AMERICAN PSYCHOLOGICAL ASSOCIATION
American Psychology-Law Society
Law and Human Behavior

ISSN: 0147-7307
2025, Vol. 49, No. 5, 473–489
https://doi.org/10.1037/lhb0000618

Differential Impacts of Investigation and Witness Delays on Layperson Perceptions of Witness Credibility

Madison B. Harvey[1], Heather L. Price[2], and Deborah A. Connolly[1]
[1] Department of Psychology, Simon Fraser University
[2] Department of Psychology, Thompson Rivers University

Objective: Prior research has reported a consistent negative impact of delayed disclosure on sexual assault complainants' perceived credibility. The present studies investigated the impact of delay in new contexts. ***Hypotheses:*** Only when a delay was a result of a witness's actions would it negatively impact credibility. Any delay would decrease convictions and guilt likelihood ratings. ***Method:*** Study 1a ($N = 412$; $M_{age} = 20.13$ years; 60.00% women; 31.70% Caucasian; 3 [delay: 1 day, 2 years, 15 years] × 4 [report consistency: consistent, contradiction, omission, reminiscence]) examined the impact of an investigation delay and report inconsistencies on witness credibility in a physical assault case. Participants watched a witness's report, rated their credibility, and made legal decisions (e.g., verdict). Study 1b ($N = 240$; $M_{age} = 19.67$ years; 71.55% women; 28.75% Caucasian; 2 [delay: 1 day, 2 years] × 3 [report inconsistency: contradiction, omission, reminiscence]) partially replicated Study 1a. Study 2 ($N = 200$; $M_{age} = 33.94$ years; 45.50% women; 65.48% Caucasian; no reason–no delay, no reason-delay, witness-delay, investigation-delay) and Study 3 ($N = 363$; $M_{age} = 34.65$ years; 50.14% women; 53.17% Caucasian; 2 [delay: immediate, 2 years] × 3 [contextual information: no details, witness, investigation]) explored how the reason for the delay (i.e., witness or investigation) impacted credibility and legal decisions. ***Results:*** Studies 1a and 1b found no impact of delay on witness credibility, but a delay resulted in reduced likelihood of guilt in Study 1a. Study 2 showed that only a witness-delay negatively impacted credibility. Any explained delay reduced guilt ratings and, for an investigation-delay, convictions, relative to no delay. Study 3 demonstrated only an effect of contextual information on honesty, suggesting negative views of witness reluctance at any delay. Any delay led to fewer convictions. ***Conclusion:*** Witness behavior, regardless of delay, may negatively impact their perceived credibility, whereas a delay outside of the witness's control might not. Delays also impact legal decisions. This work provides important new insight into how delays may be treated in the justice system.

Public Significance Statement
Delays are frequent in the justice system, but it might be that only when a delay is a result of a witness's actions does the delay negatively impact their perceived credibility. Delays as a result of police investigations may not negatively impact the perceived credibility of a witness in the same way as delays resulting from a witness's actions. The present findings suggest that a delay that is a result of a witness's actions may bias decision makers (e.g., judges, juries) against them. Safeguards may be necessary to help counteract such bias. Further, the present work offers insight for future researchers to build on by challenging previously held assumptions about delays (i.e., that delays always negatively impact credibility) and highlighting instances in which delays may not impact credibility and decision making.

Keywords: perceived credibility, delay, witness, legal decisions

Supplemental materials: https://doi.org/10.1037/lhb0000618.supp

Bradley D. McAuliff served as Action Editor.
Madison B. Harvey https://orcid.org/0000-0002-9736-6230
Heather L. Price https://orcid.org/0000-0001-6109-6198
Deborah A. Connolly https://orcid.org/0000-0002-2276-7646
The authors have no conflicts of interest to disclose. This work was supported by a Social Sciences and Humanities Research Council of Canada Joseph-Armand Bombardier Canada Graduate Scholarship awarded to Madison B. Harvey and a Social Sciences and Humanities Research Council of Canada Insight Grant (435-2014-0916) awarded to Deborah A. Connolly. Heather L. Price was also supported by the Canada Research Chairs program. This work was supported by Internal Funds awarded to Heather L. Price from Thompson Rivers University. A portion of this work has been submitted for the fulfillment of degree requirements for Madison B. Harvey. A portion of this work was also presented at the American Psychology-Law Society Annual Conference.

continued

Witness testimony after a long delay is possible, as many jurisdictions around the world do not have a statute of limitations on most crimes (Connolly et al., 2017). In delayed cases, a witness's testimony may be more likely than in timely cases to be the only evidence presented at trial because of the loss of corroborating evidence over time (e.g., biological evidence, video evidence). When a report is delayed, there are factors that can influence the perceived credibility of a witness, such as reasons for the delay. Investigating how factors that contribute to delayed reports influence a witness's perceived credibility is essential in furthering our knowledge of how such cases may be perceived in the justice system.

There are several reasons that a case may not be prosecuted until after a long delay, including delays in reporting an offense and delays in prosecuting a reported offense. Delayed reporting of some types of offenses is very common; many victims of sexual assault do not report until long after the alleged crime occurred, if at all (London et al., 2008). A review of adults reporting child sexual assault in Ireland revealed that a majority of cases were reported 11 or more years after the crime occurred (Bunting, 2014). Similarly, Monroe et al. (2005) found that more than half of respondents who had experienced sexual assault as an adult waited years before disclosing. Other crimes may also lead to a delayed disclosure from a witness or victim. Domestic violence victims may delay reporting their experience, with one survey finding that only 62.8% of victims had previously reported a violent incident (Birdsey & Snowball, 2013). Similarly, only approximately one quarter of fraud or attempted fraud victims reported the crime to authorities or other administrative entities (e.g., Better Business Bureau; Mason & Benson, 1996). Witnesses of gang-related criminal activity may also delay disclosing (Whitman & Davis, 2007).

Reasons for delayed disclosure of a crime are varied, including feelings of responsibility or shame (Reich et al., 2022), fear of negative reaction or failure to believe the report (Reich et al., 2022), social stigma (Tsang et al., 2021), fear of the perpetrator (Mason & Benson, 1996; Reich et al., 2022), a belief that the crime was not serious enough to warrant a report (Birdsey & Snowball, 2013), or not realizing that a crime even took place (Mason & Benson, 1996). Although a central aim of most criminal justice systems is to prosecute cases in a timely manner (see *R. v. Jordan*, 2016), crimes may not be prosecuted until after a delay because of factors inherent in the justice system (Euale & Turtle, 1999). These factors include heavy caseloads, inadequate resources, complex cases that require additional investigative time (Standing Senate Committee on Legal and Constitutional Affairs, 2017), and failure to identify a suspect.

When evaluating the impact of delay on perceived credibility, we may focus on perceptions of cognitive competence, honesty, and suggestibility. Ross et al. (2003) suggested that cognitive competence and honesty are two primary factors contributing to perceptions of credibility. Cognitive competence references a witness's ability to understand and accurately remember an event and to understand and appropriately respond to questions (Ross et al., 2003). Honesty refers to the likelihood that an individual believes they are telling the truth (Ross et al., 2003). Kehn et al. (2016) have proposed a three-factor model of children's credibility, suggesting that evaluations of credibility rely not only on perceived honesty and cognitive ability but also on perceived suggestibility—that is, how open a witness is to suggestion from outside sources such as improper questioning techniques and how likely the outside source can influence a witness's report. Of course, adults can also be suggestible (e.g., Eisen & Lynn, 2001); thus, we focus on this three-factor model in the current work.

Research on the effect of delay on perceived credibility and verdicts has primarily been conducted in instances of delayed disclosure of sexual assault (e.g., Balogh et al., 2003; Ellison & Munro, 2009; Franiuk et al., 2020). Generally, this research demonstrates that, when compared with an immediate report, a delayed report decreases a witness's perceived credibility and decreases guilty verdicts (Balogh et al., 2003; Ellison & Munro, 2009; Franiuk et al., 2020; Miller et al., 2022; Pozzulo et al., 2010; Thompson et al., 2021). The impact of a delay on credibility and verdict decisions has been investigated across a variety of intervals from less than a year (e.g., Ellison & Munro, 2009; Miller et al., 2022) to up to 35 years (e.g., Fraser et al., 2022). However, this research has focused on instances of delayed disclosure of sexual assault and sexual misconduct; to our knowledge, there has been no investigation into delays of nonsexual assault cases with adults (e.g., theft; see Danby et al., 2022, for an investigation into delayed disclosure of child physical abuse) or when the delay is a consequence of factors inherent in the justice system. This is an important omission because delays occur in many types of criminal cases, and there are many reasons that a criminal case may not come to trial in a timely manner. Because the extant literature has focused on sex-based cases when the delay was the result of witness reluctance, conclusions about the impact of delay on witness credibility more broadly cannot be drawn. It might be that delays caused by factors other than witness reluctance have a different impact on witness credibility.

It is also important to understand how other investigation features might interact with delay. One common outcome of a delay is that witnesses may be interviewed multiple times. These repeated interviews can result in inconsistencies reported across a witness's statements, including contradictions, reminiscences, and omissions. Contradictions are details that directly conflict with previous details reported by the witness (Krix et al., 2015). Witnesses who provide some contradictory details may not be inaccurate in all, or even most, other details reported about the event (Brewer et al., 1999). Reminiscent details are details that an individual did not report initially but did report at a later date (Krix et al., 2015). Reminiscence is common, and such details are often accurate (Gilbert & Fisher, 2006; Krix et al., 2015). Details that were reported in an early interview but not in a later interview are known as omissions. As the delay between interviews increases, more omissions are expected (Peace et al., 2015), and omissions may occur more frequently with peripheral details than central details (see Peterson & Whalen, 2001; Wells et al., 2014). Omissions may be a result of forgetting and/or

Madison B. Harvey played a lead role in formal analysis and writing–original draft and an equal role in conceptualization, data curation, investigation, methodology, and writing–review and editing. Heather L. Price played an equal role in conceptualization, methodology, supervision, and writing–review and editing. Deborah A. Connolly played an equal role in conceptualization, methodology, supervision, and writing–review and editing.

The preregistered design is available at https://osf.io/jqrh7/.

Correspondence concerning this article should be addressed to Heather L. Price, Department of Psychology, Thompson Rivers University, 805 TRU Way, Kamloops, BC V2C 0C8, Canada. Email: hprice@tru.ca

social reasons, such as a witness's belief that the interviewer already knows the information (Orbach et al., 2012).

Despite their low diagnostic value, inconsistencies in a witness's testimony are often seen as indicative of low credibility (Brewer et al., 1999). In a mock juror study, Berman et al. (1995) demonstrated that a single inconsistency on a central or peripheral detail diminished the witness's credibility. Inconsistencies are so well known to be problematic for witness credibility that legal professionals have been advised to attempt to elicit inconsistent answers from a witness in order to discredit them (Glissan, 1991). Inconsistencies in witness testimony may also impact verdict decisions. Berman and Cutler (1996) compared perceptions of consistent testimony with contradictory and reminiscent testimony in a mock juror study. Participants in the consistent witness condition were more likely to convict (69%) than those in the reminiscent (37%) or contradictory (20%) conditions. These results suggest that, although any form of inconsistency can lower conviction rates, contradictions might be the most damaging.

Potential jurors may not have the knowledge and contextual understanding to evaluate memory-based testimony. Schmechel et al. (2006) surveyed mock jurors and found that a large percentage had misunderstandings about the way memory works. Many laypeople believed that memory works much like a camera (52%), and a witness could easily play back a memory in their mind (46%; see also Simons & Chabris, 2011). Simons and Chabris (2011) found that a portion of potential jurors believed that memories will not change after initial encoding (48%). Further, although the forgetting curve has been established in research (Murre & Dros, 2015), it appears that many potential jurors are unaware of its effect. In one survey of more than 100 jurors, only roughly 30% agreed with the statement, "The rate of memory loss for an event is greatest right after the event" (Benton et al., 2006). Laypeople also appear to struggle to accurately estimate how memory may decay over time. Oeberst (2012) compared the results of a memory task with participants' estimations of how peers would perform on the task. Participants were given a brief description of the task (i.e., recalling information after an hour, day, and week) and sample graphs of how memory might decay. Participants believed that memory performance would drop drastically over a period of time (up to a week), whereas actual memory performance stayed significantly higher than estimations.

The Present Studies

The present studies explored perceptions of witnesses after varied delays for a nonsexual crime (i.e., assault and theft). Studies 1a and 1b focused on a delay caused by the investigation. Study 2 compared investigation and witness reasons for a delay. Study 3 further compared investigation and witness reasons for a delay in a fully crossed design.

Study 1a

This was a 3 (delay: 1 day, 2 years, 15 years) × 4 (report consistency: consistent, contradiction, reminiscence, omission) between-subjects design. Participants watched a portion of a mock interview between a lawyer and a witness to an assault and theft (both actors) that had occurred 1 day, 2 years, or 15 years earlier. The witness was initially interviewed by the police immediately after the crime; this initial interview was not shown to participants but was referenced in the interview. In the delay conditions (i.e., 2 years or 15 years), the case was described as delayed because no suspect was identified. In all conditions, the witness's testimony varied in terms of consistency; the witness was consistent or had contradictions, reminiscences, or omissions. Participants rated the witness's cognitive competence, honesty, and suggestibility on the basis of a three-factor model of credibility (Kehn et al., 2016). Participants then took the role of a mock juror and made final legal decisions (e.g., dichotomous verdict, likelihood of suspect guilt).

We hypothesized the following:

Hypothesis 1: Ratings of perceived credibility (i.e., cognitive competence, honesty, suggestibility), conviction rates, and likelihood of suspect guilt would decline as delay increased.

Hypothesis 2: Any inconsistency would discredit the witness, relative to a consistent witness, as reflected in lower cognitive competence, honesty, suggestibility, and suspect guilt ratings and fewer convictions.

Hypothesis 3: Of the three types of inconsistencies, contradictions would be the most detrimental to cognitive competence, honesty, and suggestibility.

Although prior work has examined the impact of delays in the context of delayed disclosures of sexual misconduct, we believed that findings would be similar in a different context, as memory may be diminished after a delay and, thus, would result in diminished credibility.

Method

Participants

A total of 465 undergraduate participants took part in the study in exchange for course credit. Participants were older than 18 years to represent a jury-eligible sample. A large number of participants failed one ($n = 137$), and a small number of participants failed both ($n = 16$), manipulation check questions (see the Results section of Study 1a for discussion of manipulation check questions). The majority of participants (74.4%) who failed one manipulation check did not accurately identify the type of inconsistency present. Participants frequently correctly identified that an inconsistency occurred (61.8%) but did not accurately identify which type. Only participants who failed the question regarding delay were excluded from analyses, resulting in a total of 412 participants. This was an adequate sample, based on an a priori power analysis, to detect a medium effect ($f = .25$, $\alpha = .05$, power = .80, $N = 279$). We further conducted a sensitivity analysis after completion of data collection and found that Study 1a had power of .80 to detect an effect as small as $f > .15$, power of .90 to detect an effect as small as $f > .19$, and power of .95 to detect an effect as small as $f > .20$.

Participants ($M_{age} = 20.13$ years) were primarily women (60.00%; 39.60% men; two participants did not provide gender) and Caucasian (31.70%), East Asian (30.50%), and South Asian (21.90%).

Materials

Videos. Participants watched one of 24 videos (two versions were created for each condition of the 3 × 4 design) of a 50-year-old

male witness describing a crime to a prosecuting lawyer. The witness described observing a man being assaulted and robbed late one evening. The witness indicated that he was interviewed by the police immediately after the crime took place, 1 day, 2 years, or 15 years before speaking with a prosecuting lawyer. The 2- or 15-year delay occurred because a suspect had not been identified. These delays were selected to represent delays that might occur in the real world, at both a shorter and longer delay (Connolly et al., 2015; *R. v. Jordan*, 2016; Rotenberg & Cotter, 2018). The witness's statement included one of four types of consistency about the weapon the perpetrator used during the assault (alternative option in parentheses). The witness was consistent in his recall (in the first interview, it was a gun [knife]; in the second interview, it was a gun [knife]), was contradictory (in the first interview, it was a knife [gun]; in the second interview, it was a gun [knife]), was reminiscent (in the first interview, he did not mention a weapon; in the second interview, it was a gun [knife]), or omitted previously reported information (in the first interview, it was a gun [knife]; in the second interview, he did not mention a weapon). The lawyer drew attention to the (in)consistency (e.g., "Last time, you said the man had a knife; now you say he had a gun?"). Two versions of the set of 12 videos were created to counterbalance the type of weapon discussed. Videos that mentioned only one weapon (i.e., consistent, reminiscence, omission) mentioned a gun in half the videos and a knife in the other half. Videos that mentioned both weapons (i.e., contradictory) were counterbalanced as to which weapon was mentioned first. The same actor portrayed the witness in all videos. All videos were approximately 2 min long.

Questionnaire. After viewing the video, participants first answered two manipulation check questions. One question was focused on the delay and asked participants to identify how long after the initial interview and crime the second interview took place (1 day, 2 years, or 15 years). The second question asked participants if the witness was consistent across the two interviews. Participants could select that the witness was consistent, reported contradictory details (contradiction), left something out of their second interview (omission), or reported something new in their second interview (reminiscence). This allowed us to identify whether participants recognized if there was an inconsistency and what type was present within one question.

The credibility scale (see Appendix) used in the present study was adapted from the Child Credibility Assessment Scale (Kehn et al., 2016). This scale includes an established set of questions that, when averaged, produce a single rating for perceived cognitive competence, honesty, and suggestibility. Limited materials have been developed to assess the perceived credibility of adults, and those established tend to focus on specific circumstances outside the context of the present work (e.g., Brodsky et al., 2010). It has been theorized that adults' credibility is assessed similarly to children in terms of perceived cognitive competence and honesty. We thus chose to use an adapted version of this scale to assess these factors, while also exploring perceived suggestibility in adults. As the scale was developed to assess the perceived credibility of children, the only adaptation made was that references to a child or adult were replaced with witness and interviewer. The adapted scale consisted of 15 questions, and each question was rated on a 7-point Likert-type scale. We averaged seven questions to produce the rating of perceived cognitive competence (e.g., ability to remember and answer questions), four questions to provide the participants' rating of perceived honesty (e.g., likelihood of lying; three were reverse coded), and four questions to provide a rating of perceived suggestibility (e.g., likelihood of being influenced by individuals' questions). This scale was developed by Kehn et al. (2016), who conducted a confirmatory factor analysis to determine whether the scale appropriately modeled the three factors of cognitive competence, honesty, and suggestibility.

Participants were then asked to make final legal decisions by providing a dichotomous verdict and rating the likelihood of the suspect's guilt from 0% to 100%, with a higher percentage reflecting a greater likelihood of guilt. Participants were also asked to rate their confidence in their verdict decision on a scale from 1 to 100 as well as to answer an open-ended question regarding the reasons for their verdict decision. Although we had no a priori hypotheses regarding reasons for the verdict, this question was included to provide additional insight into participants' decision making. Finally, participants provided demographic information (i.e., age, sex, primary language, nationality, ethnicity, occupation).

Procedure

This study was approved by the Simon Fraser University Research Ethics Board (Protocol No. 20180214). We recruited participants from undergraduate psychology courses. Participation took place in person, in groups of up to four, although all aspects of the study took place independently (i.e., no deliberation took place and questionnaires were completed independently). After participants signed the consent form, each group was randomly assigned to watch one of 24 videos on a projector screen in a mock jury room. After viewing the video, participants answered the manipulation check questions, completed the credibility scale, and were then asked to pretend that the case went forward to trial and to make a verdict decision, rate the likelihood of the suspect's guilt, rate their confidence in their verdict decision, explain their reasoning for their verdict decision, and provide demographic information. Participants were debriefed and thanked for their participation.

Coding Reasons for Verdict

Two coders read all responses, identified common categories, and created a coding guide based on the categories. They identified 15 categories (e.g., amount of evidence presented, witness demeanor, and witness consistency; see the online Supplemental Materials for all categories). If multiple reasons were provided, only the first response from each participant was used in the present analyses as it was believed to be the participant's most salient reason. All responses were double coded, and disagreements were discussed. Interrater agreement was established at $\kappa > .80$.

Results

We first conducted one-way analyses of variance (ANOVAs) to examine effects of the weapon order on cognitive competence, honesty, suggestibility, verdict, likelihood of suspect guilt, and confidence in verdict. As no significant differences were found between responses for participants who watched videos with a knife mentioned first and those who watched videos with a gun mentioned first ($ps > .37$), we collapsed across weapon order for all subsequent analyses.

Participants may have been confused about which response option accurately described the type of inconsistency they heard, resulting in a high failure rate on one of the manipulation check questions. Thus, we believe that only one manipulation check question (i.e., how long ago did the crime occur?) assessed the effectiveness of manipulations as intended. We analyzed data both with and without participants who failed the question about type of inconsistency. In both analyses, the conclusions were the same; therefore, we present the results including participants who failed the question here (see the online Supplemental Materials for presentation of results excluding those who failed the consistency manipulation check).

First, we present the results of 3 (delay) × 4 (consistency) ANOVAs that examined the effects of delay and consistency on perceived cognitive competence, honesty, and suggestibility, as well as likelihood of guilt and confidence in verdict (see Table 1 for descriptives of delay conditions). Although we intended to examine interactions, the present sample size does not provide adequate power to detect interaction effects. Thus, we explored only main effects in the present study. In some ANOVAs, the assumption of normality or equal variances was not met, and so a Kruskal–Wallis test was performed. In all cases, the patterns were identical to the ANOVA test, and so the ANOVA results are reported. Then, we present a binary logistic regression analysis of the dichotomous verdict decision (see Table 2 for frequencies of verdict decisions in delay conditions). See the online Supplemental Materials for post hoc analyses and descriptives of consistency conditions.

Cognitive Competence

A significant effect of consistency was found, $F(3, 397) = 16.19$, $p < .001$, $\eta_p^2 = .11$. Post hoc analyses revealed that participants found a consistent witness to be significantly more cognitively competent ($M = 5.07$) than witnesses with any type of inconsistency ($ps < .001$), and types of inconsistency did not differ from one another ($ps > .797$; $M_{contradiction} = 4.21$, $M_{omission} = 4.25$, $M_{reminiscence} = 4.35$). Delay did not have a significant effect on the perceived cognitive competence of the witness, $F(2, 397) = 0.71$, $p = .492$, $\eta_p^2 < .01$.

Honesty

There was an effect of consistency on perceived honesty, $F(3, 399) = 7.10$, $p < .001$, $\eta_p^2 = .05$. Follow-up tests indicated that participants viewed a consistent witness as significantly more honest ($M = 5.23$) than a witness with any level of inconsistency ($p < .01$; $M_{contradiction} = 4.71$, $M_{omission} = 4.64$, $M_{reminiscence} = 4.53$). There were no differences between levels of inconsistency ($p > .67$), nor was there a significant effect of delay, $F(2, 399) = 0.71$, $p = .490$, $\eta_p^2 < .01$.

Suggestibility

There was an effect of consistency on perceived suggestibility, $F(3, 398) = 4.25$, $p = .006$, $\eta_p^2 = .03$. Post hoc analyses revealed that participants rated a witness who had an omission ($M = 4.73$) as significantly more suggestible than a consistent witness ($M = 4.21$; $p = .004$). There were no other significant differences between consistency conditions. There was no significant impact of delay on participants' ratings of witness suggestibility, $F(2, 398) = 0.95$, $p = .389$, $\eta_p^2 < .01$.

Dichotomous Verdict

Participants rendered a dichotomous verdict decision (guilty, not guilty). A binary logistic regression was conducted to investigate the impact of our independent variables on verdict decisions. Delay and consistency were entered into the model as predictors. The overall model was not significant, $\chi^2(2) = 4.63$, $p = .10$.

Likelihood of Suspect Guilt

Participants also rated the likelihood of the suspect's guilt on a scale from 1 to 100. A significant effect of delay was found on likelihood of suspect guilt, $F(2, 397) = 5.05$, $p = .007$, $\eta_p^2 = .03$. Follow-up tests showed that participants rated the likelihood of suspect guilt lower when the witness had a 2-year delay ($p = .047$) or 15-year delay ($p = .011$) between the crime and their interview with the prosecutor, compared with a 1-day delay ($M = 71.39$). There was no difference between the 2-year delay condition and 15-year delay condition ($p = .839$). There was no significant impact of consistency, $F(3, 397) = 0.06$, $p = .981$, $\eta_p^2 < .01$.

Verdict Confidence

Participants rated their confidence in their verdict on a scale from 1 to 100. There was a significant impact of delay on participants' confidence in their verdict decision, $F(2, 395) = 4.18$, $p = .016$, $\eta_p^2 = .02$. Post hoc analyses indicated higher confidence in verdicts for a witness retelling their story after a 1-day delay compared with a 2-year delay ($p = .015$). No significant difference was found

Table 1

Study 1a Means, Standard Deviations, and Confidence Intervals of Credibility Ratings, Likelihood of Suspect Guilt, and Verdict Confidence

	1-day delay		2-year delay		15-year delay		Total	
Dependent variable	*M* (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]
Cognitive competence	4.55 (1.01)	[4.36, 4.71]	4.51 (1.11)	[4.32, 4.67]	4.37 (1.15)	[4.20, 4.57]	4.48 (1.09)	[4.37, 4.57]
Honesty	4.89 (1.20)	[4.68, 5.08]	4.72 (1.23)	[5.04, 4.91]	4.76 (1.27)	[4.55, 4.98]	4.79 (1.23)	[4.67, 4.90]
Suggestibility	4.68 (1.13)	[4.33, 4.69]	4.35 (1.04)	[4.17, 4.53]	4.50 (1.15)	[4.31, 4.70]	4.46 (1.12)	[4.35, 4.56]
Likelihood of suspect guilt	71.39 (20.29)	[67.26, 75.48]	65.68 (22.73)	[61.41, 69.88]	63.87 (20.59)	[59.72, 68.03]	67.03 (21.38)	[64.56, 69.37]
Verdict confidence	71.67 (17.55)	[67.61, 75.52]	64.49 (22.04)	[50.55, 68.39]	64.56 (19.19)	[60.77, 68.46]	66.97 (19.86)	[64.71, 69.16]

Note. Participants rated cognitive competence, honesty, and suggestibility on a scale from 1 to 7. Participants rated likelihood of suspect guilt and verdict confidence out of 100. CI = confidence interval.

Table 2
Study 1a Frequencies (Percentage of Total Decision in Condition) of Guilty Verdicts

1-day delay	2-year delay	15-year delay	Total
108 (75.00%)	92 (66.19%)	81 (63.28%)	281 (68.36%)

between the 1-day and 15-year delay conditions ($p = .168$) or the 2-year and 15-year delay conditions ($p = .616$). There was no significant impact of consistency on verdict confidence, $F(3, 395) = 1.46$, $p = .226$, $\eta_p^2 = .01$.

Reasons for Verdict Decision

Participants were asked an open-ended question regarding reasons for their verdict decision, and their first reasons given were analyzed. Participants primarily discussed the amount of evidence provided (e.g., not enough) and the witness's consistency. See the online Supplemental Materials for all categories and frequencies.

Study 1b

Given the unexpected null effects of delay on the measures of perceived credibility as well as no observed differences between the types of inconsistencies on our dependent variables, we sought to replicate central elements of Study 1a in Study 1b. Study 1b was a partial replication of Study 1a, in which we streamlined our conditions to focus on two levels of delay and three levels of inconsistency. As we found no difference between 2-year and 15-year delays in Study 1a, we focused only on the shorter delay. Study 1b was a 2 (delay: 1 day, 2 years) × 3 (report inconsistency: contradiction, reminiscence, omission) between-subjects design.

Method

Participants

Two hundred eighty-eight undergraduate participants took part in the study in exchange for course credit. This study was approved by the Simon Fraser University Research Ethics Board (Protocol No. 20180214). Participants were older than 18 years to represent a jury-eligible sample. Participants who did not pass the manipulation check regarding the length of delay were removed for analysis ($n = 29$), but those who did not pass the manipulation check regarding the type of inconsistency were kept as part of the final sample ($n = 49$). Again, we believe that participants struggled to identify the type of inconsistency on the basis of the options presented in the manipulation check question. Thus, these participants were included, as we believed that only one manipulation check worked as intended, and to remain consistent with the inclusion of participants in Study 1a. An additional 19 participants were removed from the final sample, as they did not complete the study, leaving a total of 240 participants for analysis. According to a sensitivity analysis, Study 1b had power of .80 to detect an effect as small as $f > .20$, power of .90 to detect an effect as small as $f > .23$, and power of .95 to detect an effect as small as $f > .26$. Participants ($M_{age} = 19.67$ years) were primarily women (71.55%; 28.45% men) and White (28.75%), East Asian (26.25%), and South Asian (25.00%).

Materials

Materials were very similar to those in Study 1a, with the removal of two conditions: consistent statements and 15-year delay. There were a few other changes to the materials. Participants were not asked to generate a reason for their verdict decision in Study 1b but were asked to select which of four options (generated from participant responses in Study 1a and our variables of interest) had the biggest impact on their verdict decision: the witness's story, details of the investigation, timing of the investigation, or other. As we predicted that similar reasons for the verdict decisions would be reported in the present study and in Study 1a, we chose to simplify responses by providing options for participants to choose from, rather than ask them to respond to an open-ended question. Participants then rated their confidence in their verdict decision on a scale from 1 to 10, rather than 1 to 100.

Procedure

The procedure was identical to that in Study 1a with the exception that it took place online and individually. Participants viewed a video of a witness being interviewed, completed two manipulation check questions and an adapted version of the Child Credibility Assessment Scale (Kehn et al., 2016), and made final legal decisions before providing demographic information.

Results

We conducted a series of 2 (delay: 1 day, 2 years) × 3 (report inconsistency: contradiction, reminiscence, omission) ANOVAs to evaluate the effect of delay and inconsistency on the perceived cognitive competence, honesty, and suggestibility of the witness, along with likelihood of suspect guilt ratings and confidence in verdict decisions. See Table 3 for means and standard deviations of delay conditions. In some ANOVAs, the assumption of normality or equal variances was not met, and so a Kruskal–Wallis test was performed. In all cases, the patterns were identical to the ANOVA test, and so the ANOVA results are reported. Again, the present study was not adequately powered to test for interaction effects. A binary logistic regression examined the impact of the independent variables on dichotomous verdict decisions. See Table 4 for frequencies of guilty verdicts for delay conditions. See the online Supplemental Materials for post hoc analyses and descriptives of report inconsistency conditions.

Cognitive Competence

There was no significant effect of delay, $F(1, 234) = 0.002$, $p = .962$, $\eta_p^2 < .001$, or inconsistency, $F(2, 234) = 0.75$, $p = .476$, $\eta_p^2 < .001$, on the witness's perceived cognitive competence.

Honesty

There was no significant impact of delay on the perceived honesty of the witness, $F(1, 234) = 0.55$, $p = .46$, $\eta_p^2 = .002$. There was a significant impact of inconsistency, $F(2, 234) = 6.07$, $p = .002$, $\eta_p^2 = .05$. Tukey's post hoc analyses revealed that a witness with an omission was seen as more honest than a witness with a reminiscence ($p < .002$, $d = -0.57$, 95% confidence interval [CI; −0.89, −0.24]). There were no other significant differences ($ps > .18$).

Table 3
Study 1b Means, Standard Deviations, and Confidence Intervals of Credibility Ratings, Likelihood of Suspect Guilt, and Verdict Confidence

	1-day delay		2-year delay		Total	
Dependent variable	*M* (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]
Cognitive competence	4.28 (0.83)	[4.14, 4.43]	4.27 (0.80)	[4.12, 4.42]	4.28 (0.82)	[4.17, 4.38]
Honesty	4.70 (1.07)	[4.51, 4.89]	4.78 (1.06)	[4.59, 4.98]	4.74 (1.06)	[4.61, 4.88]
Suggestibility	4.47 (1.01)	[4.29, 4.65]	4.30 (0.97)	[4.12, 4.48]	4.39 (0.99)	[4.26, 4.51]
Likelihood of suspect guilt	66.00 (21.30)	[62.20, 69.80]	70.30 (18.30)	[66.90, 73.70]	68.10 (20.00)	[65.50, 70.70]
Verdict confidence	6.19 (2.03)	[5.83, 6.55]	6.37 (1.87)	[6.03, 6.71]	6.28 (1.95)	[6.03, 6.53]

Note. Participants rated cognitive competence, honesty, and suggestibility on a scale from 1 to 7. Participants rated likelihood of suspect guilt out of 100 and verdict confidence out of 10. CI = confidence interval.

Suggestibility

There was no significant effect of delay on the perceived suggestibility of the witness, $F(1, 234) = 1.77$, $p = .184$, $\eta_p^2 = .007$, nor was there an effect of consistency, $F(2, 234) = 0.21$, $p = .81$, $\eta_p^2 = .002$.

Dichotomous Verdict

A binary logistic regression with delay and inconsistencies entered as predictors revealed that the overall model was not significant, $\chi^2(3) = 1.71$, $p = .636$. Thus, there was no impact of delay or inconsistencies on dichotomous verdict decisions.

Likelihood of Suspect Guilt

There was no significant impact of delay, $F(1, 231) = 2.87$, $p = .091$, $\eta_p^2 = .01$, or inconsistency, $F(2, 231) = 0.06$, $p = .94$, $\eta_p^2 < .01$, on the likelihood of suspect guilt ratings.

Verdict Confidence

There was no impact of delay, $F(1, 232) = 0.48$, $p = .49$, $\eta_p^2 = .002$, or inconsistency, $F(2, 232) = 0.08$, $p = .92$, $\eta_p^2 < .01$, on confidence in verdict.

Factors Influencing Verdict

Participants were asked to identify which of the following contributed most to the verdict decision: the witness's story, details of the investigation, timing of the investigation, or other. The majority of participants (60%) selected the witness's story, followed by details of the investigation (32%), timing of the investigation (5%), and other (3%).

Table 4
Study 1b Frequencies (Percentage of Total Decision in Condition) of Guilty Verdicts

1-day delay	2-year delay	Total
88 (74.57%)	88 (75.86%)	63 (73.64%)

Studies 1a and 1b Discussion

In Study 1a, we examined effects of inconsistencies and delay on perceptions of witness credibility and suspect guilt. A witness who showed any type of inconsistency (i.e., contradiction, omission, reminiscence) was seen as less cognitively competent and less honest than a consistent witness, supporting our hypothesis. Contrary to our hypothesis, there was no significant difference in the impact of different types of inconsistencies on perceived cognitive competence and honesty. We predicted that contradictions would be more detrimental to measures of perceived credibility than other types of inconsistencies; however, this pattern was not observed. Nor did we find that witness consistency affected perceptions of guilt. We conducted Study 1b to explore the null findings between types of inconsistencies and found the same pattern of results, with the exception of a significant effect of inconsistency on honesty. Participants found a witness with an omission to be more honest than a witness with reminiscence. These consistent observations—no significant effect of delay on credibility and little difference between varying types of inconsistencies in their effect on credibility—raise questions for our understanding of credibility and offer a pressing motivation for additional research.

The most interesting and potentially important finding from Studies 1a and 1b was that delay had no significant effect on perceptions of witness credibility (i.e., cognitive competence, honesty, suggestibility). An important difference between our studies and previous studies that found a detrimental effect of delay on credibility (e.g., Balogh et al., 2003; Ellison & Munro, 2009; Miller et al., 2022) was that the reason for delay in the present study was beyond the witness's control (i.e., no suspect had been identified). It is possible that the effect of delay on perceived credibility is impacted by the reason for delay. That is, perhaps a delay caused by the witness impacts their credibility, whereas a delay that is outside the witness's control has no significant effect on their perceived credibility. In other words, the impact of delay might be related to perceived witness motivation rather than perceptions of the effect of delay on memory for the event. We tested this possibility in Study 2 by examining how different reasons for a delay impact the perceived credibility of a witness and legal decisions.

Another possible explanation for the null effect of delay on perceived credibility in Studies 1a and 1b is the inclusion of an (in)consistency manipulation. When judging a delayed report, participants may be assessing possible memory errors, revealed through (in)consistent reports. When consistency is addressed by including an (in)consistency manipulation, as in Studies 1a and 1b,

effects of delay may be fully accounted for through evaluations of report consistency. One way to test this possibility is to remove inconsistencies from the design. If the effect of delay on perceptions of credibility can be accounted for with report (in)consistencies, the negative effect should return when inconsistencies are removed from the design. Thus, we removed the inconsistencies in Study 2.

A long delay in Study 1a reduced the likelihood of guilt ratings and participants' confidence in their verdict, relative to a delay of only 1 day, consistent with our hypothesis. These differences were not observed in Study 1b. Given that delay conditions were identical in each study (i.e., 1 day, 2 years; there was no difference between 2-year and 15-year delay conditions), the removal of the consistent condition in Study 1b may partially explain the differences. In the analyses of delay main effects, all consistency conditions were collapsed. The higher ratings in the no delay consistent condition of Study 1a (a condition not present in Study 1b because the consistent condition was removed) may have raised the average ratings in the no delay condition enough to be statistically different from the long delay in Study 1a. Given that this is the only design difference between the studies, this potential difference between short and long delays should be interpreted with caution.

As expected, in Study 1a, any inconsistency diminished perceptions of the witness's honesty and cognitive competence (e.g., Berman & Cutler, 1996). However, in both Studies 1a and 1b, contradictions were not more detrimental to perceived credibility than other types of inconsistencies. All inconsistencies had a similar negative impact on perceived cognitive competence and honesty in Study 1a, reminiscences negatively impacted honesty compared with omissions in Study 1b, and omissions were the most detrimental to perceived suggestibility in Study 1a. One possible explanation is that participants, although aware that an inconsistency was present, did not recognize the type of inconsistency. To explore this possibility, we reran the analyses with only participants who correctly identified the type of inconsistency. This did not alter the results; contradictions did not have a more negative impact on perceived credibility than other types of inconsistencies. A second possibility relates to the subject of the inconsistent detail, the weapon. Participants may have believed that the witness should have been able to remember weapon type, and any type of inconsistency about such a central detail may have been seen as equally damaging.

Although an inconsistent witness was discredited relative to a consistent witness, consistency did not impact final legal decisions. The null effect of consistency on verdict may be a result of the threshold at which participants decided guilt. The strength of a variable needed to affect dichotomous verdict decisions will vary as a function of how close participants are to the threshold to convict. For instance, if a participant is 60% sure that the suspect is guilty and a consistent witness raises this to 70% certainty, it is still not close enough to threshold to convict to change the dichotomous decision from acquit to convict. Although inconsistencies resulted in lowered perceived credibility, participants may not have been close enough to threshold to alter their dichotomous verdict decisions. Prior research has found that triers of fact, in practice, are aware of potential negative consequences of inconsistencies in a witness's testimony, but it is not a strong predictor of trial outcome (Connolly et al., 2009). We saw similar patterns in our own data: Participants were aware of the inconsistencies and found them harmful for the perceived credibility of the witness, but this did not significantly impact their final legal decisions.

Study 2

In Studies 1a and 1b, the reason for the delay was failure to identify a suspect, and delay had no significant effect on the witness's perceived credibility. In Study 2, we explored different reasons for delay, given the unexpected results in the first two studies. Participants viewed a witness describing a crime they saw and were asked to rate the witness's perceived cognitive competence, honesty, and suggestibility as well as to make legal decisions. The witness was interviewed immediately after the crime or 15 years later. We included only the longer delay condition from Study 1a to exaggerate the difference from the immediate condition, as there were no notable differences observed between 2 and 15 years. In this study, the reason for delay was either not provided, described as due to the witness's reluctance to be interviewed, or because a suspect had not been identified by the police. This was a one-way between-subjects design with four conditions: no reason–no delay, no reason-delay, witness-delay, and investigation-delay.

We developed the following hypotheses:

Hypothesis 1: There would be no significant effect of delay on the measures of perceived credibility of the witness when the reason for the delay was outside of the witness's control (as was the case in Studies 1a and 1b).

Hypothesis 2: A delay due to the witness's reluctance would have a negative effect on perceived credibility (hypothesized on the basis of extant literature).

Hypothesis 3: When any delay was present, it would result in fewer guilty verdicts and a lower likelihood of suspect guilt (hypothesized on the basis of the results of Studies 1a and 1b).

Method

Participants

We recruited participants older than 18 years via the online data collection platform, Prolific, from both the United States and Canada, to represent a jury-eligible sample. Whereas Studies 1a and 1b used an undergraduate sample, we believe the inclusion of both undergraduate and community participants is a benefit to the generalizability of our findings. Thus, Study 2 used a community sample. A total of 297 participants took part in exchange for monetary compensation. Twenty-three participants did not complete the study, and an additional 74 failed one or both manipulation check questions (17 participants did not correctly identify the delay, and 54 did not correctly identify the reason for the delay). If participants failed any of the manipulation check questions, they were removed, leaving a total of 200 participants for analyses. This manipulation check failure rate is not outside the norm; Abbey and Meloy (2017) found a failure rate ranging from 4% to 45% (M = 19.56%) for online participation based on manipulation checks across 30 sets of data. Our failure rate of 27.01% is well within that range. Two hundred participants was an adequate sample, based on an a priori power analysis, to detect a medium effect (f = .25, α = .05, power = .80, N = 180). According to a sensitivity analysis,

Study 2 had power of .80 to detect an effect as small as $f > .20$, power of .90 to detect an effect as small as $f > .22$, and power of .95 to detect an effect as small as $f > .25$.

Participants ($M_{age} = 33.94$ years) were 45.50% women (50.50% men, 4.00% nonbinary) and were primarily Caucasian (65.48%), East Asian (8.12%), South Asian (3.55%), and Black (3.55%).

Procedure

This study was approved by the Simon Fraser University Research Ethics Board (Protocol No. 30000001). Participation took place online. As in Studies 1a and 1b, participants watched a video depicting a witness describing a crime they saw. After viewing the video, participants completed an adapted version of the Child Credibility Assessment Scale (Kehn et al., 2016), made final legal decisions (verdict, likelihood of suspect guilt), rated their confidence in their verdict decision, provided reasoning for their verdict decision, and provided demographic information.

Materials

Videos. Videos were adapted from the materials used in Studies 1a and 1b. A video was edited to include only basic information about what the witness saw (i.e., the crime and a description of the perpetrator) and removed information about the prior interview and inconsistencies between interviews. The same video was used across all conditions, with relevant information about the delay and reason for delay provided in both text and audio form, presented at the beginning of the video. Participants were told that the crime occurred either immediately before the interview or 15 years prior. If a delay occurred, participants were either not provided with a reason, told that the delay was due to the witness choosing to not come forward, or told that the police had not identified a suspect.

Questionnaire. The questionnaire was identical to that used in Studies 1a and 1b with one exception. The manipulation check questions were updated to reflect the new design. All participants were asked if a delay was present, and only those who reported a delay were asked the reason for the delay. If participants identified the presence of a delay by selecting that the crime had occurred 15 years prior (as opposed to immediately before the interview), they were presented with a question asking why the interview was only now taking place. Participants could select a response from the following options: The reason for the delay was not provided; the police had not identified a suspect until now, or the witness did not want to be interviewed until now. Participants responded to the manipulation check questions prior to completing the remainder of the questionnaire. Given the addition of new variables, we asked participants to once again provide reasons for their verdict decision to provide additional insight into their decision making.

Coding

Coding for reasons for verdict was done as described in Study 1a. Coders identified a total of nine categories. Interrater agreement was established at $\kappa > .80$. Two coders then coded all responses and discussed and came to an agreement on all disagreements. See the online Supplemental Materials for all categories.

Results

We computed scores for cognitive competence, honesty, and suggestibility as described in Study 1a. We first present the results of ANOVAs that examined the effects of the reason for delay on perceived cognitive competence, honesty, and suggestibility. When the assumption of normality or equal variances was not met, Kruskal–Wallis analyses were run. In all cases, the patterns were identical, and so only ANOVA results are reported. We also present a binary logistic regression to analyze dichotomous verdict decisions and ANOVAs that analyzed the effect of reason for delay on likelihood of suspect guilt and confidence in verdict decisions. Means and standard deviations for the continuous variables are in Table 5, and frequencies for the verdict decisions are in Table 6. Post hoc analyses can be found in the online Supplemental Materials.

Cognitive Competence

There was a significant effect of reason for delay on perceived cognitive competence, $F(3, 195) = 2.86$, $p = .038$, $\eta_p^2 = .04$. Tukey's post hoc analysis revealed a significant difference between the no reason–no delay and witness-delay conditions ($p = .033$, $d = 0.52$, 95% CI [0.13, 0.90]). When the witness was the reason for the delay (i.e., did not want to talk to the police), they were seen as less cognitively competent compared with a witness who had no delay before their interview. There were no other significant differences ($ps > .23$).

Honesty

There was a significant effect of reason for delay on perceived honesty, $F(3, 195) = 6.53$, $p < .001$, $\eta_p^2 = .09$. Tukey's post hoc analyses revealed a significant difference between the no reason–no delay and witness-delay conditions ($p < .001$, $d = 0.82$, 95% CI [0.41, 1.21]). A witness who delayed their interview was viewed as less honest than a witness with no delay. There was also a significant difference between the witness-delay and investigation-delay conditions ($p = .042$, $d = 0.48$, 95% CI [0.06, 0.89]). When the witness, rather than the investigation, was at fault for the delay, they were seen as less honest. There were no other significant differences ($ps > .90$).

Suggestibility

There was no significant effect of the reason for delay on perceived suggestibility, $F(3, 195) = 1.15$, $p = .33$, $\eta_p^2 = .02$.

Dichotomous Verdict

A binary logistic regression was conducted with reason for delay entered as a predictor and verdict decision as an outcome variable. The overall model was significant, $\chi^2(3) = 10.95$, $p = .012$ (see the online Supplemental Materials for full model information). Follow-up z tests revealed a significant difference between the no reason–no delay and investigation-delay conditions ($z = 3.34$, $p < .001$). When the delay was due to the investigation, there were significantly fewer convictions. The difference between the no reason–no delay condition and both the witness-delay ($z = 1.83$, $p = .067$) and no reason-delay ($z = 1.73$, $p = .083$) conditions was not significant, although both showed a pattern of a delay resulting in fewer guilty verdicts.

Table 5
Study 2 Means, Standard Deviations, and Confidence Intervals of Credibility Ratings, Likelihood of Suspect Guilt, and Verdict Confidence

Dependent variable	No reason–no delay		No reason-delay		Witness-delay		Investigation-delay		Total	
	M (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]
Cognitive competence	5.49 (0.79)	[5.28, 5.69]	5.33 (0.87)	[5.07, 5.59]	4.98 (1.17)	[4.65, 5.30]	5.11 (1.06)	[4.79, 5.44]	5.23 (0.99)	[5.09, 5.37]
Honesty	5.92 (0.89)	[5.69, 6.15]	5.48 (0.90)	[5.21, 5.75]	5.05 (1.23)	[4.71, 5.39]	5.63 (1.19)	[5.26, 6.00]	5.52 (1.10)	[5.38, 5.68]
Suggestibility	3.44 (1.05)	[3.17, 3.71]	3.55 (1.07)	[3.22, 3.87]	3.75 (1.08)	[3.46, 4.05]	3.34 (1.51)	[2.87, 3.80]	3.52 (1.18)	[3.36, 3.69]
Likelihood of suspect guilt	73.60 (19.20)	[68.50, 78.60]	65.50 (23.20)	[58.50, 72.40]	61.00 (25.90)	[53.80, 68.20]	55.00 (29.50)	[45.90, 64.10]	64.40 (25.20)	[60.90, 67.90]
Verdict confidence	6.66 (2.28)	[6.07, 7.25]	5.91 (2.39)	[5.18, 6.64]	6.47 (2.06)	[5.90, 7.04]	6.00 (2.31)	[5.27, 6.73]	6.30 (2.26)	[5.99, 6.62]

Note. Participants rated cognitive competence, honesty, and suggestibility on a scale from 1 to 7. Participants rated likelihood of suspect guilt out of 100 and verdict confidence out of 10. CI = confidence interval.

Likelihood of Guilt

Participants were asked to imagine that the case went to trial and rate the accused's likelihood of guilt from 1 to 100. There was a significant effect of reason for delay on participants' ratings of likelihood of guilt, $F(3, 194) = 5.24$, $p < .002$, $\eta_p^2 = .07$. Tukey's post hoc analysis showed a significant difference between the no reason–no delay condition and both the witness–delay ($p = .037$, $d = 0.56$, 95% CI [0.16, 0.96]) and investigation-delay ($p < .002$, $d = 0.77$, 95% CI [0.34, 1.19]) conditions. When there was a delay by fault of the witness or investigation, the suspect's likelihood of guilt was rated lower than when no delay was present. There were no other significant differences ($ps > .18$).

Confidence in Verdict

Participants rated their confidence in their verdict decision on a scale from 1 to 10. There was no significant impact of the conditions on verdict confidence, $F(3, 193) = 1.29$, $p = .28$, $\eta_p^2 = .02$.

Reasons for Verdict Decisions

Participants were also asked to provide reasoning for their verdict decision. Coders identified themes based on open-ended responses and then coded responses into one of nine categories. Participants identified the witness being generally credible (e.g., general demeanor, actions) as their primary reason for their verdict decision (37.90%), followed by requiring additional evidence beyond the witness (e.g., identification procedure; 11.10%) and the delay negatively impacting the witness's memory (10.10%). A portion of comments was also categorized as "other" (e.g., stating facts about the case, unclear reasons provided; 16.70%).

Discussion

When the delay was the fault of the witness, participants found them to be less cognitively competent and less honest than when no delay was present, consistent with our hypothesis. Also consistent with our hypothesis, an investigation-delay resulted in fewer convictions, and when a reason for the delay was provided (witness or investigation), likelihood of guilt ratings were lower than when there was no delay. The most interesting finding was that delay diminished credibility only when the delay was a result of the witness's actions. A delay with no reason provided or a delay caused by the investigation had no significant effect on the perceived credibility of the witness.

Study 3

In the prior studies, we were underpowered for exploring interactions (Studies 1a and 1b), and we used a one-way design with four delay conditions in Study 2, which did not allow us to explore interactions and compare all possible conditions. Thus, in Study 3, we conducted a fully crossed design that was powered to detect interactions. Study 3 investigated the impact of delay and contextual information (i.e., reason for delay) on the perceived credibility of a witness and legal decision making. This was a 2 (delay: immediate, 2 years) × 3 (contextual information: no details, witness, investigation) between-subjects design. Note that we refer to contextual information as the reason for delay and the no-details

Table 6
Study 2 Frequencies (Percentage of Total Decision in Condition) of Guilty Verdicts

No reason–no delay	No reason-delay	Witness-delay	Investigation-delay	Total
46 (77.97%)	28 (62.22%)	33 (62.26%)	20 (46.51%)	127 (86.50%)

condition as the control condition in our preregistration. For clarity, we have updated these labels to better reflect the manipulations. It is possible that our prior investigation-delay manipulation was attributed to the witness's actions if participants reasoned that a suspect was not identified because the witness was not cooperative. Thus, we altered the investigation information provided in Study 3 to state that there was confusion regarding which police force was in charge of collecting evidence. We anticipated replicating our results from prior studies while extending our knowledge regarding the impact of delay on perceived credibility. We expected that only when a delay is the result of the witness's actions would it negatively impact their perceived cognitive competence and honesty. However, any salient delay (e.g., delay that has a reason provided) would impact likelihood of guilt ratings and convict/acquit decisions. Thus, our hypotheses were as follows:

Hypothesis 1: There would be a significant interaction between delay and contextual information on the perceived cognitive competence and honesty of the witness. The witness's perceived cognitive competence and honesty would be rated lower when the delay was a result of the witness, compared with all no delay conditions. In Study 2, there was some indication that the witness-delay condition might be rated lower than the investigation-delay and no details-delay conditions on perceived honesty, but the comparison was not statistically significant for cognitive competence, so this pattern remained exploratory here.

Hypothesis 2: There would be a significant interaction of delay and contextual information on dichotomous verdict decisions. A delay with a reason provided (i.e., witness-delay, investigation-delay) would result in fewer convictions compared with no delay. There would be no other differences between conditions.

Hypothesis 3: There would be a significant interaction of delay and contextual information on likelihood of guilt ratings. A delay with a reason provided (i.e., witness-delay, investigation-delay) would result in lower likelihood of guilt ratings compared with no delay. There would be no other differences in conditions.

Hypothesis 4: There would be no impact of delay or contextual information on perceived suggestibility.

Hypothesis 5: There would be no impact of delay or contextual information on confidence in verdict.

Method

This was a 2 (delay: immediate, 2 years) × 3 (contextual information: no details, witness, investigation) between-subjects design. As our prior work found no significant differences between a delay of 2 years and a delay of 15 years, we chose to use the shorter delay in Study 3; we theorized that this might be more realistic to participants in the given context. This study was approved by the Simon Fraser University Research Ethics Board (Protocol No. 30002141). This study was preregistered on the Open Science Framework (https://osf.io/jqrh7/).

Participants

Six hundred thirty-three participants, recruited online via Prolific, took part in the present study. We collected more participants than preregistered, to account for withdrawals and exclusions based on manipulation checks, in order to obtain our registered usable full sample. Twenty-three participants withdrew early from the study, eight participants did not provide responses to all dependent variables, one participant completed the study twice, and 242 participants failed one or both manipulation checks (131 failed the investigative details manipulation check, 61 failed the delay manipulation check, and 50 failed both). Although we had a greater manipulation check failure rate than anticipated in our registered report, a failure rate of 38.72% is not outside of the range found in online studies (i.e., 4%–45%; Abbey & Meloy, 2017). We had a final total sample of 363 participants. A sample size of 360 participants (60 participants per condition) is adequate to detect a small effect (Cohen's $f = .10$) at > 90% power ($\alpha = .05$), according to a Superpower analysis. Our prior studies have been completed by predominately White and female samples. It is possible that this unbalanced sample has implications for our dependent variables, as past work has shown that there may be gender-related differences in credibility research (e.g., Hockett et al., 2016). Thus, for Study 3, we balanced gender within each cell of our design. Although other demographic factors may impact perceptions of credibility and legal decision making, we chose to focus on balancing gender in the present work, as it was anticipated to have the greatest potential impact on our findings.

Participants were 50.14% female and 0.002% nonbinary. Two participants did not report gender. Participants spoke mainly English (92.84%); self-identified as White (53.17%), Black (15.15%), and East Asian (12.12%); and had a mean age of 34.65 ($SD = 11.72$) years. Participants were primarily Canadian (45.45%) and American (39.12%) citizens.

Procedure

Participation took place online and was similar to that in Studies 1a, 1b, and 2. After providing consent, each participant was randomly assigned to one of six possible conditions. Participants viewed a video of a witness discussing a crime they saw take place, preceded by information regarding the time since the crime occurred and contextual information. After watching the video, participants were asked to complete an adapted version of the Child Credibility Assessment Scale as well as make legal decisions regarding the guilt of the suspect, a dichotomous verdict decision, their confidence in their verdict, and reasoning for their verdict decision. Participants

were asked to complete up to three questions regarding our manipulations. Finally, participants were asked to provide demographic information.

Materials

Videos. Participants viewed one of six videos. The video began with contextual information regarding how long after the crime the interview took place (i.e., just after the crime or 2 years later) and the reason for a delay, when applicable, or additional information (see the online Supplemental Materials for video transcript and condition information). Participants in the investigation conditions, delayed or immediate, were informed that there was confusion regarding which police force was responsible for gathering evidence, a situation mirroring real-world cases (e.g., Robert Pickton; The Current, 2015). Participants in the witness conditions, delayed or immediate, were informed that the witness was reluctant to be interviewed but was now being interviewed. In the no-details condition, no additional information was provided about the witness or police. The immediate conditions made it clear that the interview happened shortly after the crime, whereas the delayed conditions made it clear that the interview happened 2 years later. Aside from this contextual information, the video of the witness was identical to that in Study 2. The witness was interviewed about the crime they viewed (an assault and theft). The witness provided general details about the crime and gave a generic description of the perpetrator. The interview was the same across all conditions, and only the contextual information changed.

Questionnaire. The questionnaire provided to participants was the same as in Studies 1a, 1b, and 2 (see the online Supplemental Materials)—an adapted version of the Child Credibility Assessment Scale, along with questions regarding the guilt of the perpetrator. Participants made a dichotomous verdict decision and rated the likelihood of suspect guilt on a scale from 1 to 100. Participants rated their confidence in their verdict decision on a scale from 1 to 100 as well as provided their most important reason for making their verdict decision. Participants also provided demographic information (e.g., gender, age, nationality, ethnicity, occupation).

To assess whether participants paid attention to our manipulation, we also asked up to three questions. All responses were made by selecting from presented options. All participants were asked when the crime took place. Participants who selected a delay then selected the reason for the delay and who caused the delay. Participants who selected no delay then selected what additional information, if any, was provided when an immediate report was given (i.e., the witness was reluctant to be interviewed, there was confusion regarding which police force was in charge of gathering evidence, no additional information was provided).

Coding

Participants' responses to the open-ended question about their verdict decision were coded as described in Study 1a. Coders identified a total of seven categories. Interrater agreement was established at $\kappa > .80$. Two coders then coded all responses and discussed and came to an agreement on all disagreements. See the online Supplemental Materials for all categories.

Results

As in Studies 1a, 1b, and 2, we first computed participants' overall ratings of perceived cognitive competence, honesty, and suggestibility from their responses to the adapted Child Credibility Assessment Scale. We then analyzed the impact of our independent variables on perceived cognitive competence, honesty, and suggestibility with a series of 2 (delay: immediate, 2 years) × 3 (contextual information: no details, witness, investigation) ANOVAs. Similarly, we conducted a 2 (delay: immediate, 2 years) × 3 (contextual information: no details, witness, investigation) ANOVA on likelihood of guilt ratings and participants' confidence in their verdict decisions. Tukey's post hoc tests were used for follow-up analyses when appropriate. We also conducted a binary logistic regression to examine the impact of delay and reason for delay on dichotomous verdict decisions. See Tables 7 and 8 for descriptive information.

Reason for Delay

In addition to answering manipulation check questions for purposes of exclusion, participants who viewed a delayed report were also asked to confirm why the delay happened. This was done to determine whether our manipulation of investigation and witness delays worked as intended. Overall, participants seemed to agree with our interpretation of delays. Of participants in the witness-delay condition, 90.16% responded that the witness was the reason for the delay (4.92% responded that it was the police, and 4.92% responded that it was "other"). Of participants in the investigation-delay condition, 98.33% responded that the police were responsible for the delay (1.67% responded that it was the witness). For participants in the no-details condition, 67.21% indicated "other" for the reason behind the delay, 29.51% indicated that it was the fault of the police, and 3.28% indicated that it was the fault of the witness.

Cognitive Competence

We found no significant effect of delay, $F(1, 357) = 0.12$, $p = .73$, $\eta_p^2 = .0003$, or contextual information, $F(2, 357) = 0.56$, $p = .57$, $\eta_p^2 = .003$, on the perceived cognitive competence of the witness. Further, there was no significant interaction, $F(2, 357) = 0.70$, $p = .50$, $\eta_p^2 = .004$.

Honesty

There was a significant main effect of contextual information on ratings of perceived honesty, $F(2, 357) = 3.76$, $p = .02$, $\eta_p^2 = .02$. When the witness was described as reluctant, participants rated their perceived honesty as significantly lower compared with the no details condition ($p = .04$). Although not statistically significant, this difference between the witness and investigation conditions was trending toward significance ($p = .07$). When the witness was described as reluctant, participants rated their perceived honesty as lower compared with the investigation condition, although this pattern should be interpreted with caution. There was no significant difference between the investigation and no details conditions ($p = .97$). We did not find a significant main effect of delay, $F(1, 357) = 1.01$, $p = .31$, $\eta_p^2 = .003$, or a significant interaction, $F(2, 357) = 0.62$, $p = .54$, $\eta_p^2 = .003$.

Table 7
Study 3 Means, Standard Deviations, and Confidence Intervals of Credibility Ratings, Likelihood of Suspect Guilt, and Verdict Confidence

		Witness		Investigation		No details		Total	
Delay condition	Dependent variable	*M* (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]
Immediate	Cognitive competence	5.57 (0.89)	[5.35, 5.80]	5.68 (0.81)	[5.47, 5.89]	5.58 (0.82)	[5.50, 5.89]	5.61 (0.84)	[5.49, 5.73]
	Honesty	5.59 (0.97)	[5.35, 5.84]	5.77 (0.51)	[5.55, 5.99]	5.77 (0.84)	[5.55, 5.98]	5.71 (0.89)	[5.58, 5.84]
	Suggestibility	3.50 (0.97)	[3.17, 3.83]	3.21 (0.85)	[2.93, 3.49]	3.19 (0.84)	[2.91, 3.46]	3.30 (0.89)	[3.13, 3.47]
	Likelihood of suspect guilt	72.05 (23.35)	[66.07, 78.03]	75.92 (20.93)	[70.51, 81.32]	74.49 (21.65)	[68.95, 80.04]	74.14 (21.94)	[70.93, 77.35]
	Verdict confidence	65.61 (22.53)	[59.84, 71.38]	73.68 (20.52)	[68.38, 78.98]	69.36 (20.26)	[64.17, 74.55]	69.53 (21.27)	[66.42, 72.64]
Delay	Cognitive competence	5.49 (0.81)	[5.28, 5.70]	5.56 (0.77)	[5.36, 5.76]	6.59 (0.77)	[5.50, 5.89]	5.58 (0.78)	[5.47, 5.70]
	Honesty	5.34 (1.09)	[5.06, 5.62]	5.71 (1.00)	[5.45, 5.97]	5.77 (0.90)	[5.54, 6.00]	5.61 (1.01)	[5.46, 5.76]
	Suggestibility	3.50 (1.09)	[3.22, 3.77]	3.39 (1.00)	[3.09, 3.70]	3.37 (0.90)	[3.10, 3.65]	3.42 (1.01)	[3.26, 3.58]
	Likelihood of suspect guilt	68.70 (24.29)	[62.43, 74.97]	70.07 (22.00)	[64.38, 75.75]	71.61 (25.95)	[64.96, 78.25]	70.13 (24.04)	[66.61, 73.66]
	Verdict confidence	67.02 (22.60)	[61.18, 72.85]	68.87 (24.55)	[62.53, 75.21]	69.51 (22.92)	[63.64, 75.38]	68.47 (23.26)	[65.06, 71.88]
Total	Cognitive competence	5.53 (0.85)	[5.38, 5.69]	5.72 (0.79)	[5.48, 5.76]	5.64 (0.79)	[5.50, 5.78]	5.60 (0.81)	[5.51, 5.68]
	Honesty	5.47 (1.03)	[5.28, 5.65]	5.74 (0.93)	[5.57, 5.91]	5.77 (0.87)	[5.62, 5.93]	5.66 (0.95)	[5.56, 5.76]
	Suggestibility	3.50 (1.03)	[3.29, 3.71]	3.30 (0.93)	[3.10, 3.51]	3.28 (0.87)	[3.09, 3.47]	3.36 (0.95)	[3.24, 3.48]
	Likelihood of suspect guilt	70.39 (21.78)	[66.12, 74.67]	72.99 (21.59)	[69.09, 76.89]	73.05 (23.84)	[68.78, 77.32]	72.14 (23.07)	[69.76, 74.52]
	Verdict confidence	66.31 (22.48)	[62.26, 70.35]	71.28 (22.66)	[67.18, 75.37]	69.43 (21.54)	[65.57, 73.29]	69.00 (22.26)	[66.70, 71.30]

Note. Participants rated cognitive competence, honesty, and suggestibility on a scale from 1 to 7. Participants rated likelihood of suspect guilt and verdict confidence out of 100. CI = confidence interval.

Suggestibility

There was no significant effect of delay, $F(1, 357) = 1.09$, $p = .32$, $\eta_p^2 = .003$, or contextual information, $F(2, 357) = 1.37$, $p = .26$, $\eta_p^2 = .008$, on the perceived suggestibility of the witness. Further, there was no significant interaction, $F(2, 357) = 0.27$, $p = .76$, $\eta_p^2 = .002$.

Verdict

We conducted a binary logistic regression to examine the impact of delay and contextual information on dichotomous verdict decisions. We first included delay, contextual information, and an interaction in the model, and the overall model was not statistically significant, $\chi^2(5) = 9.73$, $p = .08$. When the interaction term was removed from the model, the overall model was significant, $\chi^2(3) = 9.70$, $p = .02$. See the online Supplemental Materials for full model information. Delay was a significant predictor of verdict decision ($p = .003$). There was a higher proportion of guilty verdicts when the report was given immediately after the crime, compared with after a delay of 2 years.

Likelihood of Guilt

We found no significant effect of delay, $F(1, 357) = 2.74$, $p = .10$, $\eta_p^2 = .008$, or contextual information, $F(2, 357) = 0.53$, $p = .59$, $\eta_p^2 = .003$, or significant interaction of delay and investigative details, $F(2, 357) = 0.14$, $p = .87$, $\eta_p^2 = .0008$, on likelihood of guilt ratings.

Confidence in Verdict

We found no significant effect of delay, $F(1, 357) = 0.21$, $p = .65$, $\eta_p^2 = .0006$, or contextual information, $F(2, 357) = 1.54$, $p = .22$, $\eta_p^2 = .009$, on the participants' confidence in their verdict decisions. There was also no significant interaction, $F(2, 357) = 0.66$, $p = .52$, $\eta_p^2 = .004$.

Reason for Verdict Decision

Participants' open-ended responses were coded into seven categories. We chose to focus on participants' first response, as in our prior studies, as we believed it to represent their most salient reason for their verdict decision. Participants primarily referenced the witness being credible (e.g., they seemed trustworthy, they provided a good number of details; 49.31%), needing more evidence (14.05%), or simply relying on the witness's testimony to make their decision (12.12%). Participants also mentioned the witness not being credible (e.g., they seemed to be lying, they did not seem to identify the accused; 7.99%), details of the crime itself (e.g., a threat was made, a theft took place; 9.64%), and the length of time since the crime happened (e.g., it happened today, it is been 2 years; 2.48%). The remaining participant responses were coded as "other" (4.41%).

Discussion

Study 3 explored how contextual information might impact a witness's perceived credibility and legal decisions. To further our understanding of this issue, we designed Study 3 to be fully crossed; that is, the same information was provided to participants in the immediate and delayed conditions (i.e., no additional information,

Table 8
Study 3 Frequencies (Percentage of Total Decision in Condition) of Guilty Verdicts

Delay condition	Witness	Investigation	No details	Total
Immediate	51 (83.61%)	51 (85.00%)	51 (83.61%)	153 (84.07%)
Delay	42 (70.00%)	44 (73.33%)	42 (68.85%)	128 (70.72%)
Total	93 (76.23%)	95 (79.17%)	93 (76.86%)	281 (77.41%)

the witness was reluctant, or it was not clear which police force had jurisdiction)—the only difference was whether this information co-occurred with a delay. We hypothesized, on the basis of our prior studies, that only when the delay was the result of the witness's behavior would it negatively impact their perceived cognitive competence and honesty (Hypothesis 1). This hypothesis was not supported in the present study. We found no significant impact of delay on the perceived cognitive competence or the honesty of the witness. Unexpectedly, we did find a main effect of contextual information on perceived honesty. In both the immediate and delay conditions, when the witness was described as reluctant, they were seen as less honest compared with the no details condition. In other words, it was not the delay that affected perceived honesty; it was the behavior of the witness.

We also hypothesized that any delay that was made salient with a reason would negatively impact legal decisions. Indeed, participants were more likely to make a convict decision when the report was immediate rather than delayed, consistent with Hypothesis 2. However, we found no significant effect of delay on likelihood of guilt ratings, contrary to Hypothesis 3.

General Discussion

Prior work regarding the perceived credibility of delayed reports has focused almost exclusively on delayed disclosure of sexual misconduct. One may conclude, on the basis of this literature, that a delay is detrimental to a witness's perceived credibility. The present work provides new and important insight that challenges this broad conclusion. We did not find a consistent negative effect of delay on the perceived credibility of a witness, contrary to the extant literature. Evaluating delayed reports is more complicated. The present studies lend support to the notion that a delay that is outside of the witness's control does not negatively impact the credibility of the witness and that it is the behavior of the witness, irrespective of delay, that might be most damaging to perceived credibility.

Prior work has demonstrated a consistent negative impact of delays on perceived credibility (e.g., Balogh et al., 2003; Franiuk et al., 2020; Miller et al., 2022). Importantly, in prior studies, the delay was a result of the witness's actions. These prior findings, combined with a null effect of delay when the delay was investigatory in Studies 1a and 1b, led us to theorize that the reason for a delay was important to decision makers. We hypothesized that only a delay as a result of the witness's actions would negatively impact their credibility; a delay as a result of the investigation would not have a negative impact. Our results from Study 2 supported this possibility.

In Study 2, we found a negative effect of delay on perceived credibility when the delay was caused by a witness who avoided attempts to be interviewed but not when the delay was caused by a stalled investigation. However, Study 2 described witness and investigation behaviors only when there was a delay. As astutely pointed out by a reviewer, this conflated delay and behavior. To disentangle these two variables, we made Study 3 a fully crossed design in which actions (witness reluctance or investigative complications) were present in both the immediate and delayed conditions. In Study 3, we found that witness reluctance diminished their perceived honesty in the delay and no delay conditions. This raises the question as to what is actually driving the negative effect of witness-related delays on perceived credibility. Is it the witness's behavior combined with the delay, as prior work may have indicated? Or is it the witness's behavior alone, irrespective of the delay, that hurts their perceived credibility? Our findings led us to speculate that the actions of the witness, rather than delay per se, are what individuals consider when making credibility decisions. Given that past work showing different patterns of results from our present studies has focused on sexual assault or misconduct (the present work focused on nonsexual assault), future research is necessary to evaluate this possibility in additional contexts.

It is possible that the present findings will not generalize to victims, as we focused on a witness to a crime, rather than a victim as in prior work. Although a witness displaying hesitation to immediately report a crime may be concerning to participants, such behavior may be more understandable for a victim. Therefore, the presently observed diminished credibility of a reluctant *witness* may not hold true across situations involving a reluctant *victim*. Perhaps for victims, reluctance harms credibility only in combination with a delay in reporting. Future research is necessary to understand potential differences in the perceived credibility of a reluctant witness compared with a reluctant victim.

It is also possible that there is something unique about witness behavior leading to a delay in a sexual crime context that differs from the current context, resulting in past findings of a delayed report diminishing credibility. However, for any crime, laypersons may focus on the behaviors of the witness, rather than the time since the crime, when evaluating their credibility. If it is, as we suspect, a witness's behavior, not the delay, that diminishes their credibility in the case of a delayed disclosure, this has important implications for professionals in the justice system. For example, lawyers may choose to provide explanations for a delay to reduce the negative impact of a witness-delay, whereas this may not prove necessary when the delay is outside of the witness's control.

Some prior work has demonstrated differing impacts of victim-related delays on perceived credibility (Thompson & Pozzulo, 2025). Thompson and Pozzulo (2025) demonstrated that a victim who delayed their report because they feared retaliation from the perpetrator or their family's response was seen as more truthful than a victim who delayed because they believed there was not enough evidence. However, Thompson and Pozzulo did not include an immediate condition with similar victim behavior. It is possible that the behavior seen as less credible from a witness after a delay would be equally diminishing to their credibility when there is no delay. Future research must take this possibility into account and include fully crossed designs when possible. Only when the same behaviors are presented in an immediate condition and a delay condition can we truly understand the independent and interactive effects of behavior and delay on perceived credibility.

Still, we have demonstrated that a delay that is outside of the witness's control, at least in the context provided here, does not

negatively impact a witness's perceived credibility. Thus, our findings indicate that participants, when evaluating the impact of a delay, appear to focus less on the potential impact of time on memory. If participants were focused on the impact of a delay on a witness's memory, we would have expected to see an impact of any delay on the perceived credibility of the witness and, specifically, on their cognitive competence. Given that we did not see this pattern, it is possible that participants either are not aware of or do not take into consideration declines in memory over time. Future research specifically exploring participants' beliefs about the impact of time on memory within the context of a delayed report of a crime is necessary to further understand this possibility. Researchers may also look to other work that has explored layperson beliefs about memory to help understand how laypeople view the impact of delay on memory (e.g., Benton et al., 2006; Simons & Chabris, 2011). From an applied perspective, lay individuals (i.e., jurors) might benefit from training and support on what to expect from memory reports after a delay as well as guidance on how to approach such evidence in evaluating for accuracy after a delay.

Our work also shows that a delay, both within and outside of the witness's control, can still have important implications for the case overall. Across all studies, we found that legal decisions made after a delay were more likely to favor the defendant (i.e., reduction of likelihood of guilt, fewer guilty verdicts). However, participants did not express that they considered the delay in their evaluation. Across all studies, only a small proportion of participants mentioned the delay (or lack thereof) when asked about their reasons for verdict decision. This could indicate that participants considered delay in their decision-making process, without being aware that delay mattered. Our findings have important implications for legal decision making. A delay can have a significant impact on a trier of fact's decisions, potentially without their awareness; this impact may not be communicated in their reasoning, reducing the possibility for safeguards (e.g., appeals) to act against potential bias. When possible, efforts should continue to be made to reduce delays in the justice system (e.g., as put forth in *R. v. Jordan*, 2016), to minimize potential negative impacts from a delay alone. If a delay is present in a case, additional safeguards to counteract negative assessments may be important. For example, jury instructions, expert testimony, or judicial education may prove beneficial to help inform triers of fact about potential bias.

Interestingly, in Study 3, we found a main effect of delay on dichotomous legal decisions but not on likelihood of suspect guilt ratings. One possible explanation for this finding is the likelihood threshold at which participants convict. If participants' likelihood of guilt ratings did not change across delay conditions, but convict decisions did, as in Study 3, we might speculate that participants used a different threshold to determine when a conviction was appropriate. For example, immediate condition participants may require a likelihood of guilt to be 80% or higher in order to convict, whereas delayed condition participants may require a likelihood of guilt to be 90% in order to convict. To investigate this possibility, we examined the average likelihood of guilt rating for participants who convicted for immediate and delay conditions across all studies, although this was not included in our preregistration. Participants who convicted in the immediate conditions had a slightly higher average likelihood of guilt rating across all studies ($M = 79.70$, $SD = 13.06$) than those in a delay condition (e.g., collapsed across 2 years and 15 years; $M = 77.97$, $SD = 14.68$), although this difference was not statistically significant, $t(856) = -1.83$, $p = .07$. This pattern might indicate a lower threshold for conviction for a delayed report, as the average likelihood of guilt rating was lower for participants who convicted in delayed conditions compared with immediate reports. If a lower threshold for a conviction was used in delay conditions, as indicated by the pattern described earlier, this may explain why in Study 3, we found a difference in conviction rates but not a difference in likelihood of guilt ratings. However, future research is necessary to better understand whether, and why, different thresholds exist.

Limitations

Studies 1a, 1b, and 2 had a primarily female sample. Some prior work on perceived credibility has demonstrated gender-related differences (e.g., Hockett et al., 2016). To control for this, we balanced gender in Study 3. We also acknowledge that the effect sizes found in the present work are small, which may limit the practical relevance of our findings. However, given that we showed consistent effects across all studies, these findings enhance our understanding of how witnesses are evaluated and decisions are made when a report of a crime is delayed. Further, even a small effect may be just enough to bring triers of fact closer to or away from conviction threshold, leading to a change in trial outcome. Thus, our findings have important applied implications.

In the present work, participants were presented with a small amount of information regarding the report of a crime via a video and were asked to make decisions. Triers of fact would receive additional information through a trial and investigation, observe a live witness, and face real consequences for their decisions (e.g., sentencing someone to prison) that were not present for the current participants. The current work offers an initial step into the investigation of reasons for delay on decision making, but future work imitating a real-world context (e.g., full trial) is necessary to better understand the findings.

We attributed our unexpected findings in Study 3, compared with the first studies, to the fully crossed design. The inclusion of a reluctant witness in an immediate condition demonstrated that perhaps participants are focused on the behavior of the witness, not the behavior of the witness in conjunction with a delay, when making credibility assessments. However, other differences in Study 3 may explain the unexpected findings. In the witness conditions in Study 3, the witness was described as being reluctant, whereas in Studies 1a, 1b, and 2, the witness was described as resisting attempts to be interviewed. The investigation reason provided in Study 3 was that there was confusion over which police force had jurisdiction rather than that a witness had not been identified, as in Studies 1a, 1b, and 2. Finally, our initial studies had a primarily female sample, whereas Study 3 was almost even across male and female participants.

Although a strength of the present work was the new contexts under which we explored delayed reports, our work differs from prior research that has focused on a sexual assault context. It is unknown whether the same findings regarding delays outside of a witness's control will replicate in a sexual assault context. Future research is necessary to explore this possibility. Further, prior research has primarily relied on exploring credibility evaluations of the alleged victim. Here, we explored perceptions of a witness.

There may be important differences in how the motivations of victims and witnesses are evaluated.

Conclusion

Overall, the set of studies provides novel insight into how delayed reports of a crime are assessed and challenges past assumptions regarding delayed reports. When a delay is outside of the witness's control, it does not appear to negatively impact their perceived credibility. However, the behavior of the witness does have the potential to negatively impact their perceived credibility, irrespective of delay. Any delay may also impact legal decisions, in a way that favors the defendant. This present work introduces an important new consideration into the discussion of how to anticipate effects of delays in criminal proceedings.

References

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53–56*(1), 63–70. https://doi.org/10.1016/j.jom.2017.06.001

Balogh, D. W., Kite, M. E., Pickel, K. L., Canel, D., & Schroeder, J. (2003). The effects of delayed report and motive for reporting on perceptions of sexual harassment. *Sex Roles*, *48*(7–8), 337–348. https://doi.org/10.1023/A:1022990530657

Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common sense: Comparing jurors, judges and law enforcement to eyewitness experts. *Applied Cognitive Psychology*, *20*(1), 115–129. https://doi.org/10.1002/acp.1171

Berman, G. L., & Cutler, B. L. (1996). Effects of inconsistencies in eyewitness testimony on mock-juror decision making. *Journal of Applied Psychology*, *81*(2), 170–177. https://doi.org/10.1037/0021-9010.81.2.170

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*(1), 79–88. https://doi.org/10.1007/BF01499074

Birdsey, E., & Snowball, L. (2013). *Reporting violence to police: A survey of victims attending domestic violence services*. NSW Bureau of Crime Statistics and Research. https://apo.org.au/sites/default/files/resource-files/2013-12/apo-nid37075.pdf

Brewer, N., Potter, R., Fisher, R. P., Bond, N., & Luszcz, M. A. (1999). Beliefs and data on the relationship between consistency and accuracy of eyewitness testimony. *Applied Cognitive Psychology*, *13*(4), 297–313. https://doi.org/10.1002/(SICI)1099-0720(199908)13:4<297::AID-ACP578>3.0.CO;2-S

Brodsky, S. L., Griffin, M. P., & Cramer, R. J. (2010). The Witness Credibility Scale: An outcome measure for expert witness research. *Behavioral Sciences & the Law*, *28*(6), 892–907. https://doi.org/10.1002/bsl.917

Bunting, L. A. (2014). Exploring the influence of reporting delay on criminal justice outcomes: Comparing child and adult reporters of childhood sexual abuse. *Journal of Child Sexual Abuse*, *23*(5), 577–594. https://doi.org/10.1080/10538712.2014.920457

Connolly, D. A., Chong, K., Coburn, P. I., & Lutgens, D. (2015). Factors associated with delays of days to decades to criminal prosecutions of child sexual abuse. *Behavioral Sciences & the Law*, *33*(4), 546–560. https://doi.org/10.1002/bsl.2185

Connolly, D. A., Coburn, P. I., & Chong, K. (2017). Twenty-six years prosecuting historic child sexual abuse: Has anything changed? *Psychology, Public Policy, and Law*, *23*(2), 166–177. https://doi.org/10.1037/law0000121

Connolly, D. A., Price, H. L., & Gordon, H. M. (2009). Judging the credibility of historic child sexual abuse complainants: How judges describe their decisions. *Psychology, Public Policy, and Law*, *15*(2), 102–123. https://doi.org/10.1037/a0015339

Danby, M. C., Sharman, S. J., & Klettke, B. (2022). Factors influencing the perceived credibility of children alleging physical abuse. *Psychiatry, Psychology and Law*, *29*(3), 456–470. https://doi.org/10.1080/13218719.2021.1917012

Eisen, M. L., & Lynn, S. J. (2001). Dissociation, memory and suggestibility in adults and children. *Applied Cognitive Psychology*, *15*(7), S49–S73. https://doi.org/10.1002/acp.834

Ellison, L., & Munro, V. E. (2009). Reacting to rape: Exploring mock jurors' assessments of complainant credibility. *British Journal of Criminology*, *49*(2), 202–219. https://doi.org/10.1093/bjc/azn077

Euale, J., & Turtle, J. (1999). *Police foundations: Interviewing and investigation*. Emond Montgomery.

Franiuk, R., Luca, A., & Robinson, S. (2020). The effects of victim and perpetrator characteristics on ratings of guilt in a sexual assault case. *Violence Against Women*, *26*(6–7), 614–635. https://doi.org/10.1177/1077801219840439

Fraser, B. M., Pica, E., & Pozzulo, J. D. (2022). The effect of delayed reporting on mock-juror decision-making in the era of #MeToo. *Journal of Interpersonal Violence*, *37*(13–14), NP11791–NP11810. https://doi.org/10.1177/0886260521997464

Gilbert, J. A. E., & Fisher, R. P. (2006). The effects of varied retrieval cues on reminiscence in eyewitness memory. *Applied Cognitive Psychology*, *20*(6), 723–739. https://doi.org/10.1002/acp.1232

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

Hockett, J. M., Smith, S. J., Klausing, C. D., & Saucier, D. A. (2016). Rape myth consistency and gender differences in perceiving rape victims: A meta-analysis. *Violence Against Women*, *22*(2), 139–167. https://doi.org/10.1177/1077801215607359

Kehn, A., Haugen, A., Warren, A., Schweitzer, K., & Nuñez, N. L. (2016). *The Child Credibility Assessment Scale: Validity and predictability in child sexual and non-sexual abuse cases* [Paper presentation]. American Psychology-Law Society Annual Conference, Atlanta, GA, United States.

Krix, A. C., Sauerland, M., Lorei, C., & Rispens, I. (2015). Consistency across repeated eyewitness interviews: Contrasting police detectives' beliefs with actual eyewitness performance. *PLOS ONE*, *10*(2), Article e0118641. https://doi.org/10.1371/journal.pone.0118641

London, K., Bruck, M., Wright, D. B., & Ceci, S. J. (2008). Review of the Contemporary Literature on How Children Report sexual abuse to others: Findings, methodological issues, and implications for forensic interviewers. *Memory*, *16*(1), 29–47. https://doi.org/10.1080/09658210701725732

Mason, K. A., & Benson, M. L. (1996). The effect of social support on fraud victims' reporting behavior: A research note. *Justice Quarterly*, *13*(3), 511–524. https://doi.org/10.1080/07418829600093071

Miller, Q. C., Call, A. A., & London, K. (2022). Mock jurors' perceptions of child sexual abuse cases: Investigating the role of delayed disclosure and relationship to the perpetrator. *Journal of Interpersonal Violence*, *37*(23–24), NP23374–NP23396. https://doi.org/10.1177/08862605221078812

Monroe, L. M., Kinney, L. M., Weist, M. D., Dafeamekpor, D. S., Dantzler, J., & Reynolds, M. W. (2005). The experience of sexual assault: Findings from a statewide victim needs assessment. *Journal of Interpersonal Violence*, *20*(7), 767–776. https://doi.org/10.1177/0886260505277100

Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLOS ONE*, *10*(7), Article e0120644. https://doi.org/10.1371/journal.pone.0120644

Oeberst, A. (2012). If anything else comes to mind … better keep it to yourself? Delayed recall is discrediting—Unjustifiably. *Law and Human Behavior*, *36*(4), 266–274. https://doi.org/10.1037/h0093966

Orbach, Y., Lamb, M. E., La Rooy, D., & Pipe, M. E. (2012). A case study of witness consistency and memory recovery across multiple investigative interviews. *Applied Cognitive Psychology*, *26*(1), 118–129. https://doi.org/10.1002/acp.1803

Peace, K. A., Sudra, R. Y., Forrester, D. L., Kasper, R., Harder, J., & Porter, S. (2015). Tall tales across time: Narrative analysis of true and false allegations. *Journal of Investigative Psychology and Offender Profiling*, *12*(2), 171–184. https://doi.org/10.1002/jip.1421

Peterson, C., & Whalen, N. (2001). Five years later: Children's memory for medical emergencies. *Applied Cognitive Psychology*, *15*(7), S7–S24. https://doi.org/10.1002/acp.832

Pozzulo, J. D., Dempsey, J. L., & Crescini, C. (2010). Factors affecting juror decisions in historic child sexual abuse cases involving continuous memories. *Criminal Justice and Behavior*, *37*(9), 951–964. https://doi.org/10.1177/0093854810373587

R. v. Jordan, 1 SCR 631 (CanLII) 2016.

Reich, C. M., Anderson, G. D., & Maclin, R. (2022). Why I didn't report: Reasons for not reporting sexual violence as stated on Twitter. *Journal of Aggression, Maltreatment & Trauma*, *31*(4), 478–496. https://doi.org/10.1080/10926771.2021.1912873

Ross, D. F., Jurden, F. H., Lindsay, R. C. L., & Keeney, J. M. (2003). Replications and limitations of a two-factor model of child witness credibility. *Journal of Applied Social Psychology*, *33*(2), 418–431. https://doi.org/10.1111/j.1559-1816.2003.tb01903.x

Rotenberg, C., & Cotter, A. (2018). *Police-reported sexual assaults in Canada before and after #MeToo, 2016 and 2017* (Catalogue No. 85–002-X). Statistics Canada. https://www150.statcan.gc.ca/n1/pub/85-002-x/2018001/article/54979-eng.htm

Schmechel, R. S., O'Toole, T. P., Easterly, C., & Loftus, E. F. (2006). Beyond the ken? Testing jurors' understanding of eyewitness reliability evidence. *Jurimetrics*, *46*, 177–214.

Simons, D. J., & Chabris, C. F. (2011). What people believe about how memory works: A representative survey of the U.S. population. *PLOS ONE*, *6*(8), Article e22757. https://doi.org/10.1371/journal.pone.0022757

Standing Senate Committee on Legal and Constitutional Affairs. (2017). *Delaying justice is denying justice: An urgent need to address lengthy court delays in Canada*. https://sencanada.ca/content/sen/committee/421/LCJC/reports/Court_Delays_Final_Report_e.pdf

The Current. (2015, September 10). *Robert Pickton case torments former detective Lori Shenher*. Canadian Broadcast Corporation. https://www.cbc.ca/news/canada/british-columbia/robert-pickton-case-torments-former-detective-lori-shenher-1.3223349

Thompson, L. E., Pica, E., & Pozzulo, J. (2021). Jurors' decision making in a sexual assault trial: The influence of victim age, delayed reporting, and multiple allegations. *American Journal of Forensic Psychology*, *39*, 19–46.

Thompson, L. E., & Pozzulo, J. (2025). How length of and reason for delayed reporting influence mock-jurors' judgments in a sexual assault trial. *Journal of Police and Criminal Psychology*, *40*(2), 386–397. https://doi.org/10.1007/s11896-024-09664-z

Tsang, W. H. W., Chan, T. M. S., & Cheung, M. (2021). Chinese male survivors of intimate partner violence: A three-pillar approach to analyze men's delayed help-seeking decisions. *Violence and Victims*, *36*(1), 92–109. https://doi.org/10.1891/VV-D-19-00129

Wells, C., Morrison, C. M., & Conway, M. A. (2014). Adult recollections of childhood memories: What details can be recalled? *Quarterly Journal of Experimental Psychology*, *67*(7), 1249–1261. https://doi.org/10.1080/17470218.2013.856451

Whitman, J., & Davis, R. (2007). *Snitches get stitches: Youth, gangs, and witness intimidation in Massachusetts*. National Center for Victims of Crime. https://justiceresearch.dspacedirect.org/server/api/core/bitstreams/8fc796b7-2551-4a4b-b979-98052a338a92/content

Appendix

Credibility Assessment Scale

Factor	Question
Cognitive competence	Ability as a witness
	Ability to remember and answer questions
	Accuracy in describing/reporting events
	Adequacy in recounting events
	Communication skills
	Ability to recall events
	Reliability of memory
Honesty	Honesty
	Likelihood of reporting things that did not really happen
	Likelihood of lying
	Likelihood of making up the event
Suggestibility	Likelihood that accuracy is affected by stress
	Likelihood of being influenced by individuals' questions
	Likelihood of forgetting to report things that really happened
	Likelihood of being misled by the individual asking questions

Received August 30, 2023
Revision received April 11, 2025
Accepted April 21, 2025 ■