other factors must be involved.

Jury research has demonstrated biases against outgroups in the courtroom (e.g., Brewer & Hupfeld, 2004; Daudistel, Hosch, Holmes, & Graves, 1999; Sommers & Ellsworth, 2000; Taylor & Hosch, 2004), therefore social categorization may be another factor that affects evaluations of perceived credibility or witness effectiveness. Based on

*social identity theory,* which posits individuals define themselves with respect to socially meaningful categories, *self-categorization theory* arguably offers the most comprehensive explanation of how we categorize people to maximize information and minimize cognitive effort (see Tajfel & Turner, 2001, 2004). This theory suggests that there is a human tendency to identify groups to which we belong (i.e., 'us' or the ingroup) and groups to which we do not belong (i.e., 'them' or the outgroup). Individuals considered to be more 'like' ourselves, falling into the 'us' category, generally are evaluated more favourably, whereas those in the 'them' category generally receive more negative appraisals (Robbins & Krueger, 2005). Thus, in the jury context, we may be more likely to identify or sympathize with complainants (and defendants) who we perceive to be more similar to ourselves.

Socially meaningful categories through which such us-them evaluations are made often include age, race, and gender (*cf.* Macrae & Bodenhausen, 2000). With respect to the latter, research across a variety of domains supports the existence of a gender similarity bias in perceptual judgements (e.g., assessments of discrimination claims: Elkins, Phillips, & Konopaske, 2002; Elkins, Phillips, Konopaske, & Townsend, 2001; evaluations of victims of sexual assault allegations: Johnson, Jackson, Gatto, & Nowack, 1995; recommended punishment for a convicted murderer: McKelvie, 2002; attributions regarding responsibility for abuse: Worthen & Varnado-Sullivan, 2005). For reasons introduced earlier in this dissertation, gender (of the complainant and evaluator) may be a particularly meaningful social category for evaluations of IPA allegations (see Hamel et al., 2007; Locke & Richman, 1999). Moreover, with respect to evaluations of IPA

complainants, victim versus non-victim may be another socially meaningful dimension through which such categorizations are made.

However, social categorization likely does not fully explain evaluations either. Prior research suggests that such extra-legal factors may only influence decision-making under certain circumstances. Brewer and Hupfeld's (2004) examination of the interaction of testimonial consistency and witness group identity demonstrates the importance of strength or amount of evidence, for example. Specifically, as suggested by the *heuristic-systematic processing theory* (Chaiken, Liberman, & Eagly, 1989; Chen & Chaiken, 1999; Chen, Duckworth, & Chaiken, 1999), when evidence is strong (i.e., when there is considerable information for or against), the influence of witness group categorization is minimal (*cf.* Brewer & Hupfeld, 2004; see also Giner-Sorolla, Chaiken, & Lutz, 2002); that is, a detailed analysis of information, systematic in nature, is dominant because there is sufficient evidence to inform the decision and heuristics are not necessary. In contrast, heuristic processing may be dominant in decision-making regarding IPA allegations because the complainant's report may be the only evidence available and the decision maker may make use of extra-evidential information, such as complainant gender, to come to a resolution.

*Study 2*

Extending the work of Brewer and Hupfeld (2004), Study 2 of this dissertation compared the perceived effectiveness[13] of persons reporting on IPA victimization to those of persons reporting on LA participation varying the consistency of the report, as

[13] The term *effectiveness* rather than *credibility* is used throughout the remainder of Study 2 to avoid confusion between the overall construct of credibility and the actual credibility scale that was presented to participants.

well as complainant and participant gender. Exploring the relevance of social categorization, my goal was to examine whether community respondents discriminate against persons alleging IPA victimization and are differentially judgmental of consistency as a function of the content of the report. Specifically, are inconsistencies more detrimental to evaluations of IPA complainants? If so, is this discrimination a function of social categorization; that is, does social categorization moderate or mediate the effects of report content, consistency, and gender on perceptions of complainant effectiveness?

*Hypotheses*

Overall, in accordance with the notion that evaluators use consistency as a proxy for accuracy when objective verification of facts is not possible, a significant main effect of report consistency was expected across domains, such that participants would judge complainants[14] who give consistent reports as more effective than those who give inconsistent reports. Participant and complainant gender also were expected to affect evaluations, especially for reports of IPA victimization. It was anticipated that female IPA complainants would be evaluated more favourably than male complainants (for reasons related to gender congruency) and that female participants generally would assign more positive evaluations than would male participants. With respect to the latter prediction, some research suggests that women generally demonstrate more empathy for or are more favourable towards victims of abuse or assault (e.g., Bottoms, Golding, Stevenson, Wiley, & Yozwiak, 2006; Golding, Sego, Sanchez, & Hasemann, 1995;

[14] Although the terms *interviewee* or *target* may be more appropriate for LA conditions, *complainant* is used throughout this dissertation for the sake of simplicity.

Schuller & Hastings, 1996). However, male complainants may be evaluated more favourably than female complainants in LA conditions, if women are perceived as less effective communicators.

Biases against outgroups were expected with respect to both victim (i.e., victim vs. non-victim, varied as a function of report content) and gender (i.e., male vs. female) categories, such that the greatest ratings of similarity between the participant and complainant would be observed for conditions in which a same-gender complainant (i.e., female participant and female complainant; male participant and male complainant) is reporting on LA participation, and the least similarity would be observed for cases in which an other-gender complainant (i.e., female participant and male complainant; male participant and female complainant) is reporting on IPA victimization. Other-gender LA complainants were expected to receive greater ratings of similarity than same-gender IPA complainants.

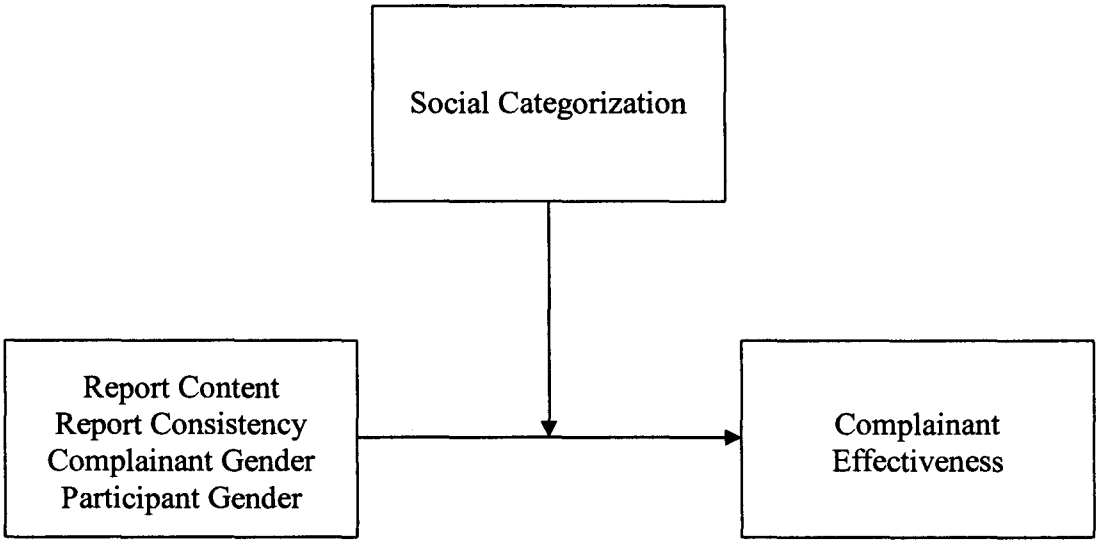
If direct associations between report characteristics (i.e., consistency and content), as well as complainant and participant gender, are established, I will examine whether social categorization moderates or mediates these associations. Whereas a moderator hypothesis addresses the question of *when* a variable predicts an outcome, a mediator hypothesis addresses the questions of *how* or *why* a variable predicts an outcome, describing a causal network (Baron & Kenny, 1986; Frazier, Tix, & Barron, 2004; Kraemer, Kiernan, Essex, & Kupfer, 2008; Kraemer, Stice, Kazdin, Offord, & Kupfer, 2001; Kraemer, Wilson, Fairburn, & Agras, 2002). Some research has conceptualized social categorization as a moderator of decision-making, driving the extent to which the decision-relevant evidence is evaluated as compared to decision-irrelevant information,

such as heuristics. As described by Brewer and Hupfeld (2004), for example, the extent to which inconsistencies affect witness effectiveness may be determined by prejudices resulting from social identity; that is, evidential information, such as report consistency, may be the dominant influence on juror decision-making for ingroup members, whereas extra-evidential information may be the dominant influence for evaluations of outgroup members.

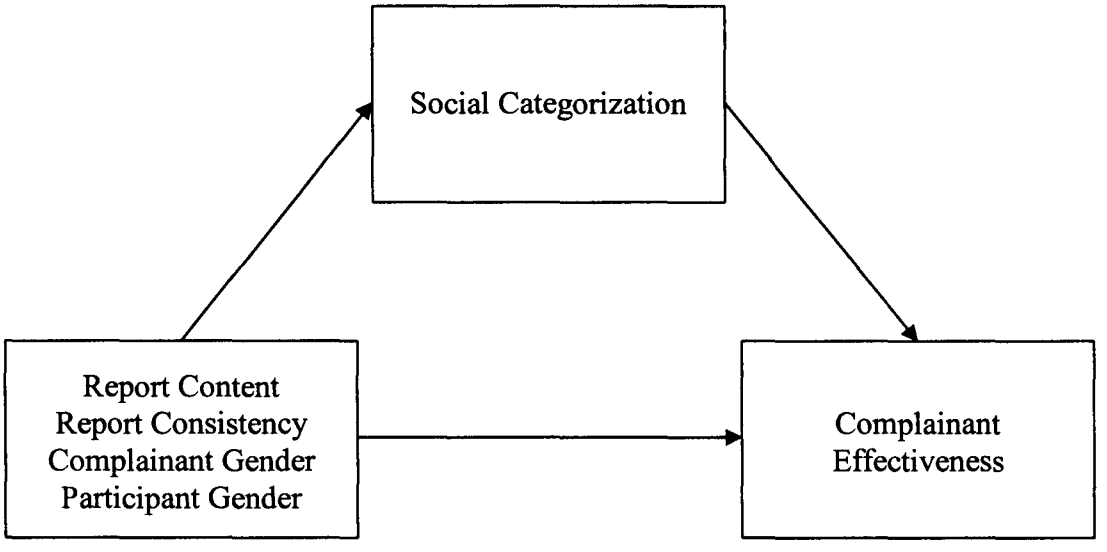
Alternatively, other research has treated social categorization as a mediator of bias in intergroup decision-making. For example, the categorization-elaboration model (CEM) proposed by van Knippenberg, De Dreu, and Homan (2004) holds that "intergroup biases flowing from social categorization disrupt the elaboration (in-depth processing) of task-relevant information and perspectives" (p. 1008). Thus, in the present study, social categorization may fulfill a mediation role, such that report consistency and content, and participant and complainant gender influence evaluations of complainant effectiveness, but only indirectly, by biasing perceptions of similarity and group categorization. In other words, an intermediary judgement of social categorization may play a mediating role in evaluations of complainant effectiveness. The potential moderating and mediating roles of social categorization are depicted in Figure 2.1 and will be compared in the present study.

***Figure 2.1. Potential Moderating and Mediating Roles of Social Categorization in Evaluations of Complainant Effectiveness***

*Social Categorization as a Moderator*




*Social Categorization as a Mediator*



## *Method*

### *Participants*

Participants were 401 (174 male, 226 female, 1 gender not disclosed) community members ranging in age from 19 to 88 years ($M = 37.37$, $SD = 15.96$) recruited from tables at two shopping centres in the Lower Mainland of British Columbia. All participants were jury-eligible (see sections 3(1) and (2) of the Jury Act & Section 4 of the Jury Regulations). Most participants (91%) had at least completed their high school education and 41% had attended a post-secondary institution. The majority (70%) indicated that English was their first language. Of the remaining participants, the average number of years they had spoken English was 17.79 ($SD = 9.31$). Participants were primarily of Caucasian (56%) or Asian (34%) descent. Comparisons with Canadian census data revealed that the sample closely approximated the population demographics on average age, gender representation, and education.[15]

### *Design*

This study was a 2 (Consistency: Consistent, Inconsistent) x 2 (Report Content: IPA, LA) x 2 (Complainant Gender: Male, Female) x 2 (Participant Gender: Male, Female) between-subjects design, for a total of 16 study conditions.

[15] Based on the results of the 2006 census, Statistics Canada reports that the median age in British Columbia is 40.8 years; men represent 49% of the population and women 51%; and English is the first language for 71% of British Columbians. Although 2006 education data are not yet released, results of the 2001 census demonstrated that 81% of British Columbians had at least completed their high school education, with 43% having completed a college or university degree.

*Materials*

*Vignettes.* To increase the generalizability of results, two IPA vignettes (i.e., rather than one) were created based on a review of Canadian domestic violence cases in which complainant credibility was at issue (see Appendix 2.1). In the vignettes, an adult describes during two interviews conducted 12 months apart an incident of IPA that occurred in the context of escalating conflict. In one vignette a complainant reports being hit by his or her intimate partner following a party, and in the other vignette a complainant reports having a coffee cup thrown at his or her face during an argument about their finances. For each vignette, complainant gender and report consistency were manipulated. Gender of the intimate partner was varied with complainant gender, such that all reports described heterosexual relationships.

Following the methodologies of Brewer and Hupfeld (2004) whose manipulations in turn draw on those of Berman and Cutler (1996), report consistency was operationalized as *consistent* (same information reported during both interviews) and *inconsistent* (four contradictions in information reported across interviews and one novel piece of information introduced during the second interview). Across inconsistent vignettes the following four pieces of information were inconsistent: the date of the incident (Vignette 1 – Saturday August 2$^{nd}$ vs. Saturday August 9$^{th}$; Vignette 2 – Sunday November 6$^{th}$ vs. Sunday November 13$^{th}$), location (Vignette 1 – front entrance vs. kitchen; Vignette 2 – kitchen vs. dining room), time of day (Vignette 1 – late evening vs. late afternoon; Vignette 2 – afternoon vs. evening), and context (Vignette 1 – alcohol involved, yes vs. no; Vignette 2 – history of violence, yes vs. no ). For Vignette 1, the novel piece of information involved the complainant reporting hearing a loud noise outside during the incident, and for Vignette 2 the complainant reported that the phone

rang during the incident. Criminal charges require a level of specificity as to time and place of an alleged criminal event; thus, some of the variations herein reflect changes in legally significant details.

As for the IPA vignettes, two LA vignettes were created based on a review of behaviours commonly reported by community respondents in the first study of this dissertation (see Appendix 2.2). Specifically, in Vignette 1, a complainant describes participating in a softball game, and in Vignette 2, a complainant describes going to the gym. As above, complainant gender and consistency were manipulated for each vignette. Again, across inconsistent vignettes the following four pieces of information were inconsistent: the date of the incident (Vignette 1 – Saturday August 2nd vs. Saturday August 9th; Vignette 2 – Sunday November 6th vs. Sunday November 13th), location (Vignette 1 – school vs. park; Vignette 2 – rec centre vs. gym), time of day (Vignette 1 – late evening vs. late afternoon; Vignette 2 – afternoon vs. evening), and context (Vignette 1 – pitcher was tired, yes vs. no; Vignette 2 – exercising regularly, yes vs. no ). For Vignette 1, the novel piece of information involved the complainant reporting that the pitcher fumbled with the ball, and for Vignette 2, the complainant reported that a cell phone rang.

Confidence level was held constant across all conditions, with the complainant stating that they were "pretty confident" in the accuracy of his or her report at both interview times, to control for its possible effects on evaluations of complainant effectiveness. Similarly, vignettes were constructed such that severity of the behaviours described in IPA conditions and intensity of the activities described in the LA conditions were relatively constant. Vignettes also were constructed such that word length was

approximately constant ($M_{words}$ = 706.75; $SD$ = 27.55; Range = 649 - 746). Finally, interviewer questions were identical across vignettes.

*Complainant Effectiveness Questionnaire.* Participants were asked to rate the complainant across seven 6-point Likert-type scales: intelligence, honesty, accuracy, suggestibility, confidence, likeability, and credibility. These scales were drawn from a 13-item credibility questionnaire used by Connolly and colleagues (2008). However, whereas Connolly et al. included three items assessing honesty and two items assessing each of accuracy and credibility, only single items were included for each construct in the present study. With the exception of the suggestibility scale, higher ratings reflect more positive judgments (i.e., 1 = Low; 6 = High); thus, suggestibility ratings were reverse coded for the purpose of analyses.

To measure social categorization, two additional items asked participants to rate how similar the complainant was to them and how closely they saw themselves as belonging to the same group as the complainant, again on 6-point scales (1 = Not at all similar/close; 6 = Very similar/close). A 6-point consistency scale also was included to test the success of the Report Consistency manipulation, as was a scale assessing severity of IPA behaviours or intensity of LA activities to ensure that this had indeed been held constant across vignettes and to establish whether this had any bearing on participants' evaluations. The Complainant Effectiveness Questionnaire is presented in Appendix 2.3.

*Procedure*

Once informed consent was obtained, participants were instructed that they would read a report consisting of two interviews with an adult (a man or woman depending on the Complainant Gender condition) who was interviewed twice about events or

experiences that may or may not have happened in the 12 months preceding the first interview. They were further instructed that the two interviews occurred approximately six weeks apart and that the report was randomly selected from a group of reports among which some of the adults were very consistent in their reports across the two interviews and some were not very consistent. This instruction was included to increase the saliency of the consistency manipulation and to parallel what an attorney might instruct jurors in a courtroom. Finally, participants were informed that they would be asked a series of questions about this person and specific details included in the report.

Each participant read one vignette, completed the questionnaire, and received a $5 gift certificate for the shopping centre in which data collection was conducted.

## *Results*

### *Manipulation Checks*

To determine whether participants had read and understood the report, a recall task of details presented in the vignette was included. Four 4-option multiple-choice questions assessed participants' recall of the month, date, and time the complainant reported that the incident or event occurred, as well as name of the other individual described in the report. These details were reported twice (i.e., in both interviews) across vignettes. In inconsistent study conditions (i.e., in which the complainant may have provided contradictory information for the same detail, such as event date), participant responses were identified as correct if they accurately selected at least one of the details

presented in the vignette per question.[16] On average, participants answered 3.68 questions correctly ($SD = 0.82$). Those with fewer than three of four correct responses were omitted ($n = 27$), for a final sample size of 374 participants.

To determine the effectiveness of the Report Consistency manipulation, participants' ratings of report consistency were analysed using a one-way ANOVA. Results demonstrated a significant main effect of Report Consistency, $F(1, 370) = 202.30$, $p < .001$, $\eta_p^2 = .35$, with ratings of consistency higher in the consistent ($M = 4.54$, $SD = 1.25$) than inconsistent ($M = 2.76$, $SD = 1.16$) condition. ANOVAs were conducted to examine effects, if any, of vignette version and data collection location on evaluations of complainant effectiveness. No significant effects were observed. ANOVAs also were conducted to examine whether participant ratings of IPA severity and LA intensity differed with vignette version. Again, no significant effects were observed.

In sum, analyses established the success of the Report Consistency manipulation and demonstrated that vignette version, data collection location, and severity/intensity ratings were not systematically associated with responding, affording the examination of the effects of Report Consistency, Report Content, Complainant Gender, and Participant Gender on effectiveness evaluations as independent from these possible procedural effects.

[16] Originally these recall items also were intended to serve as a consistency manipulation check. That is, participants could identify inconsistency in the report by circling multiple options per recall question. However, despite instructions indicating that participants should circle all the correct response(s), the observed low frequency of multiple selections per recall question suggests that participants assumed they could only select one multiple-choice option. Nonetheless, their answers still indicate whether they read and remembered the vignette details.

*Evaluations of Complainant Effectiveness*

Before conducting hypothesis tests, the structural reliability of the Complainant Effectiveness Questionnaire was examined. Readers are reminded that the questionnaire comprised the following seven items: intelligence, honesty, accuracy, suggestibility, confidence, likeability, and credibility. Means and standard deviations for effectiveness evaluations are presented in Table 2.1. Participant ratings general fell within the upper half of the scales, reflecting positive evaluations. For all scales, ratings ranged from 1.00 to 6.00. Analyses yielded a coefficient alpha estimate of .82, suggesting that internal consistency was good. Item homogeneity was measured using the mean inter-item correlation (MIC). The MIC (.40) fell within the range generally recognized to reflect a unidimensional questionnaire (.20 - .50). All corrected item-total correlations were positive and within the acceptable range (see Ferketich, 1991; Kline, 1993; Nunnally & Bernstein, 1994).

Overall, analyses supported the structural reliability of the questionnaire; thus, the hypothesized relationships between Report Consistency, Report Content, Complainant Gender, Participant Gender, and perceptions of complainant effectiveness are explored in the following section.

***Table 2.1. Descriptive Statistics for Ratings of Intelligence, Honesty, Accuracy, Suggestibility, Confidence, Likeability, and Credibility Overall and as a Function of Report Consistency, Content, Complainant Gender, and Participant Gender***

| Ratings | Overall | Independent Variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Report Consistency | | Report Content | | Complainant Gender | | Participant Gender | |
| | | Consistent | Inconsistent | IPA | LA | Male | Female | Male | Female |
| | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) |
| Intelligence | 3.78(1.00) | 4.14(0.88) | 3.45(1.00) | 3.66(0.98) | 3.91(1.01) | 3.73(0.95) | 3.84(1.05) | 3.79(1.06) | 3.78(0.97) |
| Honesty | 4.26(1.28) | 4.77(1.05) | 3.78(1.30) | 4.02(1.26) | 4.51(1.25) | 4.14(1.24) | 4.38(1.31) | 4.27(1.31) | 4.26(1.26) |
| Accuracy | 3.65(1.37) | 4.38(1.12) | 2.94(1.20) | 3.57(1.45) | 3.72(1.28) | 3.60(1.30) | 3.69(1.43) | 3.83(1.35) | 3.50(1.36) |
| Suggestibility | 3.27(1.43) | 3.76(1.30) | 2.82(1.39) | 3.25(1.44) | 3.30(1.42) | 3.24(1.33) | 3.31(1.52) | 3.41(1.43) | 3.17(1.42) |
| Confidence | 4.16(1.39) | 4.70(1.15) | 3.64(1.40) | 4.03(1.45) | 4.29(1.32) | 4.12(1.42) | 4.19(1.36) | 4.13(1.35) | 4.17(1.42) |
| Likeability | 3.92(1.14) | 4.17(1.09) | 3.69(1.13) | 3.64(1.10) | 4.21(1.10) | 3.85(1.06) | 3.98(1.20) | 3.89(1.09) | 3.93(1.17) |
| Credibility | 3.84(1.33) | 4.48(1.12) | 3.23(1.23) | 3.67(1.33) | 4.01(1.32) | 3.71(1.33) | 3.96(1.33) | 4.01(1.28) | 3.70(1.36) |

*Notes*. *n*'s = 361 - 374. Ratings were made on 6-point scales (1 = Low; 6 = High). A higher score reflects higher ratings.

Comparing mean ratings as a function of Report Consistency it was evident that complainants who gave consistent reports were rated more positively than those who gave inconsistent reports. Ratings for complainants reporting on LA participation were more positive than for complainants reporting on IPA victimization. Differences between ratings of female and male complainants were mixed but generally in the expected direction: Female complainants were rated more positively on intelligence, honesty, accuracy, suggestibility, confidence, and credibility. However, these differences generally were quite small and non-significant. Finally, with the exceptions of confidence and likeability, male participants generally gave more positive ratings than female participants, in contrast with expectations. As with the Complainant Gender comparisons, these differences often were very small.

Correlations between participant ratings on the effectiveness scales can be found in Table 2.2. With the exception of the association between likeability and suggestibility, all correlations were significant ($p$'s $< .001$) and in the expected direction. Although not presented in Table 2.2, the largest correlation observed was between participant ratings of consistency and accuracy ($r = .72$, $p < .001$).

***Table 2.2. Correlations between Participants' Ratings of Intelligence, Honesty, Accuracy, Suggestibility, Confidence, Likeability, and Credibility***

| Ratings | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1. Intelligence | -- | | | | | | |
| 2. Honesty | .54*** | -- | | | | | |
| 3. Accuracy | .52*** | .59*** | -- | | | | |
| 4. Suggestibility | .16*** | .14*** | .26*** | -- | | | |
| 5. Confidence | .39*** | .39*** | .54*** | -.20*** | -- | | |
| 6. Likeability | .52*** | .50*** | .35*** | -.01 | .36*** | -- | |
| 7. Credibility | .53*** | .60*** | .64*** | -.23*** | .56*** | .50*** | -- |

*Notes.* *n*'s = 365 - 374. ***$p < .001$. Ratings were made on 6-point scales (1 = Low; 6 = High). A higher score reflects higher ratings.

A multivariate analysis of variance (MANOVA) was conducted to examine the effects of Report Consistency, Report Content, Complainant Gender, and Participant Gender on evaluations of complainant effectiveness. Results are presented in Table 2.3. The analysis demonstrated significant effects of Report Consistency ($\eta_p^2 = .36$), Report Content ($\eta_p^2 = .08$), and Participant Gender ($\eta_p^2 = .05$), but not of Complainant Gender, on effectiveness evaluations. Univariate ANOVAs revealed that consistent complainants were rated significantly more positively than inconsistent complainants across all seven effectiveness scales ($\eta_p^2$'s = .04 – .28). LA complainants were rated as significantly more intelligent, honest, likeable, and credible than IPA complainants ($\eta_p^2$'s = .02 – .06), whereas differences between accuracy, suggestibility, and confidence ratings did not reach significance ($p$'s > .05). Male participants gave significantly higher ratings than female participants on accuracy, and credibility ($\eta_p^2$'s = .01 – .02), but ratings of intelligence, honesty, suggestibility, confidence, and likeability did not differ significantly ($p$'s > .05).

***Table 2.3. Multivariate and Univariate Analyses of Variance for Effectiveness Evaluations***

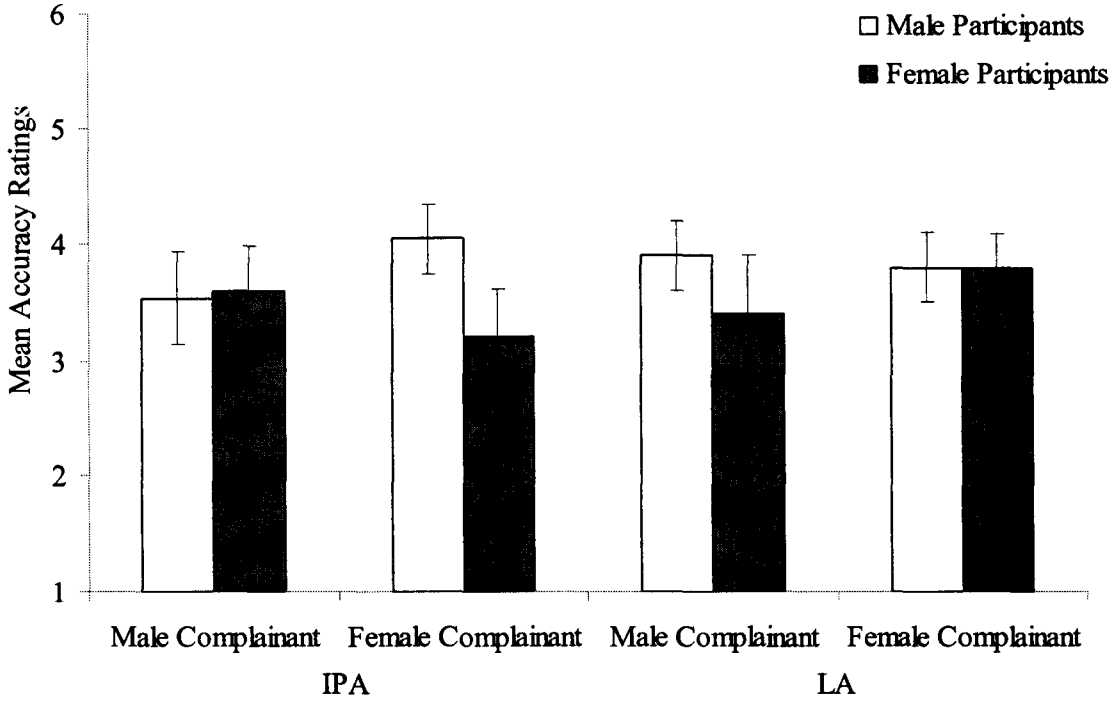
| | | Univariate | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | Multivariate $F(7, 339)$ | Intelligence $F(1, 345)$ | Honesty $F(1, 345)$ | Accuracy $F(1, 345)$ | Suggestibility $F(1, 345)$ | Confidence $F(1, 345)$ | Likeability $F(1, 345)$ | Credibility $F(1, 345)$ |
| Report Consistency (Cy) | 27.01*** | 50.11*** | 61.22*** | 136.20*** | 42.29*** | 57.52*** | 17.16*** | 100.14*** |
| Report Content (Ct) | 3.95*** | 6.09* | 14.09*** | 0.79 | 0.40 | 2.54 | 20.87*** | 6.61** |
| Complainant Gender (CG) | 0.82 | 1.48 | 3.71 | 0.60 | 0.20 | 0.54 | 0.72 | 3.77 |
| Participant Gender (PG) | 2.47* | 0.01 | 0.00 | 5.77* | 3.53 | 0.03 | 0.25 | 4.70* |
| Cy x Ct | 0.97 | 0.03 | 0.02 | 0.72 | 1.81 | 0.39 | 2.33 | 0.28 |
| Cy x CG | 0.70 | 1.45 | 0.55 | 0.00 | 1.79 | 1.25 | 1.27 | 0.39 |
| Cy x PG | 0.56 | 0.06 | 1.39 | 0.02 | 0.06 | 0.74 | 0.12 | 0.00 |
| Ct x CG | 1.35 | 0.34 | 0.89 | 0.50 | 1.47 | 0.30 | 0.75 | 7.15** |
| Ct x PG | 1.19 | 3.42 | 1.08 | 0.96 | 1.19 | 0.37 | 3.07 | 5.10* |
| CG x PG | 0.71 | 0.00 | 0.06 | 0.64 | 0.16 | 0.62 | 2.43 | 0.00 |
| Cy x Ct x CG | 0.54 | 0.03 | 0.19 | 0.27 | 0.18 | 2.15 | 0.35 | 0.03 |
| Cy x Ct x PG | 0.57 | 0.42 | 0.06 | 0.11 | 0.26 | 1.57 | 0.02 | 0.50 |
| Cy x CG x PG | 0.46 | 0.24 | 0.38 | 0.41 | 0.13 | 0.00 | 0.31 | 0.00 |
| Ct x CG x PG | 2.22* | 0.14 | 1.93 | 6.84** | 0.02 | 0.81 | 0.75 | 0.11 |
| Cy x Ct x CG x PG | 0.94 | 3.86* | 1.03 | 0.34 | 0.28 | 0.55 | 0.08 | 0.40 |

*Notes.* $n = 361$. Multivariate $F$ ratios are Wilks' Lambdas. *$p < .05$. **$p < .01$. ***$p < .001$.

The MANOVA also revealed a modest but significant Report Content x Complainant Gender x Participant Gender interaction ($\eta_p^2 = .04$). At the univariate level, the effect was significant for ratings of accuracy specifically ($\eta_p^2 = .02$). As depicted in Figure 2.2 below, *post hoc* comparisons revealed that, in contrast with hypotheses, male participants rated female IPA complainants as significantly more accurate ($M = 4.05$, $SD = 1.38$) than did female participants ($M = 3.21$, $SD = 1.55$), $t(93) = 2.76$, $p < .001$, $d = .57$, whereas accuracy ratings of male IPA complainants did not differ significantly ($p > .05$). Male and female participant ratings of male and female LA complainants did not differ significantly ($p$'s $> .05$).

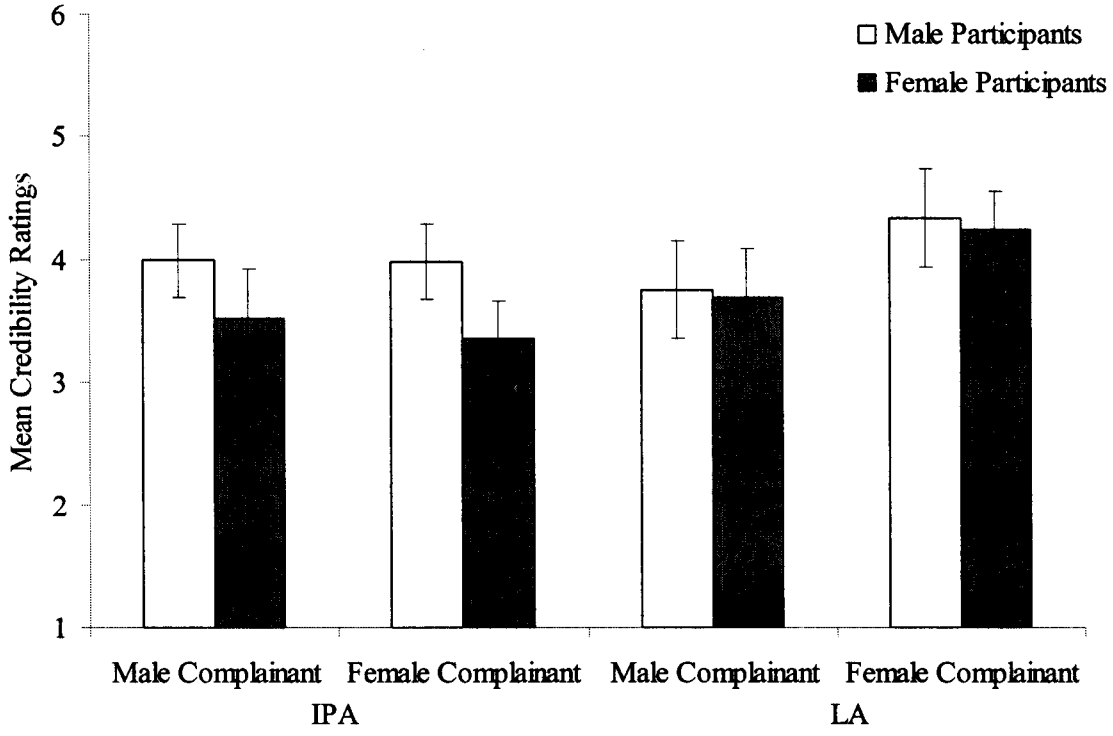
Although the interaction was significant only for ratings of accuracy, *post hoc* exploration of the other scales revealed the same pattern of responding for credibility. Figure 2.3 shows that male participants rated female IPA complainants ($M = 3.98$, $SD = 1.26$) as significantly more credible than did female participants ($M = 3.36$, $SD = 1.44$), $t(93) = 2.19$, $p < .05$, $d = .45$, whereas credibility ratings of male IPA complainants did not differ ($p > .05$). Again, male and female participant ratings of male and female LA complainants did not differ significantly.

***Figure 2.2. Report Content x Complainant Gender x Participant Gender Interaction Effect on Participant Ratings of Accuracy***




*Notes.* Bars represent 95% confidence intervals around the means. 1 = Not at all accurate; 6 = Very accurate.

***Figure 2.3. Male and Female Participants' Ratings of Credibility across Complainant Gender and Report Content***




*Notes.* Bars represent 95% confidence intervals around the means. 1 = Not at all credible; 6 = Very credible.

*Similarity and Group Categorization*

As described earlier, participants' ratings of how similar the complainant was to them and how closely they saw themselves as belonging to the same group as the complainant were used to assess social categorization. Means and standard deviations for these ratings are presented in Table 2.4. Overall, ratings fell around the scale midpoints. Comparing ratings as a function of Report Consistency, it was evident that participants rated complainants in consistent conditions as more similar and more likely to belong to the same group than complainants in inconsistent conditions. Ratings for complainants reporting on LA participation were higher than for complainants reporting on IPA victimization. Female complainants received slightly higher ratings than male complainants, but differences were very small. Male participants gave higher ratings than female participants. For both scales, participant ratings ranged from 1.00 to 6.00.

With the exception of associations with suggestibility, all correlations were significant ($p$'s < .001) and in the expected direction: More positive evaluations were associated with higher ratings of similarity and increased likelihood of belonging to the same group (see Table 2.5). A MANOVA demonstrated small but significant multivariate main effects of Report Consistency ($\eta_p^2$ = .05) and Report Content ($\eta_p^2$ = .06), but not of Participant Gender or Complainant Gender, on ratings of similarity and group categorization. Results of these analyses are presented in Table 2.6. Participants rated consistent complainants as significantly more similar and significantly more likely to belong to the same group as themselves than inconsistent complainants ($\eta_p^2$'s = .04), as was the case for LA complainants compared with IPA complainants ($\eta_p^2$'s = .02, .05; see Tables 2.4 and 2.6).

***Table 2.4. Descriptive Statistics for Ratings of Similarity and Group Categorization Overall and as a Function of Report Consistency, Report Content, Complainant Gender, and Participant Gender***

| Ratings | | Independent Variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Report Consistency | | Content | | Complainant Gender | | Participant Gender | |
| | Overall *M* (*SD*) | Consistent *M* (*SD*) | Inconsistent *M* (*SD*) | IPA *M* (*SD*) | LA *M* (*SD*) | Male *M* (*SD*) | Female *M* (*SD*) | Male *M* (*SD*) | Female *M* (*SD*) |
| Similarity | 3.04 (1.54) | 3.33 (1.57) | 2.77 (1.47) | 2.79 (1.51) | 3.29 (1.53) | 3.02 (1.52) | 3.05 (1.57) | 3.21 (1.53) | 2.90 (1.54) |
| Group Categorization | 3.06 (1.40) | 3.36 (1.32) | 2.79 (1.42) | 2.74 (1.34) | 3.39 (1.38) | 3.05 (1.39) | 3.08 (1.40) | 3.18 (1.39) | 2.97 (1.40) |

*Notes*. Similarity $n = 373$; Group Categorization $n = 372$. Ratings were made on 6-point scales (1 = Not at all similar/close; 6 = Very similar/close).

***Table 2.5. Correlations between Participants' Effectiveness Evaluations and Ratings of Similarity and Group Categorization***

| Ratings | Similarity | Group Categorization |
|---|---|---|
| Intelligence | .36*** | .41*** |
| Honesty | .37*** | .35*** |
| Accuracy | .31*** | .32*** |
| Suggestibility | .06 | .07 |
| Confidence | .30*** | .27*** |
| Likeability | .44*** | .49*** |
| Credibility | .38*** | .39*** |

*Notes.* $n$'s = 366-373. ***$p < .001$. Ratings were made on 6-point scales (1 = Low; 6 = High). A higher score reflects higher ratings.

***Table 2.6. Multivariate and Univariate Analyses of Variance for Similarity and Group Categorization Ratings***

| Variable | Multivariate | Univariate | |
|---|---|---|---|
| | | Similarity | Group Categorization |
| | $F(2, 353)$ | $F(1, 354)$ | $F(1, 354)$ |
| Report Consistency (Cy) | 9.26*** | 13.07*** | 14.97*** |
| Report Content (Ct) | 10.38*** | 8.43** | 20.36*** |
| Complainant Gender (CG) | 0.02 | 0.01 | 0.00 |
| Participant Gender (PG) | 1.76 | 3.13 | 2.13 |
| Cy x Ct | 0.32 | 0.41 | 0.01 |
| Cy x CG | 0.11 | 0.11 | 0.20 |
| Cy x PG | 0.67 | 0.17 | 1.29 |
| Ct x CG | 0.42 | 0.84 | 0.15 |
| Ct x PG | 0.18 | 0.26 | 0.29 |
| CG x PG | 0.56 | 1.11 | 0.23 |
| Cy x Ct x CG | 0.32 | 0.64 | 0.10 |
| Cy x Ct x PG | 0.52 | 0.14 | 1.02 |
| Cy x CG x PG | 2.20 | 1.62 | 0.61 |
| Ct x CG x PG | 0.56 | 1.12 | 0.28 |
| Cy x Ct x CG x PG | 0.07 | 0.12 | 0.06 |

*Notes.* $n = 370$. Multivariate $F$ ratios are Wilks' Lambdas. **$p < .01$. ***$p < .001$.

*Intimate Partner Abuse Severity and Leisure Activities Intensity*

Readers are reminded that participants rated the severity of IPA behaviours or intensity of LA activities reported to confirm whether severity and intensity had indeed been held constant across vignettes and to establish whether it had any bearing on participants' effectiveness evaluations. It is reasonable to expect, for example, that reports of more severe abuse may be more likely to be believed. Thus, analyses were replicated with severity and intensity ratings included as potential covariates.

Descriptive statistics for ratings of IPA severity in IPA conditions and ratings of LA intensity in LA conditions, overall and as a function of the independent variables (with the exception of Report Content[17]), are presented in Table 2.7. For both scales, participant ratings ranged from 1.00 to 6.00. Overall ratings of IPA severity were somewhat higher than ratings of LA intensity. With the exception of Report Consistency in LA conditions where inconsistent reports received significantly lower ratings than consistent reports, $t(182) = 2.34$, $p < .05$, $d = .35$, no significant differences in ratings of IPA severity and LA intensity were observed as a function of the independent variables ($p$'s > .05). Correlational analyses, presented in Table 2.8, demonstrated few associations between effectiveness evaluations and ratings of IPA severity and LA intensity. With the exception of the associations between ratings of IPA severity and credibility, correlations generally were small (.06 - .21).

[17] Analyses were conducted separately for IPA and LA conditions therefore Report Content could not be included as an independent variable.

***Table 2.7. Descriptive Statistics for Ratings of Intimate Partner Abuse Severity and Leisure Activities Intensity overall and as a Function of Report Consistency, Complainant Gender, and Participant Gender***

| | | Independent Variables | | | | | |
|---|---|---|---|---|---|---|---|
| | | Report Consistency | | Complainant Gender | | Participant Gender | |
| Ratings | Overall *M* (*SD*) | Consistent *M* (*SD*) | Inconsistent *M* (*SD*) | Male *M* (*SD*) | Female *M* (*SD*) | Male *M* (*SD*) | Female *M* (*SD*) |
| IPA Severity | 4.09 (1.16) | 4.20 (1.12) | 3.98 (1.20) | 4.00 (1.19) | 4.17 (1.14) | 4.21 (1.23) | 3.99 (1.10) |
| LA Intensity | 3.02 (1.12) | 3.24 (1.12) | 2.82 (1.27) | 3.07 (1.17) | 2.98 (1.26) | 3.05 (1.18) | 2.99 (1.24) |

*Notes*. IPA intensity $n = 188$; LA severity $n = 184$. Ratings were made on 6-point scales (1 = Not at all severe/intense; 6 = Very severe/intense).

***Table 2.8. Correlations between Participants' Effectiveness Evaluations, Similarity and Group Categorization Ratings, and Ratings of Intimate Partner Abuse Severity and Leisure Activities Intensity***

| Ratings | IPA Severity | LA Intensity |
|---|---|---|
| *Effectiveness Evaluations* | | |
| Intelligence | .15* | .13 |
| Honesty | .16* | .14 |
| Accuracy | .15* | .21* |
| Suggestibility | .10 | .05 |
| Confidence | .16* | .07 |
| Likeability | .06 | .11 |
| Credibility | .30*** | .18* |
| *Similarity and Group Categorization* | | |
| Similarity | -.07 | .14 |
| Group Categorization | .08 | .22** |

*Notes.* IPA conditions $n$ = 186-188; LA conditions $n$ = 180-184.*$p < .05$. **$p < .01$. ***$p < .001$. Ratings were made on 6-point scales (1 = Low; 6 = High).

Two MANCOVAs were conducted with effectiveness evaluations as the dependent variables: the first with IPA severity as the covariate for participants in IPA conditions, and the second with LA intensity as the covariate for participants in LA conditions. Results, presented in Table 2.9, generally fail to support the role of these ratings in effectiveness evaluations. No effects of LA intensity on effectiveness evaluations were found either the multivariate or univariate level ($p$'s > .05). A modest but significant multivariate effect of IPA severity was found ($\eta_p^2$ = .08), but was significant only for the credibility scale at the univariate level ($\eta_p^2$ = .07). For both IPA and LA conditions, results remained quite consistent with those previously reported. As may be seen in Table 2.9, with the exception of likeability ratings in IPA conditions, the multivariate and univariate main effects of Report Consistency remained significant ($\eta_p^2$'s = .07 - .37). Participant Gender was no longer found to affect effectiveness evaluations, nor was there a Complainant Gender x Participant Gender interaction ($p$'s > .05). However, subsequent MANOVAs revealed that these effects disappeared not as a result of controlling for IPA severity and LA ratings, but instead as a function of conducting analyses separately for IPA and LA conditions. As may be seen in Table 2.10, the multivariate and univariate main effects of Report Consistency are significant in both IPA and LA conditions. There is no effect of Participant Gender on effectiveness evaluations, nor is there a Complainant Gender x Participant Gender interaction ($p$'s > .05).

***Table 2.9. Multivariate and Univariate Analyses of Covariance for Effectiveness Evaluations Controlling for Severity and Intensity***

| | | Univariate | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | Multivariate $F(7, 168)$ | Intelligence $F(1, 174)$ | Honesty $F(1, 174)$ | Accuracy $F(1, 174)$ | Suggestibility $F(1, 174)$ | Confidence $F(1, 174)$ | Likeability $F(1, 174)$ | Credibility $F(1, 174)$ |
| *IPA Conditions* | | | | | | | | |
| IPA Severity | 2.09* | 1.54 | 1.49 | 1.42 | 0.64 | 2.38 | 0.29 | 12.96*** |
| Report Consistency (Cy) | 14.05*** | 27.39*** | 31.40*** | 70.87*** | 31.34*** | 31.88*** | 2.44 | 43.74*** |
| Complainant Gender (CG) | 0.37 | 0.11 | 0.33 | 0.01 | 0.40 | 0.00 | 0.00 | 0.80 |
| Participant Gender (PG) | 1.82 | 1.72 | 0.46 | 4.80* | 0.25 | 0.03 | 0.66 | 8.33** |
| Cy x CG | 0.73 | 0.59 | 0.05 | 0.14 | 0.43 | 3.32 | 1.41 | 0.33 |
| Cy x PG | 0.65 | 0.06 | 0.94 | 0.13 | 0.03 | 2.02 | 0.01 | 0.14 |
| CG x PG | 1.29 | 0.16 | 0.52 | 4.96* | 0.10 | 0.04 | 0.27 | 0.00 |
| Cy x CG x PG | 1.12 | 3.30 | 1.39 | 0.00 | 0.42 | 0.26 | 0.33 | 0.21 |
| *LA Conditions* | $F(7, 161)$ | $F(1, 167)$ | $F(1, 167)$ | $F(1, 167)$ | $F(1, 167)$ | $F(1, 167)$ | $F(1, 167)$ | $F(1, 167)$ |
| LA Intensity | 0.74 | 1.85 | 1.54 | 2.04 | 1.71 | 0.00 | 1.15 | 1.83 |
| Report Consistency (Cy) | 11.40*** | 19.32*** | 24.91*** | 54.72*** | 12.13*** | 21.75*** | 13.20*** | 50.04*** |
| Complainant Gender (CG) | 2.05 | 1.85 | 4.15* | 0.98 | 0.65 | 0.55 | 1.80 | 11.12*** |
| Participant Gender (PG) | 1.39 | 1.30 | 0.53 | 0.70 | 3.30 | 0.53 | 2.49 | 0.01 |
| Cy x CG | 0.38 | 0.65 | 0.65 | 0.05 | 0.86 | 0.01 | 0.09 | 0.08 |
| Cy x PG | 0.36 | 0.48 | 0.42 | 0.09 | 0.06 | 0.01 | 0.08 | 0.26 |
| CG x PG | 1.37 | 0.19 | 1.03 | 1.84 | 0.04 | 1.05 | 2.41 | 0.00 |
| Cy x CG x PG | 0.19 | 1.11 | 0.06 | 0.50 | 0.02 | 0.15 | 0.06 | 0.22 |

*Notes.* IPA conditions $n = 183$; LA conditions $n = 176$. Multivariate $F$ ratios are Wilks' Lambdas. *$p < .05$. **$p < .01$. ***$p < .001$.

***Table 2.10. Multivariate and Univariate Analyses of Variance for Effectiveness Evaluations in Intimate Partner Abuse and Leisure Activities Conditions***

| | | Univariate | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | Multivariate $F(7, 169)$ | Intelligence $F(1, 175)$ | Honesty $F(1, 175)$ | Accuracy $F(1, 175)$ | Suggestibility $F(1, 175)$ | Confidence $F(1, 175)$ | Likeability $F(1, 175)$ | Credibility $F(1, 175)$ |
| *IPA Conditions* | | | | | | | | |
| Report Consistency (Cy) | 14.25*** | 28.37*** | 33.45*** | 72.49*** | 32.20*** | 33.03*** | 2.58*** | 44.39*** |
| Complainant Gender (CG) | 0.28 | 0.21 | 0.49 | 0.00 | 0.31 | 0.02 | 0.00 | 0.26 |
| Participant Gender (PG) | 2.01 | 20.2 | 0.61 | 5.30* | 0.33 | 0.09 | 0.75 | 9.67** |
| Cy x CG | 0.73 | 0.59 | 0.05 | 0.14 | 0.43 | 3.28 | 1.41 | 0.30 |
| Cy x PG | 0.68 | 0.08 | 1.04 | 0.10 | 0.04 | 2.19 | 0.02 | 0.25 |
| CG x PG | 1.26 | 0.09 | 0.67 | 5.39* | 0.14 | 0.01 | 0.23 | 0.07 |
| Cy x CG x PG | 1.11 | 3.26 | 1.36 | 0.00 | 0.41 | 0.24 | 0.34 | 0.17 |
| *LA Conditions* | $F(7, 164)$ | $F(1, 170)$ | $F(1, 170)$ | $F(1, 170)$ | $F(1, 170)$ | $F(1, 170)$ | $F(1, 170)$ | $F(1, 170)$ |
| Report Consistency (Cy) | 13.00*** | 22.14*** | 28.86*** | 64.15*** | 12.73*** | 24.77*** | 15.15*** | 56.29*** |
| Complainant Gender (CG) | 2.05 | 1.51 | 4.03* | 1.19 | 1.32 | 0.84 | 1.55 | 10.81*** |
| Participant Gender (PG) | 1.63 | 1.44 | 0.47 | 1.11 | 4.22* | 0.33 | 2.67 | 0.00 |
| Cy x CG | 0.55 | 0.86 | 0.68 | 0.13 | 1.49 | 0.06 | 0.15 | 0.10 |
| Cy x PG | 0.33 | 0.37 | 0.43 | 0.02 | 0.27 | 0.08 | 0.13 | 0.25 |
| CG x PG | 1.45 | 0.05 | 1.31 | 1.81 | 0.04 | 1.45 | 3.08 | 0.05 |
| Cy x CG x PG | 0.22 | 1.01 | 0.08 | 0.81 | 0.01 | 0.31 | 0.04 | 0.23 |

*Notes.* IPA conditions $n = 183$; LA conditions $n = 178$. Multivariate $F$ ratios are Wilks' Lambdas. *$p < .05$. **$p < .01$. ***$p < .001$.

To examine the role of IPA severity and LA intensity ratings on similarity and group categorization judgements, two additional MANCOVAs were conducted, this time with similarity and group categorization judgements as the dependent variables. As before, the first included IPA severity as the covariate for participants in IPA conditions, and the second included LA intensity as the covariate for participants in LA conditions. In contrast with expectations, analyses revealed multivariate and univariate main effects of intensity ratings in LA conditions ($\eta_p^2$'s = .04), but no effects of severity ratings were observed in IPA conditions ($p$'s > .05). For both IPA and LA conditions, as may be seen in Table 2.11, the multivariate and univariate main effects of Report Consistency remained significant ($\eta_p^2$'s = .03 - .05).

Generally, few significant associations between severity or intensity ratings, effectiveness evaluations, and social categorization were observed. Significant effects of LA intensity on effectiveness evaluations and of IPA severity on similarity and group categorization judgements were not observed. Although analyses identified effects of IPA intensity on effectiveness evaluations and of LA intensity on similarity and group categorization judgements, inclusion of these variables as covariates did not substantially alter the data patterns. These results suggest that (1) vignettes were constructed such that severity of the behaviours described in IPA conditions and intensity of the activities described in the LA conditions were relatively constant, and (2) perceptions of the severity or intensity did not fulfill a significant role in effectiveness evaluations and social categorization.

***Table 2.11. Multivariate and Univariate Analyses of Covariance for Similarity and Group Categorization Ratings Controlling for Severity and Intensity Ratings***

| Variable | Multivariate $F(2, 177)$ | Univariate | |
|---|---|---|---|
| | | Similarity $F(1, 178)$ | Group Categorization $F(1, 178)$ |
| *IPA Conditions* | | | |
| IPA Severity | 1.30 | 1.14 | 0.57 |
| Report Consistency (Cy) | 4.71** | 4.84* | 7.79** |
| Complainant Gender (CG) | 0.15 | 0.26 | 0.15 |
| Participant Gender (PG) | 1.80 | 2.98 | 1.85 |
| Cy x CG | 0.17 | 0.26 | 0.01 |
| Cy x PG | 0.00 | 0.01 | 0.00 |
| CG x PG | 1.04 | 2.01 | 0.61 |
| Cy x CG x PG | 0.72 | 0.40 | 0.54 |
| *LA Conditions* | $F(2, 171)$ | $F(1, 172)$ | $F(1, 172)$ |
| LA Intensity | 3.41* | 1.60 | 6.47* |
| Report Consistency (Cy) | 3.65* | 7.10** | 4.70* |
| Complainant Gender (CG) | 0.27 | 0.51 | 0.14 |
| Participant Gender (PG) | 0.32 | 0.64 | 0.36 |
| Cy x CG | 0.19 | 0.38 | 0.13 |
| Cy x PG | 1.18 | 0.34 | 2.09 |
| CG x PG | 0.06 | 0.01 | 0.10 |
| Cy x CG x PG | 1.97 | 1.48 | 0.11 |

*Notes.* IPA conditions $n = 187$; LA conditions $n = 181$. Multivariate $F$ ratios are Wilks' Lambdas. *$p < .05$. **$p < .01$.

*Intimate Partner Abuse History and Leisure Activities Involvement*

As described in the Introduction, we may be more likely to sympathize with a complainant who we perceive to be more similar to ourselves, and IPA history and LA involvement may be important factors in determining the degree to which participants relate to the IPA and LA complainants, respectively. Thus, additional analyses were conducted to explore whether a history of IPA or current LA involvement had any bearing on effectiveness evaluations or social categorization. Approximately 27% of participants in IPA conditions reported having experienced physical abuse and 18% reported having used such acts against an intimate partner. In LA conditions, participants reported an average physical activity level of 4.35 ($SD$ = 1.46) on a scale from 1 (Not at all active) to 6 (Very active). Presented in Table 2.12, no significant point biserial correlations were found between effectiveness evaluations, IPA history, and LA involvement.

To examine the effects of IPA history and LA involvement on evaluations of complainant effectiveness, as for IPA severity and LA intensity, two MANCOVAs were conducted. Neither IPA history nor LA involvement predicted effectiveness evaluations ($p$'s > .05). As may be seen in Table 2.13, in IPA conditions, the multivariate and univariate main effects of Report Consistency on effectiveness evaluations remained significant ($\eta_p^2$'s = .14 - .38), with the exception of likeability ratings ($p$ > .05), as did the multivariate and univariate main effects of Report Consistency ($\eta_p^2$'s = .07 - .36) in LA conditions.

***Table 2.12. Correlations between Participants' Effectiveness Evaluations, Similarity and Group Categorization Ratings, and Intimate Partner Abuse History and Involvement in Leisure Activities***

| | IPA History | | LA Involvement |
|---|---|---|---|
| Ratings | Victimization | Perpetration | |
| *Effectiveness Evaluations* | | | |
| Intelligence | .10 | .04 | .09 |
| Honesty | .08 | .11 | .05 |
| Accuracy | .01 | -.03 | .01 |
| Suggestibility | .05 | .08 | -.07 |
| Confidence | -.07 | .12 | -.03 |
| Likeability | .02 | .01 | .09 |
| Credibility | .09 | .04 | .02 |
| *Similarity and Group Categorization* | | | |
| Similarity | .03 | .06 | .00 |
| Group Categorization | .07 | -.03 | -.06 |

*Notes*. IPA conditions $n$ = 185-188; LA conditions $n$ = 182-186. IPA Victimization: 0 = No, 1 = Yes; IPA Perpetration: 0 = No, 1 = Yes. Ratings were made on a 6-point scale (1 = Low; 6 = High).

***Table 2.13. Multivariate and Univariate Analyses of Covariance for Effectiveness Evaluations Controlling for Intimate Partner Abuse History and Involvement in Leisure Activities***

| | | Univariate | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | Multivariate $F(8, 167)$ | Intelligence $F(1, 171)$ | Honesty $F(1, 171)$ | Accuracy $F(1, 171)$ | Suggestibility $F(1, 171)$ | Confidence $F(1, 171)$ | Likeability $F(1, 171)$ | Credibility $F(1, 171)$ |
| *IPA Conditions* | | | | | | | | |
| IPA History | | | | | | | | |
| Victimization | 0.88 | 2.00 | 0.40 | 0.13 | 0.10 | 1.88 | 0.11 | 0.63 |
| Perpetration | 1.26 | 0.02 | 1.90 | 0.02 | 0.95 | 4.65* | 0.01 | 0.88 |
| Report Consistency (Cy) | 14.48*** | 29.12*** | 33.74*** | 71.98*** | 32.70*** | 33.63*** | 2.62 | 45.54*** |
| Complainant Gender (CG) | 0.26 | 0.24 | 0.30 | 0.00 | 0.43 | 0.02 | 0.00 | 0.37 |
| Participant Gender (PG) | 1.79 | 1.76 | 1.05 | 5.07* | 0.55 | 0.58 | 0.72 | 10.29** |
| Cy x CG | 0.71 | 0.77 | 0.15 | 0.17 | 0.30 | 3.50 | 1.48 | 0.48 |
| Cy x PG | 0.73 | 0.14 | 1.25 | 0.08 | 0.06 | 2.22 | 0.03 | 0.35 |
| CG x PG | 1.19 | 0.08 | 1.00 | 5.36* | 0.25 | 0.03 | 0.21 | 0.16 |
| Cy x CG x PG | 1.23 | 3.65 | 1.71 | 0.00 | 0.53 | 0.26 | 0.29 | 0.29 |
| *LA Conditions* | $F(7, 163)$ | $F(1, 169)$ | $F(1, 169)$ | $F(1, 169)$ | $F(1, 169)$ | $F(1, 169)$ | $F(1, 169)$ | $F(1, 169)$ |
| LA Involvement | 0.38 | 1.31 | 0.42 | 0.32 | 1.20 | 0.01 | 1.21 | 0.21 |
| Report Consistency (Cy) | 12.91*** | 21.92*** | 28.59*** | 63.65*** | 12.92*** | 24.59*** | 14.97*** | 55.85*** |
| Complainant Gender (CG) | 2.04 | 1.51 | 4.02* | 1.19 | 1.31 | 0.84 | 1.55 | 10.76*** |
| Participant Gender (PG) | 1.65 | 1.49 | 0.48 | 1.08 | 4.30* | 0.33 | 2.73 | 0.01 |
| Cy x CG | 0.49 | 0.67 | 0.58 | 0.17 | 1.24 | 0.07 | 0.08 | 0.08 |
| Cy x PG | 0.33 | 0.38 | 0.42 | 0.02 | 0.26 | 0.08 | 0.13 | 0.25 |
| CG x PG | 1.44 | 0.06 | 1.28 | 1.78 | 0.03 | 1.45 | 3.02 | 0.04 |
| Cy x CG x PG | 0.23 | 1.15 | 0.10 | 0.87 | 0.00 | 0.32 | 0.07 | 0.26 |

*Notes.* IPA conditions $n = 183$; LA conditions $n = 178$. Multivariate $F$ ratios are Wilks' Lambdas. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

As for effectiveness evaluations, MANCOVAs with similarity and group categorization judgements as the dependent variables and IPA history and LA intensity as the covariates demonstrated that neither IPA history nor LA involvement predicted similarity and group categorization judgements ($p$'s > .05). As may be seen in Table 2.14, the multivariate and univariate main effects of Report Consistency on effectiveness evaluations remained significant ($\eta_p^2$'s = .03 - .05) in both IPA and LA conditions.

On the whole, there was little support for the role of IPA history and LA involvement in effectiveness evaluations and social categorization. IPA history and LA involvement predicted neither similarity and group categorization judgments, nor evaluations of complainant effectiveness. Further, the data patterns remained largely unchanged controlling for these variables, suggesting that IPA and LA involvement did not inform evaluations of effectiveness and do not represent meaningful dimensions through which participants identified, or not, with complainants.

***Table 2.14. Multivariate and Univariate Analyses of Covariance for Similarity and Group Categorization Ratings Controlling for Intimate Partner Abuse History and Involvement in Leisure Activities***

| | | Univariate | |
|---|---|---|---|
| Variable | Multivariate $F$ (2, 176) | Similarity $F$ (1, 177) | Group Categorization $F$ (1, 177) |
| *IPA Conditions* | | | |
| IPA History | | | |
| Victimization | 0.81 | 0.00 | 1.44 |
| Perpetration | 1.25 | 1.29 | 0.41 |
| Report Consistency (Cy) | 4.94** | 4.60* | 8.49** |
| Complainant Gender (CG) | 0.24 | 0.49 | 0.06 |
| Participant Gender (PG) | 1.83 | 3.29 | 1.49 |
| Cy x CG | 0.14 | 0.17 | 0.03 |
| Cy x PG | 0.01 | 0.00 | 0.02 |
| CG x PG | 0.98 | 1.90 | 0.57 |
| Cy x CG x PG | 0.76 | 0.53 | 0.47 |
| *LA Conditions* | $F$ (2, 173) | $F$ (1, 174) | $F$ (1, 174) |
| LA Involvement | 0.82 | 0.01 | 0.97 |
| Report Consistency (Cy) | 4.74** | 8.84* | 6.95* |
| Complainant Gender (CG) | 0.31 | 0.49 | 0.05 |
| Participant Gender (PG) | 0.38 | 0.77 | 0.41 |
| Cy x CG | 0.23 | 0.40 | 0.36 |
| Cy x PG | 1.34 | 0.30 | 2.27 |
| CG x PG | 0.00 | 0.00 | 0.00 |
| Cy x CG x PG | 1.68 | 1.29 | 0.08 |

*Notes.* IPA conditions $n = 187$; LA conditions $n = 183$. Multivariate $F$ ratios are Wilks' Lambdas. *$p < .05$. **$p < .01$.

*Moderation and Mediation Analyses*

Having tested the effects of Report Consistency, Report Content, and Complainant and Participant Gender on participants' evaluations, a final set of analyses was conducted to examine the potential moderating and mediating roles of social categorization in effectiveness evaluations. Given the lack of association observed between suggestibility ratings and judgements of similarity and group categorization (see Table 2.5), the suggestibility scale was excluded from moderation and mediation analyses. To minimize the number of analyses, responses to each question on the effectiveness questionnaire were collapsed to create a single score that served as the outcome measure ($M = 3.94$, $SD = 0.96$; Range = 1.67 – 6.00), hereafter referred to as *overall effectiveness scores* (for a similar computation, see Connolly et al., 2008). An exploratory factor analysis of the remaining six effectiveness scales (i.e., excluding suggestibility) revealed a one factor solution, supporting this computation.[18] With regard to social categorization, analyses demonstrated that similarity and group categorization ratings were highly correlated ($r = .55$, $p < .001$) and that mean ratings did not differ significantly ($p > .05$). Thus, the similarity and group categorization ratings also were collapsed to create one combined score ($M = 3.05$, $SD = 1.30$; Range = 1.00 – 6.00), hereafter referred to as *social categorization scores*.

In the subsequent sections, I begin with an examination of social categorization as a moderator, followed by a test of its potential role as a mediator. Correlations relevant to these analyses can be found in Table 2.15.

[18] The analysis extracted one factor, accounting for 51% of the variance. Communalities ranged between .37 and .68, with a mean of .51. Factor loadings ranged between .61 and .82, with a mean of. 71.

***Table 2.15. Correlations between Report Consistency, Report Content, Complainant Gender, Participant Gender, Overall Effectiveness, and Social Categorization***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Report Consistency | -- | | | | | |
| 2. Report Content | -.01 | -- | | | | |
| 3. Complainant Gender | .01 | -.01 | -- | | | |
| 4. Participant Gender | -.01 | .00 | -.05 | -- | | |
| 5. Overall Effectiveness | .52*** | -.18** | .08 | -.05 | -- | |
| 6. Social Categorization | .21*** | -.22*** | .01 | -.10 | .53*** | -- |

*Notes.* $n$ = 373-374. **$p$ < .01. ***$p$ < .001. Report Content: 0 = LA; 1 = IPA; Complainant Gender and Participant Gender: 0 = Male; 1 = Female.

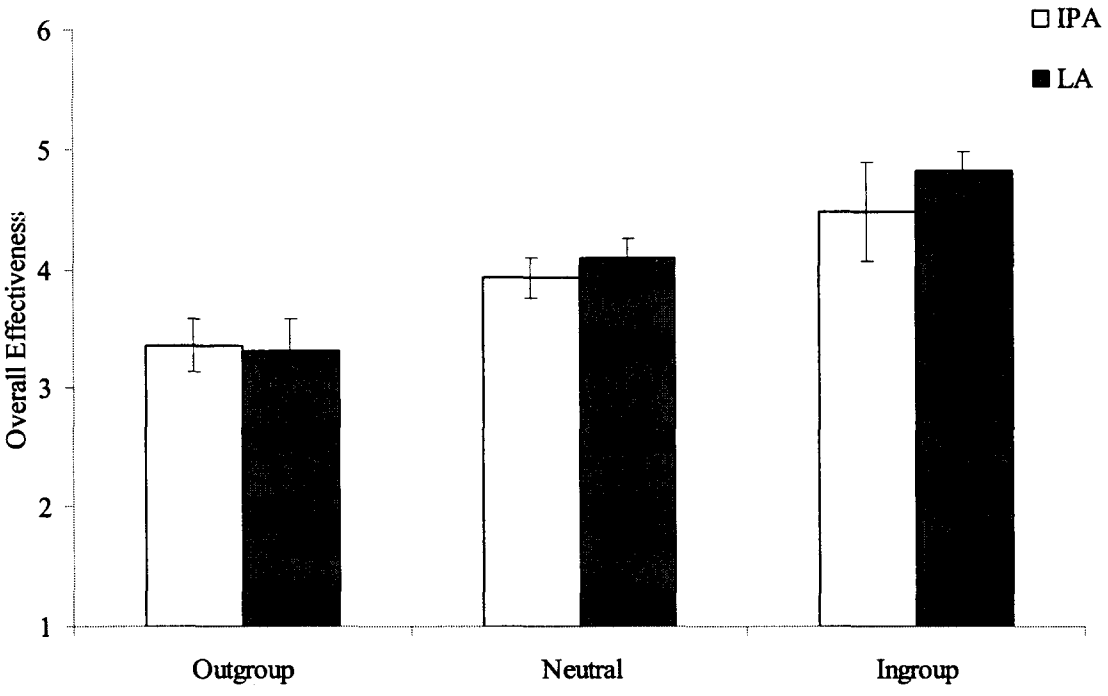
*Moderation.* I first tested whether social categorization moderated the observed effects of Report Consistency, Report Content, and Complainant and Participant Gender on evaluations of complainant effectiveness, following the criteria described by Kraemer and colleagues (2001; 2002; 2008). The model proposed here is an interaction effect where overall effectiveness scores depend on the interaction between the independent variables (i.e., Report Consistency, Report Content, and Complainant and Participant Gender) and social categorization scores. In a moderated model, social categorization scores must exist independently from the independent variables to influence the relationship between the reported event and overall effectiveness scores (Kraemer et al., 2001; 2002; 2008).

To test the moderator hypothesis, social categorization scores were used to create the grouping variable, Group Identity (outgroup: $1 \leq$ social categorization scores $< 3$; neutral: $3 \leq$ social categorization scores $\leq 4$; ingroup: $4 <$ social categorization scores $\leq 6$). An ANOVA revealed a significant effect of Group Identity on overall effectiveness scores, $F(2, 325) = 38.19$, $p < .001$, $\eta_p^2 = .19$. *Post hoc* comparisons demonstrated that participants who categorized the complainant as being a member of the outgroup had lower overall effectiveness scores ($M = 3.34$, $SD = 0.92$) than participants whose social categorization scores were neutral ($M = 4.01$, $SD = 0.81$), $t(298) = 6.63$, $p < .001$, $d = .77$, whose ratings of overall effectiveness, in turn, were lower than participants who categorized the complainant as being a member of the ingroup ($M = 4.73$, $SD = 0.69$), $t(250) = 6.73$, $p < .001$, $d = .85$. A modest but significant Group Identity x Report Content interaction effect on effectiveness evaluations was observed, $F(2, 325) = 3.71$, $p < .05$, $\eta_p^2 = .02$. As may be seen in Figure 2.4, *post hoc* comparisons, however, revealed

only one difference that approached significance: Amongst those categorized as members of the ingroup, LA complainants received more positive effectiveness evaluations than did IPA complainants, $t(72) = 1.93, p = .06, d = .45$. All other comparisons failed to reach, or even approach, significance, and no other Group Identity interaction effects were observed ($p$'s > .05).

In general, results of this analysis fail to support the moderating role of social categorization in evaluations of complainant effectiveness. Although a significant interaction between Group Identity and Report Content was observed, the data pattern violates one of the requirements of the moderation model; that is, that the moderated variable (Report Content) and the moderator (Group Identity or social categorization scores) are not associated (Kraemer et al., 2001; 2002; 2008). Examination of Table 2.15 reveals a highly significant correlation between Report Content and social categorization scores, and a subsequent chi-square analysis demonstrated that Group Identity varies systematically with Report Content, $\chi^2(1, N = 374) = 21.71, p < .001, \Phi = .24$. Thus, in the present study, social categorization does not appear to moderate evaluations of complainant effectiveness.

***Figure 2.4. Group Identity by Report Content Interaction Effect on Effectiveness Evaluations***




*Notes.* Bars represent 95% confidence intervals around the means. 1 = Low; 6 = High.

*Mediation.* I then used a mediation model to test whether similarity and group categorization mediated the observed relationships between Report Consistency, Report Content, and Complainant and Participant Gender. The question addressed here was whether the effects of Report Consistency, Report Content, and Complainant and Participant Gender on participants' judgements of complainant effectiveness could be explained as a function of how similar participants thought the complainant was to them and how closely they saw themselves as belonging to the same group as the complainant. The model was tested using regression analyses following the 4-step procedure initially outlined by Baron, Kenny, and Judd (Baron & Kenny, 1986; Judd & Kenny, 1981) and recently modified by a MacArthur Foundation Network subgroup (Kraemer et al., 2001; 2002; 2008).

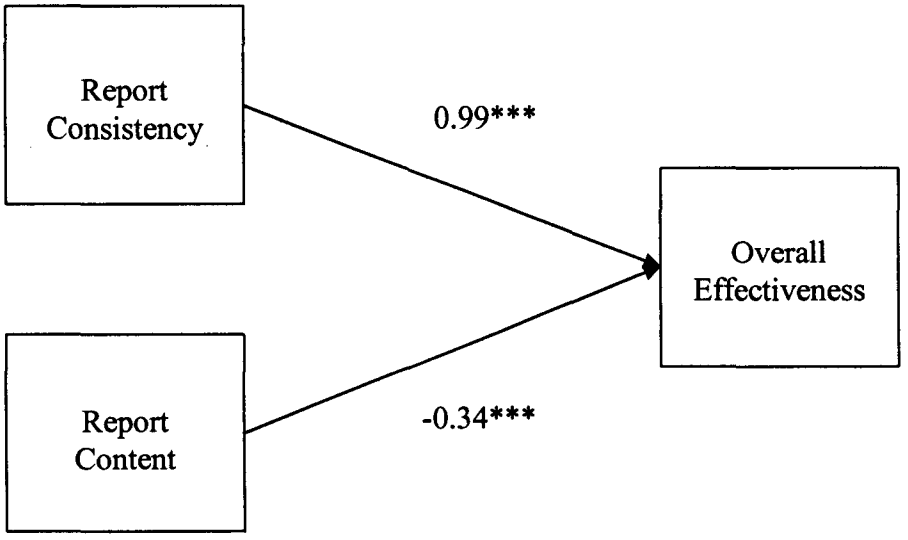
The first step of testing this mediation model required establishing that the relationship between the predictors and outcome exists. This assumption was partially met: Report Consistency and Report Content were significantly correlated with overall effectiveness scores, but Complainant Gender and Participant Gender were not, nor was their interaction. Thus, these latter predictors were excluded from the mediation model. The second step required that a relationship between the predictors (Report Consistency and Report Content) and the proposed mediator (social categorization scores) is established. This requirement was met. As may be seen in Table 2.15, Report Consistency and Report Content were significantly correlated with social categorization scores, and a regression analysis demonstrated that Report Consistency and Report Content significantly predicted social categorization scores, $F(2, 371) = 19.48$, $p < .001$, accounting for 10% of the variation. For the third step, the relationship between the

proposed mediator (social categorization scores) and outcome (overall effectiveness scores) must be established. Again, this criterion was met. These scores were significantly correlated (see Table 2.15) and a regression demonstrated that social categorization scores accounted for 28% of the variation in overall effectiveness, $F(1, 372) = 142.23, p < .001$.

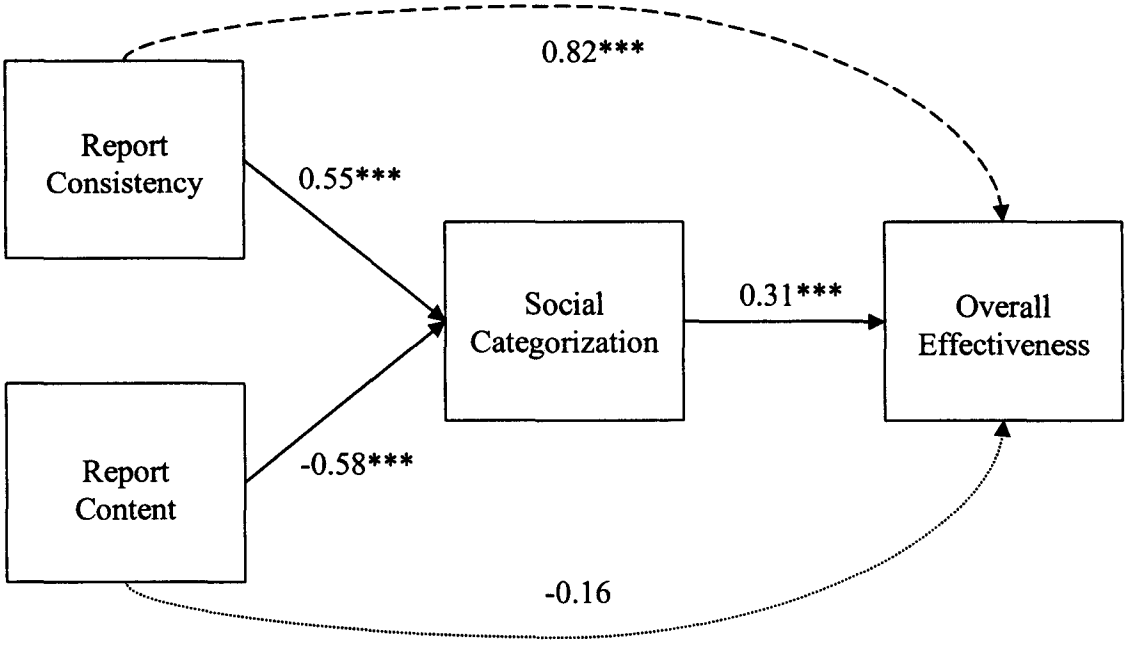
For the final step, two regression analyses were conducted: The first tested the direct effects of Report Consistency and Report Content on overall effectiveness scores, and the second included social categorization scores in the model to control for their effects. The direct effects of Report Consistency and Report Content on overall effectiveness scores were large. Together, Report Consistency and Report Content accounted for 30% of the variability in overall effectiveness, $F(2, 371) = 77.94, p < .001$ (see Figure 2.5). Including social categorization scores increased the variability in overall effectiveness scores accounted for by the model to 45%, $F(3, 370) = 102.14, p < .001$, a substantial increase (15%) over the variance accounted for in the previous model, $F(1, 370) = 106.30, p < .001$. Controlling for social categorization scores in this second model, Report Content no longer contributed significantly to the prediction of overall effectiveness ($p > .05$) and significantly decreased the effect of Report Consistency, $z = 4.02, p < .001$ (Sobel, 1982). These results indicate that social categorization scores fully mediate the observed relationship between Report Content and overall effectiveness scores, but only partially mediate the relationship between Report Consistency and overall effectiveness scores (see Figure 2.5).

***Figure 2.5. Model of Social Categorization Scores Mediating the Effects of Report Consistency and Report Content on Overall Effectiveness Scores***

*Model Excluding Social Categorization Scores*




*Model Including Social Categorization Scores*




*Notes*. Values are unstandardized regression coefficients. ***$p < .001$. The dotted line indicates full mediation; the dashed line indicates partial mediation.

In sum, there is evidence supporting the role of social categorization as a mediator, but not a moderator, of effectiveness evaluations. Although an interaction effect was observed between Report Content and social categorization, the association between them violates the criterion that the moderator exists independently from the independent variables to affect the outcome, as stipulated by the MacArthur approach (Kraemer et al., 2001; 2002; 2008). Instead, there was evidence to support the role of social categorization as mediating evaluations of complainant effectiveness: Ratings of similarity and group categorization fully explained the effects of Report Content on evaluations of complainant effectiveness such that complainants reporting on LA participation were judged to be more similar and more likely to belong to the same group as participants compared to complaints reporting IPA victimization which resulted in more positive evaluations of LA than IPA complainant effectiveness. With respect to Report Consistency, in contrast, ratings of similarity and group categorization only partially explained the effects. Although consistent complainants were judged to be more similar and more likely to belong to the same group as participants compared to inconsistent complainants, controlling for these ratings attenuated but did not eliminate differences in perceived effectiveness between IPA and LA complainants, suggesting that additional factors or processes explain the effects of Report Consistency on evaluations of complainant effectiveness.

## *Discussion*

This study examined evaluations IPA complainants, focusing on the roles of consistency and social categorization. Varying the consistency of reports, as well as complainant and participant gender, I compared effectiveness of complainants reporting IPA victimization to that of 'complainants' reporting LA participation. Overall, results strongly suggest that participants used consistency as a proxy for accuracy in evaluations of credibility. Findings of the present study, like those of Berman and Cutler (1996), Berman et al. (1995), and Brewer and Hupfeld (2004), demonstrated a negative impact of inconsistency on evaluations of perceived credibility or witness effectiveness. In fact, in the present study, this was the largest and most reliable effect across conditions and analyses. Consistency also reliably affected ratings of similarity and group categorization such that participants rated consistent complainants as significantly more similar and significantly more likely to belong to the same group as themselves.

Results additionally suggest that reports of IPA victimization *are* received with more scepticism than reports of everyday experiences: LA complainants generally were rated more favourably than were IPA complainants. Further, findings suggest that victim versus non-victim *is* a socially meaningful dimension through which group categorizations are made to the extent that similarity and group categorization ratings were higher (indicating greater similarity and greater likelihood of belonging to the same group) for LA complainants compared with IPA complainants. However, the answer to whether inconsistencies are more detrimental to effectiveness evaluations of IPA complainants appears to be 'no'. There was no evidence for an interaction between

Report Consistency and Report Content, either on evaluations of effectiveness, or on judgements of similarity and group categorization.

Although evaluations were expected to differ with participant and complainant gender and, for reports of IPA victimization in particular, analyses failed to demonstrate significant effects of Complainant Gender. Moreover, in contrast with expectations, the observed effects of Participant Gender on effectiveness evaluations revealed that male participants made more favourable evaluations than female participants, although actual differences generally were quite small. A significant Report Content by Complainant Gender by Participant Gender interaction was found, but again, differences were in contrast with expectations. Men rated female IPA complainants as more accurate and more credible than did women, whereas ratings of male IPA complainants did not differ as a function of Participant Gender. No differences between evaluations made by men versus women were found for LA complainants. Further, there was no evidence for a gender similarity bias, in contrast with past research (e.g., Elkins et al., 2002; Elkins et al., 2001; Johnson et al., 1995; McKelvie, 2002; Worthen & Varnado-Sullivan, 2005), nor was there evidence for the anticipated interaction of victim by gender categorization (i.e., that a same-gender complainant reporting on LA participation would be rated as more similar than an other-gender complainant reporting on IPA victimization).

Findings regarding the role of similarity and group categorization judgements support the role of social categorization as a mediator, rather than a moderator, of evaluations of the complainant. The mediation model, however, deviated from expectations given that the anticipated effects of Complainant Gender and Participant Gender, and the gender similarity bias specifically, were not found. Nonetheless, analyses

demonstrated full mediation of the effects of Report Content on overall effectiveness scores and partial mediation of the effects of Report Consistency. In other words, complainants reporting on LA participation were judged to be more similar and more likely to belong to the same group as participants compared to complaints reporting IPA victimization which resulted in more positive evaluations of LA than IPA complainants. Consistent complainants also were judged to be more similar and more likely to belong to the same group as participants compared to inconsistent complainants; however, social categorization only partially explained differences in evaluations of IPA and LA complainants.

Overall, the observed pattern of results is generally in accordance with the findings and explanation of Brewer and Hupfeld (2004). Obvious inconsistencies appear to have been interpreted as strong evidence that the event *did not* happen. However, in contrast with the Testimonial Consistency by Group Identity interaction observed by Brewer and Hupfeld, Group Identity only was found to interact with Report Content and did not fulfill a moderator role. Social categorization did not appear to drive the extent to which participants considered evidential versus extra-evidential information in their evaluations of complainant effectiveness and failed to predict when inconsistencies, in particular, affected evaluations. Instead, results supported a mediation model of decision-making. The effects of similarity and group categorization were observed in both consistent and inconsistent conditions, suggesting that report consistency alone, at least in the present study, is not *enough* evidence. That is, there was indication of both systematic and heuristic processing across conditions independent of variation in report consistency. Participants appeared to use whatever additional information was available to inform

their judgements, including decision-irrelevant or extra-evidential heuristics, such as similarity and group categorization.

Compared with the stimulus materials used in the Brewer and Hupfeld research, participants in the present study were provided with a reduced evidentiary basis to inform their decisions which may account for the disparate results. Specifically, participants in the present study read 2-page vignettes whereas participants in the Brewer and Hupfeld study listened to a 25-minute audiotape, as well as reading a newspaper extract. Discrepancies between the Brewer and Hupfeld findings and the present results also may be due to the variation in decision contexts (see Bodenhausen & Lichtenstein, 1987). The Brewer and Hupfeld study employed a mock jury design, asking participants to render decisions regarding not only witness effectiveness, but also verdict, in the context of a mock criminal trial. The present study, in contrast, asked participants to evaluate an adult along a variety of dimensions, free of the consequences of decisions implicit within a mock jury design (i.e., the consequences of finding defendant guilty versus not guilty).

*Limitations*

This study is not without limitations, a few of which I discuss here. First, no information was provided to participants regarding the context or consequences of their evaluations. Such ratings may not predict decision-making, and verdict specifically, in the high-stakes context of an actual court case. As the findings of Berman et al. (1995), Brewer and Burke (2002), and Potter and Brewer (1999) suggest, the impact of consistency on judgements of guilt may be negligible, despite decision makers' beliefs regarding the relationship between consistency and accuracy. Similarly, the generalizability of results to 'real-world' jury decision-making are limited by the present

study's use of fictional interviews about fictional events and the absence of deliberation. Further, although representative of the Canadian population, the present study's sample may not necessarily be reflective of the characteristics of individuals who serve as jurors in Canadian courts.

Second, differences between, and not within, individuals were examined. Including individual difference measures, such as scales assessing attitudes towards women and men, the extent to which participants believe in a just world, and knowledge or beliefs about domestic violence, may account for or explain additional variation in effectiveness evaluations. Hillier and Foddy (1993), for example, found that participants endorsing 'traditional' attitudes towards women placed greater blame for the abuse with the (female) victim of IPA and less with the (male) perpetrator than did participants reporting more egalitarian attitudes. Finally, perceptions regarding the strength of evidence were not directly assessed, limiting interpretation of the findings with regard to heuristic-systematic processing theory, nor was information regarding the temporal sequencing of participants' judgements (e.g., whether ratings of similarity and group categorization preceded or followed effectiveness evaluations).

*Conclusions and Future Directions*

Comparing complainants reporting on everyday autobiographical events to complainants reporting on IPA victimization experiences, the main goals of this study were to elucidate the effects of consistency on perceptions of complainant credibility or witness effectiveness and to determine whether these effects are greater for IPA complainants compared to persons reporting on everyday events. It was anticipated that evaluations of effectiveness would vary as a function of both the content and consistency

of the report; this prediction, however, was not supported. Findings instead endorse the role of social categorization as mediating evaluations of LA versus IPA complainants, irrespective of the consistency of reports. To the extent that the Report Content variable reflects variation along this dimension, victim versus non-victim appears to a socially meaningful category through which judgements regarding social categorization are made. With the present design, however, it is not possible to determine how or why differential similarity and group categorization judgements were made.

Future research could build upon the present study by explicitly asking participants to rate the strength of evidence. Such ratings may help clarify why effects of social categorization were observed across consistency conditions in the present study. Testimonial consistency may not be sufficient evidence to conclude that the event occurred *beyond a reasonable doubt,* increasing reliance on heuristic processing (Chaiken et al. 1989; Chen & Chaiken, 1999; Chen et al., 1999). The implications for 'real-world' adjudication of IPA cases, typically involving one party's word against the other, are such that evaluations of the allegations may be predominantly heuristic in nature. Thus, investigations of means through which we can reduce reliance on heuristic processing are worth pursuing. Further research examining characteristics of the assessor would contribute to our understanding of the socially meaningful dimensions through which us-them categorizations are made and also may elucidate why the gender similarity bias demonstrated in past research was not found in the present study. Finally, findings would be strengthened through replication in more ecologically valid research designs, such as ones using summaries from or simulations of domestic violence cases heard in court and including deliberation.

In conclusion, this study represents an extension of the theory and work of Brewer and Hupfeld (2004), demonstrating that both evidential and extra-evidential report characteristics shape evaluations of witness effectiveness or complainant credibility. The strengths of the study lie within its application of social cognitive theory to perceived effectiveness research, as well as design features including the use of multiple vignettes and the large sample of jury-eligible community participants. Findings suggest that inconsistencies in reports across repeated questioning are detrimental evaluations of the witness or complainant and contribute to the development of a process understanding of perceived credibility.

# GENERAL DISCUSSION

Memory is variable and reports of autobiographical events can change over time, from being incomplete and distorted to being more complete and accurate (*cf.* Zola, 1998). Such changes may have significant implications for evaluations of these reports. Given the sheer prevalence, associated costs (personal, social, and economic), and repetitive nature of IPA, there is a need to investigate factors which may affect intervention and potentially contribute to the winnowing of IPA complaints through the criminal justice system. Across two studies, this dissertation examined IPA report consistency and perceived effectiveness, asking whether the observed phenomenology was unique to the context of IPA or generalisable to reports of everyday events. The first study compared consistency of reports of the abusive and everyday autobiographical event prevalence (i.e., occurrence and frequency) both within and between victim and non-victim samples. The second study examined evaluations of complainant effectiveness, focusing on the roles of report consistency over repeated questioning and social categorization. Taken together, results suggest that victims' disclosures of IPA may be received with scepticism, although actual differences between the characteristics of reports of abusive and everyday autobiographical event prevalence were few. Such discrimination may increase the likelihood that IPA complaints are disregarded, potentially precluding appropriate legal intervention.

In the first project to examine IPA report consistency and its effects on perceived effectiveness, results support the value of such investigation and offer direction for

continued research on the topic. In particular, the focus in Study 1 on consistency in the occurrence and frequency of abuse reported (rather than descriptions of a particular instance or specific event) may be relevant to legal outcomes, including findings of innocence or guilt, sentencing, or financial compensation. Additionally, the calendar portion of the Danger Assessment (DA; Campbell, 2003), a commonly used domestic violence risk assessment tool, may be admitted as evidence in domestic violence cases to establish a pattern and history of abuse (Campbell, 2008). Briefly, completed by a law enforcement official, health care professional, or victim advocate and the victim together, the DA is designed to predict risk of intimate partner homicide. The first portion of the instrument assesses frequency and severity of abuse by way of a calendar-based method similar to that used in Study 1: The victim is presented with a calendar of the past year and is asked to mark the approximate days when physically abusive incidents occurred, ranking the severity of the incident on a scale from 1 (slap, pushing, no injuries and/or lasting pain) to 5 (use of weapon, wounds from weapon). The second portion of the DA comprises 20 risk factors, scored yes/no based in part of the calendar responses, associated with intimate partner homicide. As evidenced by Study 1, however, we may anticipate considerable variation in calendar responses over time.

In the following sections, these issues are explored in more detail. First, I discuss the implications of the findings with regard to the criminal justice system, focusing on the importance of understanding the reason(s) for inconsistencies, and second, offer procedural recommendations, such as the use of interview techniques designed to maximize consistency and education for triers-of-fact regarding the accuracy-consistency relationship.

### *Implications and Recommendations for the Criminal Justice System*

The prevalence of inconsistencies observed in Study 1 and the impact of inconsistencies on perceived effectiveness observed in Study 2 suggest that the use of consistency as a proxy for accuracy in criminal justice proceedings merits some discussion. Unlike most estimator variables which are not under the control of the criminal justice system (*cf.* Wells & Olson, 2003), consistency of information reported from police interview to courtroom testimony may be influenced by criminal justice system procedures. Assuming that complainants are being truthful and reporting on true events, recommendations regarding interview procedures for maximizing consistency are the same as those discussed in the literature pertaining to maximizing accuracy of reports. In particular, a semi-structured interview format, free of suggestive or leading questions, such as the Cognitive Interview (e.g., Fisher, McCauley, & Geiselman, 1994) or the Step-Wise Interview (e.g., Yuille, Hunter, Joffe, & Zaparniuk, 1993), should enhance the consistency of reports over repeated questioning (Krackow & Lynn, 2003).

There is, however, a plethora of reasons for observing inconsistencies in reports over time and it is these reasons that are of the utmost relevance to the adjudication of IPA cases. As Talarico and Rubin (2003, 2007) noted, consistency is easier to measure than accuracy, but may not be sufficient to deem the information as accurate. In fact, the body of work examining the statistical relationship between consistency and accuracy (e.g., Brewer et al., 1999; Fisher & Cutler, 1995; Penrod & Cutler, 1995) suggests that consistency is neither a sufficient nor a necessary condition of accuracy. Discussed in some detail in Study 1, decreased disclosures, for example, may not necessarily reflect false allegations but instead may reflect memory impairments, such as forgetting over

time or the effects of post-event information. However, decreased disclosures also may be unrelated to memory, resulting instead from other influences, including fear of retaliation by the accused or fear of the legal consequences of making such allegations. For instance, many North American jurisdictions have implemented policies that require police and prosecutors to charge and prosecute all IPA allegations where there is reason to believe that an offence has been committed (Dutton & Corvo, 2006; Hoctor, 1997; Miller, 2004). Such action, however, may conflict with victims' desires to reunite with their partners and may ultimately contribute to recantation of their IPA complaints (Hoyle & Sanders, 2000; Russell & Light, 2006; Ursel & Brikey, 1996).

A complainant's expectations and interpretation of the event also can contribute to "fundamental changes in the reports" (Haber & Haber, 2000, p. 1063). IPA victims do not necessarily recognize the behaviours perpetrated against them as abusive for a variety of reasons, including cognitive distortions or cultural, religious, and generational beliefs regarding normative spousal behaviours (e.g., Andrews & Brewin, 1990; Barnett, 2001; Klevens et al., 2007; O'Neill & Kerig, 2000; Pape & Arias, 2000; Ramsey-Klawsnick, 2003; Shurman & Rodriguez, 2006). Indeed, research conducted by Desmarais et al. (2006) demonstrated an association between cognitive distortions (including misattributions of responsibility for the abuse and failure to label the experiences as abusive) and report inconsistency in a sample of female victims of IPA. Specifically, results demonstrated that the greater the endorsement of cognitive distortions, the more report inconsistencies were observed in the direction of decreased disclosure. Consequently, criminal justice system professionals and decision makers should

recognize that inconsistencies may be attributable to cognitive distortions and other factors rather than necessarily indicating inaccurate or false allegations.

To this end, procedural safeguards similar to those established to prevent mistaken identifications from resulting in erroneous convictions could be implemented to attempt to reduce the likelihood that inconsistencies are misinterpreted as necessarily reflecting inaccuracy or fabrication. For example, expert testimony on the topic could be proffered. However, for such testimony to be admissible, it must be determined that the nature of the relationship between accuracy and consistency is *beyond the ken* of the trier(s)-of-fact. Although there has been a recent increase in research reassessing expert and lay beliefs about eyewitness topics (see Read & Desmarais, in press), the accuracy-consistency relationship (or lack thereof) has not been included in such surveys. Even if it is established that knowledge is deficient, the effectiveness of expert testimony in educating and sensitizing jurors would need to be established (*cf.* Cutler, Penrod, & Dexter, 1989). Jury instructions may be another procedural safeguard through which to educate jurors regarding the normative nature of inconsistencies. The research evidence regarding the effectiveness of jury instructions in sensitizing jurors, however, is equivocal and, as with expert testimony, would need to be established (*cf.* Ogloff & Rose, 2005).

## *General Conclusions and Future Directions*

Inconsistencies were operationalized quite differently in Study 1 and Study 2. Specifically, in Study 1, inconsistencies referred to disagreement in either the occurrence or frequency of the events report, whereas in Study 2, inconsistencies comprised four contradictions in the details of the event described, as well as the introduction of new

information, during the follow-up interview. Further, type of inconsistency (i.e., inconsistency in the occurrence vs. inconsistency in the frequency of events), although examined in Study 1, was not manipulated in Study 2. Thus, one future research direction to pursue is the effects of type of inconsistencies on perceived credibility. Even though Berman et al. (1995) manipulated type of inconsistency (i.e., inconsistency in central vs. peripheral report details) with minimal effect on credibility evaluations, results of Study 2 of this dissertation emphasized the importance of report content, warranting continued examination of this issue; that is, whether inconsistencies in central and peripheral details or in event occurrence and frequency differentially affect how jurors evaluate reports of abusive experiences versus everyday events. In addition to comparing the effects of different types of inconsistencies, it may be informative to explore the effects of decreased disclosures compared with increased disclosures on perceptions of complainant or witness effectiveness.

Related to the issues of type or direction of inconsistency, the effects of level of inconsistency on evaluations of credibility or effectiveness also should be explored in future research. Results of Study 1 suggest that the expectation of absolute consistency over time in reports of autobiographical memory events is unrealistic, whether the memories are of abusive experiences or everyday events; however, varying levels of inconsistency may produce lesser or greater effects. It is possible that there exists a minimal level of inconsistency which has negligible effects on perceived credibility. Thus, future research could explore the threshold at which inconsistencies become detrimental to evaluations (i.e., examining what level of inconsistency is normative) and whether the threshold varies with report content. Identification of such a threshold also

would offer an explanation for the mixed effects of inconsistency on perceptions credibility and effectiveness observed across studies.

In light of the differences observed in report consistency as a function of IPA category, another research direction worth pursuing may be to compare evaluations of a complainant's credibility or effectiveness across types of IPA (psychological, physical assault, or sexual). IPA acts described in Study 2's vignettes comprised primarily physical assaults, albeit with some features of psychological aggression. For reasons including the stigmatization of male sexual assault victims, the effects of factors (e.g., social categorization) on jurors' evaluations and decision-making may vary. Further, additional research examining (1) the characteristics of the individual and of the experiences that influence the consistency of reports of event prevalence, and (2) the impact of these characteristics on perceived credibility is needed. Although Study 2 demonstrated the relevance of social categorization, still other factors must be involved.

In summary, this dissertation used novel methodologies to explore consistency of IPA reports and the perceived credibility of IPA complainants. Results of Study 1 suggest that comparisons of report inconsistencies within categories of abusive experiences and everyday events warrant further investigation. Results of Study 2 demonstrate that evaluations of a complainant's credibility or effectiveness comprise complex and interactive processes involving characteristics of both the evaluator and complainant. Findings speak to the traumatic memory debate, providing support for the body of research suggesting that memories of abusive events, arguably traumatic in nature, appear to reflect memory processes typical of other more mundane autobiographical events, and contribute to a process understanding of perceived credibility.

# REFERENCES


Abshire, J., & Bornstein, B. H. (2003). Juror sensitivity to the cross-race effect. *Law and Human Behavior, 27,* 471-480.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (3rd ed., text rev.). Washington, DC: Author.

American Psychological Association (APA) Working Group on Investigation of Memories of Childhood Abuse. (1998). Final conclusions of the American Psychological Association Working Group on Investigation of Memories of Childhood Abuse. *Psychology, Public Policy, and Law, 4,* 933-940.

American Psychological Association (APA) Working Group on Intimate Partner Abuse and Relationship Violence. (2002). *Intimate partner abuse and relationship violence.* Retrieved March 20th, 2006 from http://www.apa.org/pi/iparv.pdf.

Anderson, S. J., & Conway, M. A. (1993). Investigating the structure of autobiographical memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1178-1196.

Andrews, B., & Brewin, C. R. (1990). Attributions of blame for marital violence: A study of antecedents and consequences. *Journal of Marriage & the Family, 52,* 757-767.

Arias, I., & Pape, K. T. (1999). Psychological abuse: Implications for adjustment and commitment to leave violent partners. *Violence and Victims, 14,* 55-67.

Bachman, R., & Coker, A. L. (1995). Police involvement in domestic violence: The interactive effects of victim injury, perpetrator's history of violence, and race. *Violence and Victims, 10,* 91-106.

Barnett, O. W. (2001). Why battered women do not leave, part 2: External inhibiting factors – social support and internal inhibiting factors. *Trauma, Violence, and Abuse, 2,* 3-35.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

Belli, R. F. (1998). The structure of autobiographical memory and the Event History Calendar: Potential improvements in the quality of retrospective reports in surveys. *Memory, 6,* 383-406.

Belli, R. F., Shay, W. L., & Stafford, F. P. (2001). Event history calendars and question list surveys: A direct comparison of interviewing methods. *Public Opinion Quarterly, 65,* 45-74.

Bennett, L., Goodman, L., & Dutton, M. A. (1999). Systematic obstacles to the criminal prosecution of a battering partner: A victim perspective. *Journal of Interpersonal Violence, 14,* 761-772.

Berman, G. L., & Cutler, B. L. (1996). Effects of inconsistencies in eyewitness testimony on mock-juror decision making. *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.

Berntsen, D., Willert, M., & Rubin, D. C. (2003). Splintered memories of vivid landmarks? Qualities and organization of traumatic memories with and without PTSD. *Applied Cognitive Psychology, 17,* 675-693.

Blackwell, B. S., & Vaughn, M. S. (2003). Police civil liability for inappropriate response to domestic assault victims. *Journal of Criminal Justice, 31,* 129-146.

Bodenhausen, G. V., & Lichtenstein, M. (1987). Social stereotypes and information-processing strategies: The impact of task complexity. *Journal of Personality and Social Psychology, 52,* 871-880.

Bogart, L. M., Walt, L. C., Pavlovic, J. D., Ober, A. J., Brown, N., & Kalichman, S. C. (2007). Cognitive strategies affecting recall of sexual behavior among high-risk men and women. *Health Psychology, 26,* 787-793.

Bohanek, J. G., Fivush, R., & Walker, E. (2005). Memories for positive and negative emotional events. *Applied Cognitive Psychology, 19,* 51-66.

Bottoms, B. L., Golding, J. M., Stevenson, M. C., Wiley, T. R. A., & Yozwiak, J. A. (2006). 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.), *Handbook of eyewitness psychology, Volume 1: Memory for events* (pp. 509-543). Mahwah, NJ: Erlbaum.

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior, 26,* 353-364.

Brewer, N., & Hupfeld, R. M. (2004). Effects of testimonial inconsistencies and witness group identity on mock-juror judgments. *Journal of Applied Psychology, 34,* 493-513.

Brewer, N., Potter, R., Fisher, R. P., Bond, N., & Luszcz, M. (1999). Beliefs and data on the relationship between consistency and accuracy of eyewitness testimony. *Applied Cognitive Psychology, 14,* 297-313.

Brewin, C. R. (2007). Autobiographical memory for trauma: Update on four controversies. *Memory, 15,* 227-248.

Brewin, C. R. (2001). A cognitive neuroscience account of posttraumatic stress disorder and its treatment. *Behaviour Research and Therapy, 39,* 373-393.

Briere, J., & Scott, C. (2006). *Principles of trauma therapy: A guide to symptoms, evaluation, and treatment.* Thousand Oaks, CA: Sage.

Brown, G. A. (2004). Gender as a factor in the response of the law-enforcement system to violence against partners. *Sexuality and Culture: An Interdiscipinary Quarterly, 8,* 3-139.

Brown, R. A., Burgess, E. S., Sales, S. D., Whiteley, J. A., Evans, D. M., & Miller, I. W. (1998). Reliability and validity of a smoking timeline follow-back interview. *Psychology of Addictive Behaviors, 12,* 101-112.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review, 106,* 676-713.

Campbell, J. C. (2008, April). *Using the Danger Assessment: Applications & dissemination.* Workshop presented at the Fifth Annual Forensic Psychiatry Conference, Vancouver, Canada.

Campbell, J. C. (2003). *The Danger Assessment.* Retrieved January 4th, 2006 from http://www.dangerassessment.com

Canada Evidence Act (R.S., 1985, c. C-5)

Canadian Judicial Council Model Jury Instructions in Criminal Matters. (2004).

Carey, M. P., Carey, K. B., Masito, S. A., & Gordon, C. M., & Weinhardt, L. S. (2001). Assessing sexual risk behaviour with the Timeline Followback (TLFB) approach: Continued development and psychometric evaluation with psychiatric outpatients. *International Journal of STD & AIDS, 12,* 365-375.

Catania, J. A., Gibson, D. R., Chiwood, D. D., & Coates, T. J. (1990). Methodological problems in AIDS behavioral research: Influences on measurement error and participation bias in studies of sexual behavior. *Psychological Bulletin, 108,* 339-362.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. Uleman & J. Bargh (Eds.), *Unintended thought* (pp. 212-252). New York: Guilford.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73-96). New York: Guilford.

Chen, S., Duckworth, K., & Chaiken, S. (1999). Motivated heuristic and systematic processing. *Psychological Inquiry, 10,* 44-49.

Coker, A. L., Davis, K. E., Arias, I., Desai, S., Sanderson, M., Brandt, H. M., et al., (2002). Physical and mental health effects of intimate partner violence for men and women. *American Journal of Preventive Medicine, 23,* 260-268.

Coluccia, E., Bianco, C., & Brandimonte, M. A. (2006). Dissociating veridicality, consistency, and confidence in autobiographical and event memories for the Columbia shuttle disaster. *Memory, 14,* 452-470.

Connecticut Department of Public Safety, Crimes Analysis Unit. (1999). *Family violence arrests, annual report.* Middletown, CT: Author.

Connolly, D. A., Price, H. L., Lavoie, J. A. A., & Gordon, H. M. (2008). Perceptions of children's reports of a unique event and an instance of a repeated event. *Law and Human Behavior, 32,* 92-112.

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.

Connolly, D. A., & Read, J. D. (2003). Remembering historical child sexual abuse. *Criminal Law Quarterly, 47,* 438-480.

Conrad, F. G., Brown, N. R., & Cashman, E. R. (1998). Strategies for estimating behavioural frequency in survey interviews. *Memory, 6,* 1998, 339-366.

Conway, M. A. (1996). Autobiographical knowledge and autobiographical memories. In D. C. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 67-93). New York: Cambridge University Press.

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1989). The eyewitness, the expert psychologist, and the jury. *Law and Human Behavior, 13,* 311-332.

Danis, F. S. (2003). The criminalization of domestic violence: What social workers need to know. *Social Work, 48,* 237-247.

Daudistel, H. C., Hosch, H. M., Holmes, M. D., & Graves, J. B. (1999). Effects of defendant ethnicity on juries' dispositions of felony cases. *Journal of Applied Social Psychology, 29,* 317-336.

Deaux, K., & Lafrance, M. (1998). Gender. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology, Vols. 1 and 2 (4th ed.)* (pp. 788-827). New York: McGraw-Hill.

Desmarais, S. L., Gibas, A., & Nicholls, T. L. (in press). Beyond violence against women: Gender-inclusiveness in domestic violence research, policy, and practice. In C. F. Ferguson (Ed.), *Violent crime: Clinical and social implications.* Thousand Oaks, CA: Sage.

Desmarais, S. L., Klein, C., Nicholls, T. L., Read, J. D., & Koch, W. J. (2006, March). *Battered women, battered thinking? Variables associated with cognitive distortions in women abused by intimate partners.* Poster presented at the American Psychology-Law Society conference, St. Petersburg, FL.

Desmarais, S. L., Nicholls, T. L., & Koch, W. J. (2005, Nov.). *Women abused by intimate partners: Prevalence of female-perpetrated aggression and psychiatric symptoms.* Paper presented at the Canadian Psychiatric Association conference, Vancouver, Canada.

Desmarais, S. L., & Yarmey, A. D. (2004). Judgements of deception and accuracy of performance in eyewitness testimony. *Canadian Journal of Police and Security Services, 2,* 13-22.

DiFranceisco, W., McAuliffe, T. L., & Sikkema, K. (1998). Influences of survey instrument format and social desirability on the reliability of self-reported high risk sexual behavior. *AIDS and Behavior, 2,* 329-337.

Dill, D. L., Chu, J. A., Grob, M. C., & Eisen, S. V. (1991). The reliability of abuse history reports: A comparison of two inquiry formats. *Comprehensive Psychiatry, 32,* 166-169.

Dutton, D. G., & Corvo, K. (2006). Transforming a flawed policy: A call to revive psychology and science in domestic violence research and practice. *Aggression and Violent Behavior, 11,* 457-483.

Dutton, D. G., & Nicholls, T. L. (2005). The gender paradigm in domestic violence research and theory: Phase 1 – The conflict of theory and data. *Aggression and Violent Behavior, 10,* 680-714.

Eagly, A., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin, 111,* 3-22.

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy, 38,* 319-345.

Elkins, T. J., Phillips, J. S., & Konopaske, R. (2002). Gender-related biases in evaluations of sex discrimination allegations: Is perceived threat the key? *Journal of Applied Psychology, 87,* 280-292.

Elkins, T. J., Phillips, J. S., Konopaske, R., & Townsend, J. (2001). Evaluating gender discrimination claims: Is there a gender similarity bias? *Sex Roles, 44,* 1-15.

Fals-Stewart, W. (2003). The occurrence of partner physical aggression on days of alcohol consumption: A longitudinal diary study. *Journal of Consulting and Clinical Psychology, 71,* 41-52.

Fals-Stewart, W., Birchler, G. R., & Kelley, M. L. (2003). The Timeline Followback Spousal Violence Interview to assess physical aggression between intimate partners: Reliability and validity. *Journal of Family Violence, 18,* 131-142.

Fals-Stewart, W., & Golden, J. (2002, September). *Use of diaries with the timeline followback spousal violence interview.* Poster presented at the 7th International Conference on Family Violence, San Diego, CA.

Fals-Stewart, W., O'Farrell, T. J., Freitas, T. T., McFarlin, S. K., & Rutigliano, P. (2000). The Timeline Followback reports of psychoactive substance use by drug-abusing patients: Psychometric properties. *Journal of Consulting and Clinical Psychology, 68,* 134-144.

Feather, N. T. (1996). Domestic violence, gender and perceptions of justice. *Sex Roles, 35,* 507-519.

Felson, R. B., Messner, S. F., & Hoskin, A. (1999). The victim-offender relationship and calling the police in assaults. *Criminology, 37,* 931-947.

Ferketich, S. (1991). Focus on psychometrics: Aspects of item analysis. *Research in Nursing & Health, 14,* 165-168.

Ferraro, K. J. (1989). Policing woman battering. *Social Problems, 36,* 61-74.

Finn, M. A., & Stalans, L. J. (1997). The influence of gender and mental state on police decisions in domestic assault cases. *Criminal Justice and Behavior, 24,* 157-176.

Fisher, R. P., & Cutler, B. L. (1995). The relation between consistency and accuracy of eyewitness testimony. In G. Davies, S. Lloyd-Bostock, M. McMurran, & C. Wilson (Eds.), *Psychology, law and criminal justice: International developments in research and practice* (pp. 21-28). Oxford: De Gruyter.

Fisher, R. P., McCauley, M. R., & Geiselman, R. E. (1994). Improving eyewitness testimony with the Cognitive Interview. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 245-269). New York: Cambridge University Press.

Fivush, R. (1993). Developmental perspectives on autobiographical recall. In G. S. Goodman & B. L. Bottoms, *Child victims, child witnesses: Understanding and improving testimony* (pp. 1-42). New York: Guildford Press.

Flin, R., Boon, J., Knox, A., & Bull, R. (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology, 83,* 323-336.

Florida Standard Jury Instructions in Criminal Cases. (1987).

Foa, E. B., Molnar, C., & Cashman, L. (1995). Change in rape narratives during exposure therapy for posttraumatic stress disorder. *Journal of Traumatic Stress, 8,* 675-690.

Follingstad, D., DeHart, D., & Green, E. (2004). Psychologists' judgments of psychologically aggressive actions when perpetrated by a husband versus a wife. *Violence and Victims, 19,* 435-452.

Follingstad, D. R., Rutledge, L. L., Berg, B. J., Hause, E. S., & Polek, D. S. (1990). The role of emotional abuse in physically abusive relationships. *Journal of Family Violence, 5,* 107-120.

Folsom, A. R., Jacobs Jr., D. R., Caspersen, C. J., Gomez-Marin, O., & Knudsen, J. (1986). Test-retest reliability of the Minnesota Leisure Time Physical Activity questionnaire. *Journal of Chronic Disease, 39,* 505-511.

Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology,* 5511, 115-1334.

Freyd, J. (1996). *Betrayal trauma: The logic of forgetting childhood abuse.* Cambridge, MA: Harvard University Press.

Geraerts, E., Kozarić-Kovačić, D., Merckelbach, H., Peraica, T., Jelicic, M., & Candel, I. (2007). Traumatic memories of war veterans: Not so special after all. *Consciousness and Cognition, 16,* 170-177.

Ghetti, S., Goodman, G. S., Eisen, M. L., Qin, J., & Davis, S. L. (2002). Consistency in children's report of sexual and physical abuse. *Child Abuse & Neglect, 26,* 977-995.

Giner-Sorolla, R., Chaiken, S., & Lutz, S. (2002). Validity beliefs and ideology can influence legal case judgments differently. *Law and Human Behavior, 26,* 507-526.

Glissan, J. L. (1991). *Cross-examination: Practice and procedure.* Sydney: Butterworths.

Goodman, L. A., Thompson, K. M., Weinfurt, K., Corl, S., Acker, P., Mueser, K. T., et al. (1999). Reliability of reports of violent victimization and posttraumatic among men and women with serious mental illness. *Journal of Traumatic Stress, 12,* 587-599.

Goldberg, C. (1999, November 23). Spouse abuse crackdown, surprisingly, nets many women. *The New York Times,* p. A16.

Golding, J. M., Sego, S. A., Sanchez, R. P., & Hasemann, D. (1995). The believability of repressed memories. *Law and Human Behavior, 19,* 596-592.

Greaves, C. L. (2005). *Phenomenological characteristics of autobiographical event memory.* Unpublished Master's thesis. Simon Fraser University, Burnaby, Canada.

Haber, R. N., & Haber, L. (2000). Experiencing, remembering and reporting events. *Psychology, Public Policy, and Law, 6,* 1057-1097.

Hamel, J., Desmarais, S. L., & Nicholls, T. L. (2007). Perceptions of motives in intimate partner violence: Expressive versus coercive violence. *Violence and Victims, 22,* 561-574.

Hartley, C. C. (2001). "He said, she said": The defense attack of credibility in domestic violence felony trials. *Violence Against Women, 7,* 510-544.

Herlihy, J., Scragg, P., & Turner, S. (2002). Discrepancies in autobiographical memories: Implications for the assessment of asylum seekers: Repeated interviews study. *British Medical Journal, 324,* 324-327.

Herman, J. L. (1997). *Trauma and recovery.* New York: Basic Books.

Herman, J. L., & Schatzow, E. (1987). Recovery and verification of memories of childhood sexual trauma. *Psychoanalytic Psychology, 4,* 1-14.

Hertel, P. (2004). Memory for emotional and nonemotional events in depression: A question of habit? In D. Reisberg & P. Hertel (Eds.), *Memory and emotion* (pp. 186-216). New York: Oxford University Press.

Hillier, L., & Foddy, M. (1993). The role of observer attitudes in judgments of blame in cases of wife assault. *Sex Roles, 29,* 629-644.

Hoctor, M. (1997). Domestic violence as a crime against the state. *California Law Review, 85,* 643.

Hoyle, C., & Sanders, A. (2000). Police response to domestic violence: From victim choice to victim empowerment? *British Journal of Criminology, 40*, 14-36.

Jenkins, P., & Davidson, B. (1990). Battered women in the criminal justice system: An analysis of gender stereotypes. *Behavioral Sciences and the Law, 8,* 161-170.

Jobe, J. B., Tourangeau, R., & Smith, A. F. (1993). Contributions of survey research to an understanding of memory. *Applied Cognitive Psychology, 7,* 567-584.

Johnson, J. D., Jackson, L. A., Gatto, L., & Nowack, A. (1995). Differential male and female responses to inadmissible sexual history information regarding a rape victim. *Basic and Applied Social Psychology, 16,* 503-513.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5,* 602-619.

Judicial Council of California Criminal Jury Instructions. (2007).

Kihlstrom, J. F. (1996). The trauma-memory argument and recovered memory therapy. In K. Pezdek & W. P. Banks (Eds.), *The recovered memory/false memory debate* (pp. 297-311). San Diego, CA: Academic Press.

Klevens, J., Shelley, G., Clavel-Arcas, C., Barney, D. D., Tobar, C., Duran, E. X., et al. (2007). Latinos' perspectives and experiences with intimate partner violence. *Violence Against Women, 13,* 141-158.

Kline, P. (1993). *The handbook of psychological testing.* London: Routledge.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts; Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123,* 297-315.

Koss, M. P., Figueredo, A. J., Bell, I., Tharan, M., & Tromp, S. (1996). Traumatic memory characteristics: A cross-validated mediational model of response to rape among employed women. *Journal of Abnormal Psychology, 105,* 421-432.

Krackow, E., & Lynn, S. J. (2003). Is there touch in the game of Twister? The effects of innocuous touch and suggestive questions on children's eyewitness memory. *Law and Human Behavior, 27,* 589-604.

Kraemer, H. C., Kiernan, M., Essex, M., & Kupfer, D. J. (2008). How and why criteria defining moderators and mediators differ between the Baron & Kenny and MacArthur approaches. *Health Psychology, 27,* S101-S108.

Kraemer, H. C., Stice, E., Kazdin, A., Offord, D., & Kupfer, D. (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry, 158,* 848-856.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediaotrs and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59,* 877-883.

Krinsley, K. E., Gallagher, J. G., Weathers, F. W., Kutter, C. J., & Kaloupek, D. G. (2003). Consistency of retrospective reporting about exposure to traumatic events. *Journal of Traumatic Stress, 16,* 399-409.

Krug, E., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (Eds.). (2002). *World report on violence and health.* Geneva: World Health Organization.

Laroche, D. (2005). *Aspects of the context and consequences of domestic violence: Situational couple violence and intimate terrorism in Canada in 1999.* Quebec, Canada: Institut de la statistique du Quebec. Retrieved June 5, 2005 from www.stat.gouv.qc.ca.

Lee, P. J., & Brown, N. R. (2003). Delay related changes in personal memories for September 11, 2001. *Applied Cognitive Psychology, 17,* 1007-1015.

Leippe, M. P., & Romanczyk, A. (1989). Reactions to child (versus adult) eyewitnesses: The influence of jurors' preconceptions and witness behaviour. *Law and Human Behavior, 13,* 103-132.

Lewin, C., Strand, S., & Belfrage, H. (2007, June). *Victims not participating in police investigations: The case of spousal assault.* Paper presented at the meetings of the International Association of Forensic Mental Health Services, Montreal, Canada.

Lindsay, R. C. L., Lim, R., Marando, L., & Cully, D. (1986). Mock-juror evaluations of eyewitness testimony: A test of metamemory hypotheses. *Journal of Applied Social Psychology, 16,* 447-459.

Lisak, D. (1994). The psychological impact of sexual abuse: Content analysis of interviews with male survivors. *Journal of Traumatic Stress, 7,* 525-548.

Locke, L. M., & Richman, C. L. (1999). Attitudes towards domestic violence: Race and gender issues. *Sex Roles, 40,* 227-247.

Loftus, E. F., Smith, K. D., Klinger, M. R., & Fiedler, J. (1992). Memory and mismemory for health events. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 102-137). New York: Sage.

Maccoby, E. E. (1988). Gender as a social category. *Developmental Psychology, 24,* 755-765.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51,* 93-120.

McKelvie, S. J. (2002). Effects of sex of judge and sex of victim on recommended punishment of a male murderer in a mock scenario. *Psychological Reports, 91,* 533-536.

McKeown, L., Noce, A., & Czerny, P. (2007, September). Factors associated with internet use: Does rurality matter? *Rural and Small Town Canada Analysis Bulletin, 7,* Catalogue no. 21-006-XIE. Retrieved December 22, 2007 from http://dsp-psd.pwgsc.gc.ca/collection_2007/statcan/21-006-X/21-006-XIE2007003.pdf.

McKimmie, B. M., Newton, C. J., Terry, D. J., & Schuller, R. A. (2004). Jurors' responses to expert witness testimony: The effects of gender stereotypes. *Group Processes & Intergroup Relations, 7,* 131-143.

McNally, R. J. (2003). *Remembering trauma.* Cambridge, MA: The Belknap Press of Harvard University Press.

Means, B., & Loftus, E. F. (1991). When personal history repeats itself: Decomposing memories for recurring events. *Applied Cognitive Psychology,* 5, 297-318.

Memon, A., Vrij, A., & Bull, R. (2003). *Psychology and Law: Truthfulness, accuracy and credibility* (2nd ed). Chichester, UK: John Wiley & Sons Ltd.

Miller, N. (2004). *Domestic violence legislation.* Retrieved August 28th, 2005 from www.ilj.org/dv/Papers/DV_Legislation-3.pdf

Mollica, R. F., Caridad, K. R., Massagli, M P. (2007). Longitudinal study of posttraumatic stress disorder, depression, and changes in traumatic memories over time in Bosnian refugees. *The Journal of Nervous and Mental Disease, 195,* 572-579.

Moretti, M. M., Catchpole, R. E. H., & Odgers, C. (2005). The dark side of girlhood: Recent trends, risk factors and trajectories to aggression and violence. *Canadian Child & Adolescent Psychiatry Review, 14,* 21–25.

Muller, R., Desmarais, S. L., & Hamel, J. (under review). *Do judicial responses to restraining order requests violate the Equal Protection Clause? Evidence for gender bias against male victims of domestic violence.*

New York Criminal Jury Instructions 2d. (2007).

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3rd ed). New York: McGraw-Hill.

Odgers, C. L., Moretti, M. M., & Reppucci, N. D. (2005). Examining the science and practice of violence risk assessment in female adolescents. *Law and Human Behavior, 29,* 7-27.

O'Hare, T. (1991). Measuring alcohol consumption: A comparison of the retrospective diary and the quantity-frequency methods in a college drinking survey. *Journal of Studies on Alcohol, 52,* 500-502.

O'Keefe, D. J. (2003). Message properties, mediating states, and manipulation checks: Claims, evidence, and data analysis in experimental persuasive message effects research. *Communication Theory, 13,* 251-274.

O'Neill, M. L., & Kerig, P. K. (2000). Attributions of self-blame and perceived control as moderators of adjustment in battered women. *Journal of Interpersonal Violence, 15,* 1036-1049.

Pape, K. T., & Arias, I. (2000). The role of perceptions and attributions in battered women's intentions to permanently end their violent relationships. *Cognitive Therapy and Research, 24,* 201-214.

Paradis, C. M., Solomon, L. Z., Florer, F., & Thompson, T. (2004). Flashbulb memories of personal events of 9/11 and the day after for a sample of New York City residents. *Psychological Reports, 95,* 304-310.

Parks, K. A., & Fals-Stewart, W. The temporal relationship between college women's alcohol consumption and victimization experiences. *Alcoholism: Clinical and Experimental Research, 28,* 625-629.

Peace, K. A., Porter, S., & ten Brinke, L. (2008). Are memories for sexually traumatic events "special"? A within-subjects investigation of trauma and memory in a clinical sample. *Memory, 16,* 10-21.

Peace, K. A., & Porter, S. (2004). A longitudinal investigation of the reliability of memories for trauma and other emotional experiences. *Applied Cognitive Psychology, 18,* 1143-1159.

Pearson, R. W., Ross, M., & Dawes, R. M. (1992). Personal recall and the limits of retrospective questions in surveys. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 65-101). New York: Sage.

Penrod, S. D., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy and Law, 1,* 817-845.

Peterson, C., Moores, L., & White, G. (2001). Recounting the same events again and again: Children's consistency across multiple interviews. *Applied Cognitive Psychology, 15,* 353-371.

Pezdek, K., & Taylor, J. (2002). Memory for traumatic events in children and adults. In M. L. Eisen, J. A., Quas, & G. S. Goodman (Eds.), *Memory and suggestibility in the forensic interview.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Porter, S., & Birt, A. R. (2001). Is traumatic memory *special*? A comparison of traumatic memory characteristics with memory for other emotional life experiences. *Applied Cognitive Psychology, 15,* S101-S1117.

Porter, S., & Peace, K. A. (2007). The scars of memory: A prospective, longitudinal investigation of the consistency of traumatic and positive emotional memories in adulthood. *Psychological Science, 18,* 435-441.

Potter, R., & Brewer, N. (1999). Perceptions of witness behaviour-accuracy relationships held by police, lawyers and jurors. *Psychiatry, Psychology and Law, 6,* 97-103.

*R. v. Burke*, [1996] 1 S.C.R. 474.

*R. v. Gagnon*, [2006] S.C.C. 17.

Ramsey-Klawsnik, H. (2003). Elder sexual abuse within the family. *Journal of Elder Abuse & Neglect, 15*, 43-58.

Read, J. D. (2001). Introduction: Trauma, stress, and autobiographical memory. *Applied Cognitive Psychology, 15,* S1-S5.

Read, J. D., & Connolly, D. A. (2007). The effects of delay on long-term memory for witnessed events. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *Handbook of eyewitness psychology, Volume 1: Memory for events* (pp. 117-155). Mahwah, NJ: Erlbaum

Read, J. D., & Desmarais, S. L. (in press). Lay knowledge of eyewitness issues: A Canadian survey. *Applied Cognitive Psychology.*

Read, J. D., & Lindsay, D. S. (2000). The amnestic consequences of summer camps and high school graduations: Effects of memory retrieval upon reported amnesia. *Journal of Traumatic Stress, 13,* 129-147.

Read, J. D., & Lindsay, D. S. (Eds.). (1997). *Recollections of trauma: Scientific evidence and clinical practice.* New York: Plenum Press.

Rebovich, D. J. (1996). Prosecution response to domestic violence: Results of a survey of large jurisdictions. In E. S. Buzawa & C. G. Buzawa (Eds.), *Do arrests and restraining orders work?* Thousand Oaks, CA: Sage.

Richardson, M. T., Leon, A. S., Jacobs Jr., D. R., Ainsworth, B. E., & Serfass, R. (1994). Comprehensive evaluation of the Minnesota Leisure Time Physical Activity questionnaire. *Journal of Clinical Epidemiology, 47,* 271-281.

Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroup and outgroups: A review and meta-analysis. *Personality and Social Psychology Review, 9,* 32-47.

Robinson, A. L., & Chandek, M. S. (2000). Philosophy into practice? Community policing units and domestic violence victim participation. *Policing, 23,* 280-302.

Room, R. (1990). Measuring alcohol consumption in the United States: Methods and rationales. In L. T. Kozlowski et al. (Eds.), *Research advances in alcohol and drug problems* (pp. 39-80). New York: Plenum Press.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103,* 734-760.

Russell, M., & Light, L. (2006). Police and victim perspectives on empowerment of domestic violence victims. *Police Quarterly, 9,* 375-396.

Ruva, C. L., & Bryant, J. B. (2004). The impact of age, speech style, and question form on perceptions of witness credibility and trial outcome. *Journal of Applied Social Psychology, 34,* 1919-1944.

Salazar, L. F., Baker, C. K., Price, A. W., & Carlin, K. (2003). Moving beyond the individual: Examining the effects of domestic violence policies on social norms. *American Journal of Community Psychology, 32,* 253-264.

Salhany, R. E. (1991). *Cross examination: The art of the advocate* (Revised Edition). Toronto, Canada: Butterworth.

Saywitz, K., & Camparo, L. (1998). Interviewing child witnesses: A developmental perspective. *Child Abuse & Neglect, 22,* 825-843.

Schafran, L. H. (1997). Credibility in the courts: Why is there a gender gap? In N. V. Benokraitis (Ed.), *Subtle sexism: Current practice and prospects for change* (pp. 216-225). Thousand Oaks, CA: Sage.

Scheppele, K. L. (1992). Just the facts, ma'am: Sexualized violence, evidentiary habits, and the revision of truth. *New York Law School Law Review, 37,* 123-172.

Schuller, R. A., & Hastings, P. A. (1996). Trials of battered women who kill: The impact of alternative forms of expert evidence. *Law and Human Behavior, 20,* 518-606.

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation, 22,* 127-160.

Shaw, J. S., McClure, K. A., & Dykstra, J. A. (2007). Eyewitness confidence from the witnessed event through trial. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *Handbook of eyewitness psychology, Volume 1: Memory for events* (pp. 371-397). Mahwah, NJ: Erlbaum.

Shobe, K. K., & Kihlstrom, J. F. (1997). Is traumatic memory special? *Current Directions in Psychological Science, 6,* 70-74.

Sixth Circuit Criminal Pattern Jury Instructions. (2005).

Smith, A. F., Jobe, J. B., & Mingay, D. J. (1991). Retrieval from memory of dietary information. *Applied Cognitive Psychology, 5,* 269-296.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models, *Sociological Methodology, 13,* 290-313.

Sobell, L. C., & Sobell, M. B. (1996). *Timeline FollowBack user's guide.* Toronto, Canada: Addiction Research Foundation.

Sobell, L. C., & Sobell, M. B. (1992). Timeline follow-back: A technique for assessing self-reported alcohol consumption. In R. Z. Litten & J. P. Allen (Eds.), *Measuring alcohol consumption: Psychosocial and biochemical methods* (p. 41-72). Totowa, NJ: Humana Press, Inc.

Sommers, S. P., & Ellsworth, P. C. (2000). Race in the courtroom: Perceptions of guilty and dispositional attributions. *Personality and Social Psychology Bulletin, 26,* 1367-1379.

Southwick, S. M., Morgan, C. A., Nocolaou, A. L., & Charney, D. S. (1997). Consistency of memory for combat-related traumatic events in veterans of Operation Desert Storm. *American Journal of Psychiatry, 154,* 173-177.

Spinhoven, P., Bean, T., & Eurelings-Bontekoe, L. (2006). Inconsistencies in the self-report of traumatic experiences by unaccompanied refugee minors. *Journal of Traumatic Stress, 19,* 663-673.

Stangor, C., Lynch, L., Duan, C., & Glass, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology, 62,* 207-218.

Stanko, E. A. (1982). Would you believe this woman? Prosecutorial screening for "credible" witnesses and a problem of justice. In N. H. Rafter & E. A. Stanko (Eds.), *Judge, lawyer, victim, thief: Women, gender roles, and criminal justice* (pp. 63-82). Boston: Northeastern University Press.

State of California. (1999). Report on arrest for domestic violence in California, 1998 (Office of the Attorney General). *Criminal Justice Statistics Center Report Series, 1*(3), 1-20.

Statistics Canada. (1993). *Violence Against Women Survey.* Ottawa, Canada: Canadian Centre for Justice Statistics.

Stets, J., & Straus, M. (1992). Gender differences in reporting marital violence. In M. Straus & R. Gelles, *Physical violence in American families* (pp. 151-166). New Brunswick, NJ: Transaction Publishers.

Stewart, A., & Maddren, K. (1997). Police officers' judgements of blame in family violence: The impact of gender and alcohol. *Sex Roles, 37,* 921-933.

Straus, M. A. (1990). The Conflict Tactics Scale and its critics: An evaluation and new data on validity and reliability. In M. A. Straus & R. J. Gelles (Eds.), *Physical violence in American families* (pp. 49-73). New Brunswick, NJ: Transaction Books.

Straus, M. A. (1993). Physical assaults by wives: A major social problem. In R. Gelles & D. Loseke (Eds.), *Current controversies on family violence* (pp. 67-80). Thousand Oaks, CA: Sage.

Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised Conflict Tactics Scale (CTS2): Development and preliminary psychometric data. *Journal of Family Issues, 17,* 283-316.

Straus, M. A., Hamby, S. L., & Warren, W. L. (2003). *The Conflict Tactics Scales handbook.* Los Angeles, CA: Western Psychological Services.

Stuesser, L. (1993). *An introduction to advocacy.* Sydney: The Law Book Company.

Tajfel, H., & Turner, J. C. (2004). The Social Identity Theory of intergroup behaviour. In J. T. Jost & J. Sidanius (Eds.), *Political psychology: Key readings* (pp. 276-293). New York: Psychology Press.

Tajfel, H., & Turner, J. (2001). An integrative theory of intergroup conflict. In M. A. Hogg & D. Abrams (Eds.), *Intergroup relations: Essential readings* (pp. 94-109). New York: Psychology Press.

Talarico, J. M., & Rubin, D. C. (2007). Flashbulb memories are special after all; in phenomenology, not accuracy. *Applied Cognitive Psychology, 21,* 557-578.

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science, 14,* 455-461.

Taylor, H. L., Jacobs, D. R. Jr, Schucker, B., Knudsen, J., Leon, A. S., & Debacker, G. (1978). A questionnaire for the assessment of leisure time physical activities. *Journal of Chronic Disease, 31,* 741–755.

Taylor, T. S., & Hosch, H. M. (2004). An examination of jury verdicts for evidence of a similarity-leniency effect, an out-group punitiveness effect or a black sheep effect. *Law and Human Behavior, 28,* 587-598.

Terr, L. (1994). *Unchained memories: True stories of traumatic memories, lost and found.* New York: Basic Books.

Thompson, C. P., & Mingay, D. (1991). Estimating the frequency of everyday events. *Applied Cognitive Psychology, 5,* 497-510.

Tjaden, P., & Thoennes, N. (2000). Prevalence and consequences of male-to-female and female-to-male intimate partner violence as measured by the National Violence Against Women Survey. *Violence Against Women, 6,* 142-161.

Ursel, J. E., & Brickey, S. (1996). The potential of legal reform reconsidered: An examination of Manitoba's zero-tolerance policy on family violence. In T. O'Reilly-Fleming (Ed.), *Post-critical criminology* (pp. 56-77). Scarborough, Canada: Prentice-Hall.

van der Kolk, B. A., & Fisler, R. (1995). Dissociation and the fragmentary nature of traumatic memories: Overview and exploratory study. *Journal of Traumatic Stress, 8,* 505-525.

van Giezen, A. E., Arensman, E., Spinhoven, P., & Wolters, G. (2005). Consistency in memory for emotionally arousing events: A review of prospective and experimental studies. *Clinical Psychology Review, 25,* 935-953.

van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology, 89,* 1008-1002.

Vendetta, K., Stappenbeck, C., & Fals-Stewart, W. (2004). *The timeline followback spousal violence interview: A user's guide.* Buffalo, NY: Addiction and Family Research Group.

Weaver, C. A., (1993). Do you need a "flash" to form a flashbulb memory? *Journal of Experimental Psychology: General, 122,* 39-46.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54,* 277-295.

Whitaker, D. J., Halleyesus, T., Swahn, & Saltzman, L. S. (2007). Differences in frequency of violence and reported injury between relationships with reciprocal and nonreciprocal intimate partner violence. *American Journal of Public Health, 97,* 941-947.

Whobrey, L., Sales, B. D., & Elwork, A. (1981). Witness credibility law. *Applied Social Psychology Annual, 2,* 189-210.

Wight, D., & West, P. (1999). Poor recall, misunderstandings and embarrassment: Interpreting discrepancies in young men's reported heterosexual behaviour. *Culture, Health & Sexuality, 1,* 55-78.

Worthen, J. B., & Varnado-Sullivan, P. (2005). Gender bias in attributions of responsibility for abuse. *Journal of Family Violence, 20,* 305-312.

Wright, D. B., & Loftus, E. F. (1998). How memory research can benefit from CASM. *Memory, 6,* 467-474.

Yuille, J. C., Hunter, R., Joffe, R., & Zaparniuk, J. (1993). Interviewing children in sexual abuse cases. In G. Goodman & B. Bottoms (Eds.), *Child victims, child witness: Understanding and improving children's testimony* (pp. 95-115). New York: The Guilford Press.

Zola, S. M. (1998). Memory, amnesia, and the issue of recovered memory: Neurobiological aspects. *Clinical Psychology Review, 18,* 915-832.

# APPENDICES

## *Appendix 1.1. Relationship Behaviours Questionnaire*

This survey involves recalling things that have happened in your life over the last year. We know that it can be challenging at times to remember specific events or behaviours in the past. To make it easier, we are providing you with an online calendar covering the last 12 months. You will type your answers in this calendar when appropriate (when instructed to do so). Because this survey requires using two internet windows at the same time, we strongly recommend closing all other windows and programs before beginning this survey.

Please start by indicating on the calendar in the other internet window any dates that are personally significant and easy for you to remember. Please include the categories of events listed below. You can include other categories of events as well (e.g., starting/losing jobs, weddings, deaths). You can use short forms or abbreviations to save time (e.g., birthdays = BDAY; vacations/holidays = VH). Please separate multiple responses for a single day with a semicolon (;) which is found next to the L key. Once you have finished, please confirm whether you have included these events and if not, explain the reason (e.g., no break-ups).

birthdays (BDAY)
vacations/holidays (VH)
relationship break-ups (RBU)
pay days (PD)
major community/social events (CE)
Other1
Other2
Other3

We are now going to ask you questions about the romantic relationship in which there was conflict in the past 12 months. If you have been in more than one such relationship (3 months in length or longer) in the past 12 months, please answer all subsequent questions with respect to the LONGEST relationship. Please indicate the date you and that partner started dating.

From the list below, please select the phrase that best describes the status or seriousness of that relationship.

Casual dating
Steady dating, but not living together
Living together, but not married

Married
Other

Are you and this partner still together?
Yes
No

Please indicate the date you broke up:

Do you have any children?

How many children do you have?

Do they reside with you?

*Category: Psychological Aggression*[19]
No matter how well a couple gets along, there are times when they disagree, get annoyed with the other person, want different things from each other, or just have spats or fights because they are in a bad mood, are tired, or for some other reason. Couples also have many different ways of trying to settle their differences and below is a list of behaviours that you or your partner might do when you have differences. Please read through this list of behaviours: -insulted or swore -shouted or yelled -stomped out of room -did something to spite or make partner mad -accused of being lousy lover -called fat or ugly (or other name) -destroyed something that belonged to partner -threatened to hit or throw something In the past 12 months, did your partner do any of these things to you during the course of a disagreement or conflict?
Yes
No

How confident are you in the accuracy of this answer?

Thinking back to those days that stand out, such as holidays, starting or losing a job, and important events with family or friends, please indicate on the calendar in the other internet window when one or more of these things happened. Please answer as accurately as possible, giving your best estimate when you can't remember precisely. Please mark EACH DAY on the calendar that your partner did that behaviour by typing the number that represents that behaviour (e.g., 1 = insulted or swore). If your partner did more than one of these types of behaviours on a given day, please write down the multiple numbers for that day, separating multiple responses for a single day with a semicolon (;) which is found next to the L key. Once you have finished, please confirm whether you have included these behaviours. Please only exclude behaviours that your partner did not engage in.
1 = insulted or swore
2 = shouted or yelled

[19] Italicized headings were not presented to participants.

3 = stomped out of room
4 = did something to spite partner or make partner mad
5 = accused of being lousy lover
6 = called fat or ugly (or some other name)
7 = destroyed something that belonged to partner
8 = threatened to hit or throw something

How confident are you in the accuracy of these dates?

*Category: Physical Assault*
Please read through this list of behaviours: -threw something that could hurt -twisted arm or hair -pushed or shoved -grabbed -slapped -used knife or gun -punched or hit with something that could hurt -choked -slammed against wall -beat up -burned or scalded on purpose -kicked In the past 12 months, did your partner do any of these things to you during the course of a disagreement or conflict?

Yes
No

How confident are you in the accuracy of this answer?

Thinking back to those days that stand out, such as holidays, starting or losing a job, and important events with family or friends, please indicate on the calendar in the other internet window when one or more of these things happened. Please mark EACH DAY on the calendar that your partner did that behaviour by typing the number that represents that behaviour (e.g., 17 = threw something that could hurt). If your partner did more than one of these types of behaviours on a given day, please write down the multiple numbers for that day, separating multiple responses for a single day with a semicolon (;) which is found next to the L key. Once you have finished, please confirm whether you have included these behaviours. Please only exclude behaviours that your partner did not engage in.

17 = threw something that could hurt
18 = twisted arm or hair
19 = pushed or shoved
20 = grabbed
21 = slapped
22 = used a knife or gun
23 = punched or hit with something that could hurt
24 = choked
25 = slammed against wall
26 = beat up
27 = burned or scalded on purpose
28 = kicked

How confident are you in the accuracy of these dates?

*Category: Sexual Coercion*
This is the last section that asks about relationship behaviours. Please read through this list of behaviours: -made partner have sex without a condom -insisted that we have sex (no physical force) -insisted that we have oral/anal sex (no physical force) -physically forced oral/anal sex -physically forced sex -used threats to force oral/anal sex -used threats to force sex In the past 12 months, did your partner do any of these things to you during the course of a disagreement or conflict?

Yes
No

How confident are you in the accuracy of this answer?

Thinking back to those days that stand out, such as holidays, starting or losing a job, and important events with family or friends, please indicate on the calendar in the other internet window when one or more of these things happened. Please mark EACH DAY on the calendar that your partner did that behaviour by typing the number that represents that behaviour (e.g., 41 = threw something that could hurt). If your partner did more than one of these types of behaviours on a given day, please write down the multiple numbers for that day, separating multiple responses for a single day with a semicolon (;) which is found next to the L key. Once you have finished, please confirm whether you have included these behaviours. Please only exclude behaviours that your partner did not engage in.

41 = made partner have sex without a condom
42 = insisted that we have sex (no physical force)
43 = insisted that we have oral/anal sex (no physical force)
44 = physically forced oral/anal sex
45 = physically forced sex
46 = used threats to force oral/anal sex
47 = used threats to force sex

How confident are you in the accuracy of these dates?

Please indicate your level of agreement with the following statement. Please respond as accurately and honestly as possible.

Overall, I am certain of the accuracy of my answers to the questions about relationship behaviours my partner engaged in.

## *Appendix 1.2. Leisure Activities Questionnaire*

This survey involves recalling things that have happened in your life over the last year. We know that it can be challenging at times to remember specific events or behaviours in the past. To make it easier, we are providing you with an online calendar covering the last 12 months. You will type your answers in this calendar when appropriate (when instructed to do so). Because this survey requires using two internet windows at the same time, we strongly recommend closing all other windows and programs before beginning this survey.

Please start by indicating on the calendar in the other internet window any dates that are personally significant and easy for you to remember. Please include the categories of events listed below. You can include other categories of events as well (e.g., starting/losing jobs, weddings, deaths). You can use short forms or abbreviations to save time (e.g., birthdays = BDAY; vacations/holidays = VH). Please separate multiple responses for a single day with a semicolon (;) which is found next to the L key. Once you have finished, please confirm whether you have included these events and if not, explain the reason (e.g., no break-ups).

birthdays (BDAY)
vacations/holidays (VH)
relationship break-ups (RBU)
pay days (PD)
major community/social events (CE)
Other1
Other2
Other3

*Category: Walking*[20]

We are now going to ask you questions about leisure activities (including sports, recreational, yard, and household work) that you performed during the last 12 months. Please answer as accurately as possible, giving your best estimate when you can't remember precisely. To help, you will again use the calendar provided. Please read through this list of behaviours: -walking for pleasure -cross country hiking -back packing -mountain climbing -bicycling for pleasure -dancing, lessons or classes -dancing, unstructured (e.g., at a club) -horseback riding In the past 12 months, did you do participate in any of these activities?

Yes
No

How confident are you in the accuracy of this answer?

[20] Italicized headings were not presented to participants.

Thinking back to those days that stand out, such as holidays, starting or losing a job, and important events with family or friends, please indicate on the calendar in the other internet window when one or more of these things happened. Please mark EACH DAY on the calendar that you participated in that activity by typing the number that represents that behaviour (e.g., 101 = walking for pleasure). If you did more than one of these types of activities on a given day, please write down the multiple numbers for that day, separating multiple responses for a single day with a semicolon (;) which is found next to the L key. Once you have finished, please confirm whether you have included these activities. Please only exclude activities that you did not participate in.

101 = walking for pleasure
102 = cross country hiking
103 = back packing
104 = mountain climbing
105 = bicycling for pleasure
106 = dancing, lessons or classes
107 = dancing, unstructured (e.g., at a club)
108 = horseback riding

How confident are you in the accuracy of these dates?

*Category: Conditioning Exercise*
Please read through this list of activities: -home exercise -health club exercise -jogging -running -weight lifting -yoga -pilates -stretching In the past 12 months, did you participate in any of these activities?

Yes
No

How confident are you in the accuracy of this answer?

Thinking back to those days that stand out, such as holidays, starting or losing a job, and important events with family or friends, please indicate on the calendar in the other internet window when one or more of these things happened. Please mark EACH DAY on the calendar that you participated in that activity by typing the number that represents that behaviour (e.g., 109 = home exercise). If you did more than one of these types of activities on a given day, please write down the multiple numbers for that day, separating multiple responses for a single day with a semicolon (;) which is found next to the L key. Once you have finished, please confirm whether you have included these activities. Please only exclude activities that you did not participate in.

109 = home exercise
110 = health club exercise
111 = jogging
112 = running
113 = weight lifting
114 = yoga
115 = pilates
116 = stretching

How confident are you in the accuracy of these dates?

*Category: Sports*
Please read through this list of activities: -bowling -volley ball -table tennis or ping pong -tennis, singles or doubles -softball -badminton -racket ball -basketball: non-game (e.g., free throwing, drills) -basketball: game play -football -squash -soccer In the past 12 months, did you participate in any of these activities?

Yes
No

How confident are you in the accuracy of this answer?

Thinking back to those days that stand out, such as holidays, starting or losing a job, and important events with family or friends, please indicate on the calendar in the other internet window when one or more of these things happened. Please mark EACH DAY on the calendar that you participated in that activity by typing the number that represents that behaviour (e.g., 117 = bowling). If you did more than one of these types of activities on a given day, please write down the multiple numbers for that day, separating multiple responses for a single day with a semicolon (;) which is found next to the L key. Once you have finished, please confirm whether you have included these activities. Please only exclude activities that you did not participate in.

117 = bowling
118 = volley ball
119 = table tennis or ping pong
120 = tennis, singles or doubles
121 = softball
122 = badminton
123 = racket ball
124 = basketball: non-game (e.g., free throwing, drills)
125 = basketball: game play
126 = football
127 = squash
128 = soccer

How confident are you in the accuracy of these dates?

*Category: Outdoor Activities* (Water & Winter activities combined)
This is the last section that asks about leisure activities. Please read through this list of activities: -water skiing or wake boarding -sailing -canoeing, rowing, or kayaking -swimming lengths in a pool -swimming at the beach -scuba diving -snorkelling -snow skiing or boarding, downhill -snow skiing, cross country -ice, roller, or inline skating -sledding or tobogganing -snow shoeing In the past 12 months, did you participate in any of these activities?

Yes
No

How confident are you in the accuracy of this answer?

Thinking back to those days that stand out, such as holidays, starting or losing a job, and important events with family or friends, please indicate on the calendar in the other internet window when one or more of these things happened. Please mark EACH DAY on the calendar that you participated in that activity by typing the number that represents that behaviour (e.g., 129 = water skiing or wake boarding). If you did more than one of these types of activities on a given day, please write down the multiple numbers for that day, separating multiple responses for a single day with a semicolon (;) which is found next to the L key. Once you have finished, please confirm whether you have included these activities. Please only exclude activities that you did not participate in.

129 = water skiing or wake boarding
130 = sailing
131 = canoeing, rowing, or kayaking
132 = swimming lengths in a pool
133 = swimming at the beach
134 = scuba diving
135 = snorkelling
136 = snow skiing or boarding, downhill
137 = snow skiing, cross country
138 = ice, roller, or inline skating
139 = sledding or tobogganing
140 = snow shoeing

How confident are you in the accuracy of these dates?

Please indicate your level of agreement with the following statement. Please respond as accurately and honestly as possible.

Overall, I am certain of the accuracy of my answers to the questions about relationship behaviours my partner engaged in.

## *Appendix 2.1. Intimate Partner Abuse Vignettes*

*Intimate Partner Abuse Vignette 1 – Consistent Version*

**Interview 1**

Interviewer: "Can you tell me what happened?"

Respondent: "The night started out fine. We were at a friend's house for a get together and everyone was having a great time. A few hours into the night, Joanne/Tom had had a bit too much to drink. I mean we both had been drinking, but you could tell Joanne/Tom had had too much. She/He saw me talking to a girl/guy from work and thought we were flirting. Joanne/Tom can get really jealous. She/He came over and told me it was time to leave. I could see how upset Joanne/Tom was, so I agreed even though I wanted to stay. We took a cab home and we didn't say one word to each other. It was probably at least 11pm when we got home. I could tell that Joanne/Tom was pretty angry. We were standing in the front entrance and she/he started yelling at me and accused me of cheating on her/him, which of course, wasn't true. That made me upset with Joanne/Tom for thinking that I would cheat on her/him because I just would never do anything like that. We were both getting more and more angry, yelling louder and louder. And then all of the sudden, she/he raised her/his hand and hit me hard across the face near my eye. I ended having a really bad black eye."

Interviewer: "And do you remember the date that this happened?"

Respondent: "I think it was the first Saturday in August, so August 2nd."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "No. That's about it."

Interviewer: "Finally, how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

**Interview 2**

Interviewer: "Thinking back to the event you described in your first interview six weeks ago, can you tell me what happened?"

Respondent: "A friend of ours was having a party, it was Saturday August 2nd, and a bunch of our friends were there. Everything seemed to be fine when all of the sudden Joanne/Tom started to get mad at me because I was talking to one of my coworkers that she/he didn't know. It was just someone that I worked with but Joanne/Tom is the jealous type, and thought I was flirting with this girl/guy, which wasn't the case at all. Anyway, we left the party early and grabbed a cab when we got outside. By the time we got home, Joanne/Tom was fuming mad, accusing me of cheating on her. That pissed me off and I started getting upset because I had been nothing but loyal to Joanne/Tom and so I starting yelling back. We were just standing there in the front entrance screaming at each other. Things just seemed to be getting worse and worse. That's when Joanne/Tom hit me hard, right on the bone near the eye. The bruise lasted for days."

Interviewer: "In your first interview you stated that the event occurred on August 2nd in the front entrance and today you again said that it occurred on August 2nd in the front entrance. Is that correct?"

Respondent: "Correct."

Interviewer: "And do you remember what time it was when you got home?"

Respondent: "I think it was about 11pm or so."

Interviewer: "Before you also said it was about 11pm."

Respondent: "Yes. That's right."

Interviewer: "Had you and Joanne/Tom had any alcohol to drink that night?"

Respondent: "Yes. We'd both been drinking."

Interviewer: "Last time you also suggested that Joanne/Tom had had too much to drink."

Respondent: "Yes. I remember. She/He had been drinking all night."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "No."

Interviewer: "Are you sure there is no more information you can provide about that event?"

Respondent: "Yes. I'm sure. I haven't thought of anything else."

Interviewer: "This is the last question - how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

*Intimate Partner Abuse Vignette 1 – Inconsistent Version*

**Interview 1**

Interviewer: "Can you tell me what happened?"

Respondent: "The night started out fine. We were at a friend's house for a get together and everyone was having a great time. A few hours into the night, Joanne/Tom had had a bit too much to drink. I mean we both had been drinking, but you could tell Joanne/Tom had had too much. She/He saw me talking to a girl/guy from work and thought we were flirting. Joanne/Tom can get really jealous. She/He came over and told me it was time to leave. I could see how upset Joanne/Tom was, so I agreed even though I wanted to stay. We took a cab home and we didn't say one word to each other. It was probably at least 11pm when we got home. I could tell that Joanne/Tom was pretty angry. We were standing in the front entrance and she/he started yelling at me and accused me of cheating on her/he, which of course, wasn't true. That made me upset with Joanne/Tom for thinking that I would cheat on her/he because I just would never do anything like that. We were both getting more and more angry, yelling louder and louder. And then all of the sudden, she/he raised her/his hand and hit me hard across the face near my eye. I ended having a really bad black eye."

Interviewer: "And do you remember the date that this happened?"

Respondent: "I think it was the first Saturday in August, so August 2nd."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "No. That's about it."

Interviewer: "Finally, how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

**Interview 2**

Interviewer: "Thinking back to the event you described in your first interview six weeks ago, can you tell me what happened?"

Respondent: "A friend of ours was having a party, it was the beginning of August, Saturday August 9th I think, and a bunch of our friends were there. Everything seemed to be fine when all of the sudden Joanne/Tom started to get mad at me because I was talking to one of my coworkers that she/he didn't know. It was just someone that I worked with but Joanne/Tom is the jealous type, and thought I was flirting with this girl/guy, which wasn't the case at all. Anyway, we left the party early and grabbed a cab when we got outside. By the time we got home, Joanne/Tom was fuming mad, accusing me of cheating on her/him. That pissed me off and I started getting upset because I had been nothing but loyal to Joanne/Tom and so I starting yelling back. We were just standing there in the kitchen screaming at each other. Things just seemed to be getting worse and worse. That's when Joanne/Tom hit me hard, right on the bone near the eye. The bruise lasted for days."

Interviewer: "In your first interview you stated that the event occurred on August 2nd in the front entrance and today you said that it occurred on August 9th in the kitchen. Which is correct?"

Respondent: "Sorry. That's right. We were in the front entrance and it was August 2nd."

Interviewer: "And do you remember what time it was when you got home?"

Respondent: "I think it was about 8pm."

Interviewer: "Before you said it was about 11pm. Which is it?"

Respondent: "It was later, about 11pm. I'm not sure why I just said that."

Interviewer: "Had you and Joanne/Tom had any alcohol to drink that night?"

Respondent: "No. Not really."

Interviewer: "Last time you suggested that Joanne/Tom had had too much to drink."

Respondent: "Yes. I remember now, she/he had been drinking all night."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "Actually, I remember hearing a noise outside – like someone was in our yard – and thinking that the neighbours were coming to see what was going on."

Interviewer: "Why didn't you provide this information in your first interview?"

Respondent: "I don't know. It just occurred to me now."

Interviewer: "This is the last question - how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

*Intimate Partner Abuse Vignette 2 – Consistent Version*

**Interview 1**

Interviewer: "Can you tell me what happened?"

Respondent: "It was the weekend, the middle of the afternoon - about 2pm I think. Susan/Jeff and I both had the weekend off, which is pretty rare. We were sitting at the table in the kitchen, going over the bills that we needed to pay in the next couple of weeks. We don't have a lot of money and we were both pretty stressed, worrying about how we were going to pay them all. Susan/Jeff started getting mad at me, telling me I need to spend less money and insisting that I should have to pay all of our bills for the month. I didn't think that was fair because it wasn't just what I spent that needed to be paid off and I told her/him that maybe we just need to figure out how to bring more money in. Susan/Jeff made some comment under her breath, which made me mad, so I started to scream at her/him. She/He was getting patronizing, telling me I needed a "time out." That made me even more angry so I swore at her/him. That's when she/he picked up her/his coffee cup and threw it across the table at me. I tried to turn away, but it hit me right on the side of the face and gave me a really bad bruise. I just couldn't believe she/he had actually thrown the cup even though she/he'd done something like that before."

Interviewer: "And do you remember the date that this happened?"

Respondent: "I think it was the first Sunday in November – November 6th."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "No. That's about it."

Interviewer: "Finally, how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

**Interview 2**

Interviewer: "Thinking back to the event you described in your first interview six weeks ago, can you tell me what happened?"

Respondent: "Susan/Jeff and I actually both had the day off, I think it was Sunday November 6th, and we were at the kitchen table sorting through bills. It was the beginning of the month, so we were figuring out what needed to be paid for the rest of the month. Money's tight so things were kind of tense. I had done some shopping in the previous month and Susan/Jeff was pissed at me for spending more money, which was fair I guess, but then she/he said that I should have to pay off all the bills. That's just ridiculous. So I started to get mad too and told her/him that maybe we just needed to make more money. Susan/Jeff just wouldn't let it go and mumbled something about me, like she/he was mocking me. I can't stand it when she/he does that so I started to yell at her/him. I remember she/he told me I needed a "time out" because she/he said I couldn't have an adult conversation and was acting like a kid. I finally told her/him to shut up, which obviously set her/him off, because the next thing I knew, she/he threw her coffee cup at me and it hit my cheekbone really hard. I ended up having a bruise for days."

Interviewer: "In your first interview you stated that the event occurred on November 6th in the kitchen and today you again said that it occurred on November 6th in the kitchen. Is that correct?"

Respondent: "Yes. That's right."

Interviewer: "And do you remember what time it was when this happened?"

Respondent: "I think it was 2pm or so."

Interviewer: "Before you also said it was about 2pm."

Respondent: "Correct."

Interviewer: "Had anything like this happened before?"

Respondent: "Yes."

Interviewer: "Last time you also suggested that this wasn't first time anything like this had happened."

Respondent: "Yes. That's correct. She/He's been violent like that a few times."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "No."

Interviewer: "Are you sure there is no more information you can provide about that event?"

Respondent: "Yes I'm sure. I don't remember anything else."

Interviewer: "This is the last question - how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

*Intimate Partner Abuse Vignette 2 – Inconsistent Version*

**Interview 1**

Interviewer: "Can you tell me what happened?"

Respondent: "It was the weekend, the middle of the afternoon - about 2pm I think. Susan/Jeff and I both had the weekend off, which is pretty rare. We were sitting at the table in the kitchen, going over the bills that we needed to pay in the next couple of weeks. We don't have a lot of money and we were both pretty stressed, worrying about how we were going to pay them all. Susan/Jeff started getting mad at me, telling me I need to spend less money and insisting that I should have to pay all of our bills for the month. I didn't think that was fair because it wasn't just what I spent that needed to be paid off and I told her/him that maybe we just need to figure out how to bring more money in. Susan/Jeff made some comment under her/his breath, which made me mad, so I started to scream at her/him. She/He was getting patronizing, telling me I needed a "time out." That made me even more angry so I swore at her/him. That's when she/he picked up her/his coffee cup and threw it across the table at me. I tried to turn away, but it hit me right on the side of the face and gave me a really bad bruise. I just couldn't believe she/he had actually thrown the cup even though she/he'd done something like that before."

Interviewer: "And do you remember the date that this happened?"

Respondent: "I think it was the first Sunday in November – November 6th."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "No. That's about it."

Interviewer: "Finally, how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

**Interview 2**

Interviewer: "Thinking back to the event you described in your first interview six weeks ago, can you tell me what happened?"

Respondent: "Susan/Jeff and I actually both had the day off, I think it was Sunday November 13th, and we were at the dining room table sorting bills. It was the beginning of the month, so we were figuring out what needed to be paid for the rest of the month. Money's tight so things were kinda tense. I had done some shopping and Susan/Jeff was pissed at me for spending more money, which was fair I guess, but then she/he said that I should have to pay off all the bills. That's just ridiculous. So I started to get mad too and told her that maybe we just needed to make more money. Susan/Jeff just wouldn't let it go and mumbled something, like she/he was mocking me. I can't stand it when she/he does that so I started to yell at her/him. I remember she/he told me I needed a "time out" because she/he said I couldn't have an adult conversation and was acting like a kid. I finally told her/him to shut up, which obviously set her/him off, because the next thing I knew, she/he threw her/his coffee cup at me and it hit my cheekbone really hard. I had a bruise for days."

Interviewer: "In your first interview you stated that the event occurred on November 6th in the kitchen and today you said that it occurred on November 13th in the dining room. Which is correct?"

Respondent: "It was November 6th and we were in the kitchen. I'm not sure why I just said that."

Interviewer: "And do you remember what time it was when this happened?"

Respondent: "I think it was 5pm or so."

Interviewer: "Before you said it was about 2pm."

Respondent: "Sorry. That's right. It was earlier, about 2pm."

Interviewer: "Had anything like this happened before?"

Respondent: "No. Nothing like that."

Interviewer: "Last time you suggested that this wasn't first time anything like this had happened."

Respondent: "That's actually correct. She/He has been violent like that a few times. I don't really like thinking about it too much."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "Yes. Actually, I remember that the phone rang right after but I didn't answer it because I was in shock."

Interviewer: "Why didn't you provide this information in your first interview?"

Respondent: "I'm not sure. I didn't remember until just now."

Interviewer: "This is the last question - how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

## *Appendix 2.2. Leisure Activities Vignettes*

*Leisure Activities Vignette 1 – Consistent Version*

**Interview 1**

Interviewer: "Can you tell me what happened?"

Respondent: "The game started out slow but things picked up at the end. We were at the diamond in the park playing softball – it was a group of my friends - and everyone was having a great time. A few hours into the night, Joanne/Tom, the pitcher, was getting tired. I mean everyone was getting tired, but you could tell Joanne/Tom was really getting worn out. She/He was the only one on their team that pitched during the whole game. We just needed one more run to win the game and I was up to bat. I stepped up to the plate and my whole team was cheering and clapping. I remember planting my feet and just trying to stay focused on hitting that ball. I swung and missed on the first pitch, but connected with the ball on the second one. I remember dropping the bat and started to run as fast as I could. Both teams were getting more and more excited, yelling louder and louder. After I came around the corner, I realized it might be close so I dove for second base. I hit the base hard right near my eye when I landed and I ended having a really bad black eye, but it was totally worth it because it was a great game. It probably was at least 11pm when the game was over."

Interviewer: "And do you remember the date that this happened?"

Respondent: "I think it was the first Saturday in August, so August 2nd."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "No. That's about it."

Interviewer: "Finally, how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

**Interview 2**

Interviewer: "Thinking back to the event you described in your first interview six weeks ago, can you tell me what happened?"

Respondent: "A friend of mine had organized a softball game at the diamond in the local park, it was the beginning of August, Saturday the 2nd I think, and a bunch of my friends were playing. The game was a bit slow to start but by the end we had a good match going. The teams were tied and it was my turn to hit. I went up to bat and remember hearing my team cheering me on. I looked over at Joanne/Tom, who was pitching, and I remember trying to focus on what I needed to do because we just needed one run to win the game. I didn't connect on the first pitch and realized the pressure was on. I hit the ball on the second pitch and the cheering was getting louder and louder. I took off for first base, dropping the baseball bat when I started to run. I decided to try to make it second base. I didn't know if I was going to make it and actually dove for the base. I fell hard and cracked my cheek on the base, right on the bone near my eye. The bruise lasted for days."

Interviewer: "In your first interview you stated that the event occurred on August 2nd in the front entrance and today you again said that it occurred on August 2nd in the front entrance. Is that correct?"

Respondent: "Correct."

Interviewer: "And do you remember what time it was when you got home?"

Respondent: "I think it was about 11pm."

Interviewer: "Before you also said it was about 11pm."

Respondent: "Yes. That's right."

Interviewer: "Was everyone tired by this point in the game?"

Respondent: "Yes. The game had been going on for a while."

Interviewer: "Last time you also suggested that the pitcher was especially worn out."

Respondent: "Yes. I remember. No one else pitched for their team."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "No."

Interviewer: "Are you sure there is no more information you can provide about that event?"

Respondent: "Yes. I'm sure. I haven't thought of anything else."

Interviewer: "This is the last question - how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

*Leisure Activities Vignette 1 – Inconsistent Version*

**Interview 1**

Interviewer: "Can you tell me what happened?"

Respondent: "The game started out slow but things picked up at the end. We were at the diamond in the park playing softball – it was a group of my friends - and everyone was having a great time. A few hours into the night, Joanne/Tom, the pitcher, was getting tired. I mean everyone was getting tired, but you could tell Joanne/Tom was really getting worn out. She/He was the only one on their team that pitched during the whole game. We just needed one more run to win the game and I was up to bat. I stepped up to the plate and my whole team was cheering and clapping. I remember planting my feet and just trying to stay focused on hitting that ball. I swung and missed on the first pitch, but connected with the ball on the second one. I remember dropping the bat and started to run as fast as I could. Both teams were getting more and more excited, yelling louder and louder. After I came around the corner, I realized it might be close so I dove for second base. I hit the base hard right near my eye when I landed and I ended having a really bad black eye, but it was totally worth it because it was a great game. It probably was at least 11pm when the game was over."

Interviewer: "And do you remember the date that this happened?"

Respondent: "I think it was the first Saturday in August, so August 2nd."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "No. That's about it."

Interviewer: "Finally, how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

**Interview 2**

Interviewer: "Thinking back to the event you described in your first interview six weeks ago, can you tell me what happened?"

Respondent: "A friend of mine had organized a softball game at the diamond at the local school, it was the beginning of August, Saturday the 9th I think, and a bunch of my friends were playing. By the end of the game we had a good match going. The teams were tied and it was my turn to hit. I went up to bat and remember hearing my team cheering me on. I looked over at Joanne/Tom, who was pitching, and I remember trying to focus on what I needed to do because we just needed one run to win the game. We just needed one run to win the game. I didn't connect on the first pitch and realized the pressure was on. I hit the ball on the second pitch and the cheering was getting louder and louder. I took off for first base, dropping the baseball bat when I started to run. I decided to try to make it second base. I didn't know if I was going to make it and actually dove for the base. I landed hard and cracked my cheek on the base, right on the bone near my eye. The bruise lasted for days."

Interviewer: "In your first interview you stated that the event occurred on August 2nd at the diamond in the park and today you said that it occurred on August 9th at the diamond at the local school. Which is correct?"

Respondent: "Sorry. That's right. We were in the park and it was August 2nd."

Interviewer: "And do you remember what time it was when you got home?"

Respondent: "I think it was about 8pm."

Interviewer: "Before you said it was about 11pm. Which is it?"

Respondent: "It was later, about 11pm. I'm not sure why I just said that."

Interviewer: "Was everyone tired by this point in the game?"

Respondent: "No. I don't think so."

Interviewer: "Last time you suggested that everyone was tired and that the pitcher was especially worn out."

Respondent: "Yes. I remember now, she/he'd been pitching all night."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "Yes. Actually, I remember seeing the pitcher fumble with the ball after I passed first base and that's when I decided to try to make it to second."

Interviewer: "Why didn't you provide this information in your first interview?"

Respondent: "I don't know. It just occurred to me now."

Interviewer: "This is the last question - how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

*Leisure Activities Vignette 2 – Consistent Version*

**Interview 1**

Interviewer: "Can you tell me what happened?"

Respondent: "It was the weekend, the middle of the afternoon - about 2pm I think. Susan/Jeff and I both had the weekend off, which is pretty rare. We decided to go to the rec centre for a work-out. It was raining out that day, otherwise I think I would've just gone for a run or something outside to take advantage of the weather. Anyway, when we got to there, we stretched a bit together, but then split up because we wanted to use different equipment. I started on the bike and did that for about 20 minutes or so to warm up. I was starting to get pretty warm, so I dropped my sweatshirt off in the locker room and grabbed a drink at the water fountain. Then I went to the weight area to do some strength training. I did leg and arm exercises for the next half hour. I started with some exercises on the machines, but spent most of my time using some free weights and the bench. I remember getting annoyed because the guy next to me on the machines was singing along with the music he was listening to. I think we had been there for about an hour when I was about ready to go. Even though I was pretty tired by the time I got home, it was totally worth it. I felt really good about going, because I hadn't done any exercise in a while."

Interviewer: "And do you remember the date that this happened?"

Respondent: "I think it was the first Sunday in November – November 6th."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "No. That's about it."

Interviewer: "Finally, how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

**Interview 2**

Interviewer: "Thinking back to the event you described in your first interview six weeks ago, can you tell me what happened?"

Respondent: "Susan/Jeff and I actually both had the day off, I think it was Sunday November 6th. I remember that we wanted to get some exercise, maybe go for a run or something, but it was raining outside so we decided to go to the rec centre. We started with some stretching together when we first got there, but then we each did our own thing the rest of the time. First, I did some cardio on the bike for about 20 minutes because I wanted to get warmed up. It worked, because I remember taking a quick break to get some water and put my sweatshirt in the locker. I moved on to strength training for my arms and legs for about 30 minutes or so – I think we were there for about an hour total. The weight area had machines, which I did first, and also an area with free weights and benches. I didn't stay with the machines for very long though because there was this guy singing out loud to whatever music he was listening to, which was irritating. I remember being beat afterwards when I was done, but feeling good about myself because I had gotten some exercise.

Interviewer: "In your first interview you stated that the event occurred on November 6th at the rec centre and today you again said that it occurred on November 6th at the rec centre. Is that correct?"

Respondent: "Yes. That's right."

Interviewer: "And do you remember what time it was when this happened?"

Respondent: "I think it was 2pm or so."

Interviewer: "Before you also said it was about 2pm."

Respondent: "Correct."

Interviewer: "Had you been exercising regularly before that?"

Respondent: "No. Not really."

Interviewer: "Last time you also suggested that you hadn't been exercising regularly before that."

Respondent: "Yes. That's correct. It had been a while since I had been working out regularly."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "No."

Interviewer: "Are you sure there is no more information you can provide about that event?"

Respondent: "Yes I'm sure. I don't remember anything else."

Interviewer: "This is the last question - how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

*Leisure Activities Vignette 2 – Inconsistent Version*

**Interview 1**

Interviewer: "Can you tell me what happened?"

Respondent: "It was the weekend, the middle of the afternoon - about 2pm I think. Susan/Jeff and I both had the weekend off, which is pretty rare. We decided to go to the rec centre for a work-out. It was raining out that day, otherwise I think I would've just gone for a run or something outside to take advantage of the weather. Anyway, when we got to there, we stretched a bit together, but then split up because we wanted to use different equipment. I started on the bike and did that for about 20 minutes or so to warm up. I was starting to get pretty warm, so I dropped my sweatshirt off in the locker room and grabbed a drink at the water fountain. Then I went to the weight area to do some strength training. I did leg and arm exercises for the next half hour. I started with some exercises on the machines, but spent most of my time using some free weights and the bench. I remember getting annoyed because the guy next to me on the machines was singing along with the music he was listening to. I think we had been there for about an hour when I was about ready to go. Even though I was pretty tired by the time I got home, it was totally worth it. I felt really good about going, because I hadn't done any exercise in a while."

Interviewer: "And do you remember the date that this happened?"

Respondent: "I think it was the first Sunday in November – November 6th."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "No. That's about it."

Interviewer: "Finally, how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

**Interview 2**

Interviewer: "Thinking back to the event you described in your first interview six weeks ago, can you tell me what happened?"

Respondent: "Susan/Jeff and I actually both had the day off, I think it was Sunday November 13th. I remember that we wanted to get some exercise, maybe go for a run or something, but it was raining outside so we decided to go to the gym. We started with some stretching together when we first got there, but then we each did our own thing the rest of the time. First, I did some cardio on the bike for about 20 minutes because I wanted to get warmed up. It worked, because I remember taking a quick break to get some water and put my sweatshirt in the locker. I moved on to strength training for my arms and legs for about 30 minutes or so – I think we were there for about an hour total. The weight area had machines, which I did first, and also an area with free weights and benches. I didn't stay with the machines for very long though because there was this guy singing out loud to whatever music he was listening to, which was irritating. I remember being beat afterwards when I was done, but feeling good about myself because I had gotten some exercise.

Interviewer: "In your first interview you stated that the event occurred on November 6th at the rec centre and today you said that it occurred on November 13th at the gym. Which is correct?"

Respondent: "It was November 6th and we were at the rec centre. I'm not sure why I just said that."

Interviewer: "And do you remember what time it was when this happened?"

Respondent: "I think it was 5pm or so."

Interviewer: "Before you said it was about 2pm."

Respondent: "Sorry. That's right. It was earlier, about 2pm."

Interviewer: "Had you been exercising regularly before that?"

Respondent: "Yes. Fairly regularly."

Interviewer: "Last time you suggested that you hadn't been exercising regularly before that."

Respondent: "That's actually correct. Come to think of it, it had been a while since I had been working out regularly."

Interviewer: "Is there anything else that you remember about what happened?"

Respondent: "Yes. Actually, I remember that my cell phone rang while I was at the gym, but I didn't pick it up because I just wanted to get the workout done."

Interviewer: "Why didn't you provide this information in your first interview?"

Respondent: "I'm not sure. I didn't remember until just now."

Interviewer: "This is the last question - how confident are you in the accuracy of your report?"

Respondent: "Pretty confident."

## *Appendix 2.3. Complainant Effectiveness Questionnaire*

**Instructions: Please answer all of the following questions about the respondent based on the report you just read by circling a number between 1 and 6.**

1. How **intelligent** do you think this person was?
   Not at all intelligent 1 2 3 4 5 6 Very intelligent

2. How **honest** or **truthful** do you think this person was?
   Not at all honest 1 2 3 4 5 6 Very honest

3. How **accurately** do you think the person recalled the events?
   Not at all accurate 1 2 3 4 5 6 Very accurate

4. How **similar** was this person to you?
   Not at all similar 1 2 3 4 5 6 Very similar

5. How **suggestible** or **easily influenced** do you think this person was by (mis)leading questions?
   Not at all suggestible 1 2 3 4 5 6 Very suggestible

6. How **consistent** do you think this person was?
   Not at all consistent 1 2 3 4 5 6 Very consistent

7. How **confident** do you think this person was?
   Not at all confident 1 2 3 4 5 6 Very confident

8. How **likeable** do you think this person was?
   Not at all likeable 1 2 3 4 5 6 Very likeable

9. Imagine the person belonged to a group of people similar to him(her)self. How closely would you see yourself as **belonging** to that same **group**?
   Not at all close 1 2 3 4 5 6 Very close

10. What is the **severity** or **intensity** of the behaviours reported?
    Not at all severe 1 2 3 4 5 6 Very severe

11. Overall, how **credible** or **believable** do you think this person was?
    Not at all credible 1 2 3 4 5 6 Very credible