



UNIVERSITY OF
GLOUCESTERSHIRE

This is a peer-reviewed, final published version of the following in press document and is licensed under Creative Commons: Attribution 4.0 license:

**Zhang, Yikang, Segal, Aleksandr, Pompedda, Francesco
ORCID: 0000-0001-9253-0049, Haginoya, Shumpei and
Santtila, Pekka (2022) Confirmation bias in simulated CSA
interviews: How abuse assumption influences interviewing
and decision-making processes? Legal and Criminological
Psychology. doi:10.1111/lcrp.12213 (In Press)**

Official URL: <http://doi.org/10.1111/lcrp.12213>

DOI: <http://dx.doi.org/10.1111/lcrp.12213>

EPrint URI: <https://eprints.glos.ac.uk/id/eprint/10792>

Disclaimer

The University of Gloucestershire has obtained warranties from all depositors as to their title in the material deposited and as to their right to deposit such material.

The University of Gloucestershire makes no representation or warranties of commercial utility, title, or fitness for a particular purpose or any other warranty, express or implied in respect of any material deposited.

The University of Gloucestershire makes no representation that the use of the materials will not infringe any patent, copyright, trademark or other property or proprietary rights.

The University of Gloucestershire accepts no liability for any infringement of intellectual property rights in any material deposited but will remove such material from public view pending investigation in the event of an allegation of any such infringement.

PLEASE SCROLL DOWN FOR TEXT.

ORIGINAL ARTICLE

Confirmation bias in simulated CSA interviews: How abuse assumption influences interviewing and decision-making processes?

Yikang Zhang^{1,2}  | Aleksandr Segal³  | Francesco Pompemma⁴  |
Shumpei Haginoya⁵  | Pekka Santtila^{2,6} 

¹Maastricht University, Maastricht, The Netherlands

²NYU-ECNU Institute for Social Development, Shanghai, China

³Mykolas Romeris University, Vilnius, Lithuania

⁴University of Gloucestershire, Cheltenham, UK

⁵Meiji Gakuin University, Tokyo, Japan

⁶New York University Shanghai, Shanghai, China

Correspondence

Yikang Zhang, Faculty of Psychology and Neuroscience, Maastricht University, Universiteitssingel 40, 6229 ER, Maastricht, The Netherlands.

Email: kang.y.zhang@outlook.com

Abstract

Purpose: Research has shown that confirmation bias plays a role in legal and forensic decision-making processes and, more specifically, child interviews. However, previous studies often examine confirmation bias in child interviews using non-abuse-related events. We enrich the literature by examining interviewers' behaviours in simulated child sexual abuse (CSA) cases.

Method: In the present study, we used data from a series of experiments in which participants interviewed child avatars to examine how an assumption of abuse based on preliminary information influenced decision-making and interviewing style. Interview training data ($N_{\text{interview}} = 2084$) from eight studies with students, psychologists and police officers ($N = 377$) were included in the analyses.

Results: We found that interviewers' preliminary assumption of sexual abuse having taken place predicted 1) a conclusion of abuse by the interviewers after the interview; 2) higher confidence in their judgement; 3) more frequent use of not recommended question types and 4) a decreased likelihood of reaching a correct conclusion given the same number of available relevant details.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Legal and Criminological Psychology* published by John Wiley & Sons Ltd on behalf of British Psychological Society

Conclusion: The importance of considering how preliminary assumptions of abuse affect interview behaviour and outcomes and the implications for the training of investigative interviewers were discussed.

KEYWORDS

child interviewing, child sexual abuse, confirmation bias, emotions, mega-analysis

INTRODUCTION

It has long been known that people have a tendency to confirm their hypotheses rather than to falsify them. While this is an effective strategy in everyday life where heuristics help us in dealing with a huge amount of information, such confirmation bias can have deleterious effects in some contexts (Mynatt et al., 1977; Nickerson, 1998). This bias has also been examined in the legal and forensic settings, showing, for example, that trait confirmation bias was negatively associated with evaluator-rated interview performance among police officers who were receiving interview training (Melinder et al., 2020). Compared with exonerating tactics, interrogators who presume the suspect to be guilty tend to use more guilt-confirming tactics (Rassin et al., 2010). Mock jurors' interpretation of the evidence is biased by prior beliefs (Carlson & Russo, 2001). Moreover, the effects of confirmation bias also extend to aspects of forensic science that have been previously thought to be immune from subjectivity such as fingerprint identification as discussed by Kassin et al. (2013).

Limited studies have also examined how confirmation bias manifests itself when interviewing not the suspect but the alleged victim. For many crimes, it is usually clear at the beginning of the investigation that a crime has occurred. However, it is less so in cases of (sexual) assault and particularly child sexual abuse. When a suspected child sexual abuse case is reported, whether the abuse has actually happened is to be determined during the investigation. This makes the investigative interview in child sexual abuse (CSA) cases central. Even though extensive research has been carried out to improve child interview quality, confirmation bias has not received sufficient attention (for a recent review, see O'Donohue & Cirlugea, 2021). Moreover, child sexual abuse is an emotionally charged topic for both investigating professionals (Magnusson et al., 2021) and the public (e.g. McCartan et al., 2015), which could, potentially, amplify the potential negative effects of confirmation bias.

Confirmation bias in child interviews has been examined using experimental designs where children have been interviewed about mock events they have participated in. The results have shown that interviewers' prior experiences and presumptions bias their questioning leading to the children making false statements and the interviewers drawing incorrect conclusions (Bruck et al., 1999, as cited in O'Donohue & Cirlugea, 2021; Quas et al., 2007). A study examining police officers also found that biased information led to fewer open questions and more leading questions (Powell et al., 2012). However, the negative result was limited to interviewers not adhering to the practice of open questions. An understandable limitation of these studies is that the mock events that the child participants have experienced are neutral and unrelated to child sexual abuse. However, Koppelaar et al. (1997) found that manipulating victim credibility beforehand affected interviewing behaviour and interpretation of results in adult sexual abuse scenarios, but evidence that is more directly relevant for CSA interviews is needed. Moreover, even though past research has shown that confirmation bias affects investigative and judicial decision-making, no one has yet examined how different preliminary assumptions can influence the magnitude of confirmation bias.

In light of this, the current study aimed at examining how confirmation bias manifests itself in simulated CSA interviews and how the assumption of abuse can influence the magnitude of confirmation bias in such a context.

Confirmation bias

Confirmation bias refers to a universal tendency to seek or interpret evidence in a way that is consistent with one's pre-existing hypothesis (Frey, 1986; Nickerson, 1998). This tendency is more prominent in sequential information search, where pieces of information arrive one at a time, compared with situations where all information is given simultaneously. This is due to the decision-focus (rather than information focus) under sequential information search. Specifically, in such situations, partial information is retrieved and then processed before new information is searched for. This sequence repeats until the need for information is satisfied or a decision has been made. Therefore, when a new piece of information is discovered, the decision-maker can either heed or ignore it (Jonas et al., 2001). Whereas focussing on all available pieces of information reduces preference for supporting information, focussing on the decision increases it (Jonas et al., 2008). In addition, people also show greater confirmation bias when the information is limited (Fischer et al., 2005), which exactly corresponds to the typical scenario in alleged CSA investigations.

Even though the operational definitions of confirmation bias vary considerably between studies that have investigated it, we suggest that confirmation bias manifests itself in investigative interviews mainly in two processes, that is, information searching (Rassin et al., 2010) and interpretation of evidence (Carlson & Russo, 2001). During the information searching stage, people tend to use positive hypothesis testing to gather information (Klayman, 1995), for example, if an interviewer is investigating if someone was abused, the interviewer might ask a question that contains suggestive information related to the presence of an abuse ('Did he touch you?' or 'Where did he touch you?'), instead of asking more neutral questions (e.g. 'Tell me what happened'). This kind of questioning can be categorized as option-posing or suggestive (e.g. Orbach et al., 2000; Pompedda et al., 2015), and it may sometimes even lead to the elicitation of wrong details from the interviewee and negatively impact interview outcome (e.g. Brown et al., 2013; Bruck et al., 1999, as cited in O'Donohue & Cirlugea, 2021).

In the process of interpreting evidence, confirmation bias may manifest itself by selectively processing information, that is, interpreting ambiguous evidence in accordance with the individual's own hypothesis or belief, and ignoring evidence that might disconfirm the initial hypothesis (Klayman, 1995). A recent study has shown that evidence consistent with the observer's initial choice is processed more efficiently compared to inconsistent ones (Talluri et al., 2018). Research in forensic settings has also shown that both confession evidence (Kukucka & Kassin, 2014) and a pre-existing hypothesis (O'Brien, 2009) can bias the interpretation of evidence and influence subsequent decisions. In light of this, we expected that when given the same amount of evidence, those who exhibit greater confirmation bias might be less likely to reach a correct conclusion.

Assumption of abuse, metacognition, and confirmation bias

CSA interviews can be an emotional experience for the child and the interviewer. In a recent study specifically examining the interviewer's emotional responses to narratives of alleged cases of child sexual abuse, it was found that abuse narratives induced strong feelings of sadness, anger and disgust in participants whereas narratives of non-abuse induced feelings of relief (Segal & Santtila, 2022). These emotions (anger and disgust in particular) associated with the assumption of abuse have been shown to have an impact on decision-making via metacognitive appraisals (Briñol & DeMarree, 2012; Briñol et al., 2007, 2018; Huntsinger et al., 2012) and may influence the interviewer's attitudes during interviewing (Oxburgh et al., 2006). Anger and disgust are associated with increased confidence in predetermined thoughts and greater reliance on these thoughts when making judgements while other emotions such as surprise and awe are associated with doubt (Briñol et al., 2018). Indirect evidence shows that anger as a result of perceived injustice may also become a motivating factor resulting in punitive attitudes (Hartnagel & Templeton, 2012). Therefore, interviewers in a CSA interview who have formed a preliminary assumption of abuse (vs. no abuse) could exhibit greater confirmation bias.

THE CURRENT RESEARCH

Utilizing published data, the current research examined how confirmation bias manifests itself in the process of investigative interviews.

Hypotheses

The presence of confirmation bias

In the context of the CSA interview, we expected the interviewer's preliminary assumption about what had happened (formulated based on a brief scenario before the interview) to predict their conclusion after the interview (H1).

Confirmation bias driven questioning

Past research has shown that interviewers exposed to biased information use more leading questions and fewer open-ended questions (O'Donohue & Cirlugea, 2021; Powell et al., 2012). If the preliminary assumption was incorrect, it is less likely that the interviewers would elicit details from the avatar that would confirm their preliminary assumption. As a result, the interviewers may thus seek evidence to confirm their assumptions using leading or suggestive questions. Therefore, we expected that confirmation bias would drive those who had the wrong preliminary assumption (i.e. assumed abuse when there was no abuse or vice versa) to use more not recommended questions (e.g. option-posing and suggestive questions see Table 1, H2.1) and an increase in wrong details (H2.2).

Assumption of abuse and enhanced confirmation bias

As the assumption of abuse is associated with experiences of anger and disgust, which have been shown to increase the reliance on previous thoughts (Briñol et al., 2018), we expected that if the interviewer's preliminary assumption was that there had been sexual abuse, they would be more confident in their judgement. (H3.1) We also expected that if the interviewer's preliminary assumption was that there had been sexual abuse, they would show more confirmation bias as evidenced by the use of more not recommended questions (H3.2), the elicitation of more wrong details (H3.3), and a lower rate of reaching the correct conclusion (H3.4).

Besides affecting the hypothesis-testing behaviour of the interviewer, we expected confirmation bias to also manifest itself in the interpretation of evidence. That is, we expected those who show greater confirmation bias to underuse relevant information when the information contradicts their preliminary assumption. In the current context, we expected that those, whose preliminary assumption was one of sexual abuse, will use relevant information (defined as relevant details uncovered during the interview that revealed if the abuse happened or otherwise) to a lesser extent (H4).

METHOD

Participants

The eight studies included in this analysis collected training data ($N_{\text{interview}} = 2084$) using participant samples of European and Japanese students, psychologists and police officers ($N = 377$). Detailed information regarding all of the samples can be found in Table 2.

TABLE 1 Question-type coding used in the studies

Category	Definition
Recommended Questions	
Facilitators	Open-ended and non-suggestive questions that encourage the child to continue with the previous answer
Invitations	Open-ended and non-suggestive questions. They are broad and let the child talk freely
Directive	Open-ended and non-suggestive questions that focus the child attention on a previously mentioned detail asking for a specific explanation
Not Recommended Questions	
Option-posing	Closed-ended questions that focus on unmentioned detail (without implying a particular type of response) or on a mentioned detail asking the child to provide a yes/no answer
Specific suggestive	Open-ended or closed-ended questions that are based on an unmentioned detail and express the expected response
Unspecific suggestive	Open-ended or closed-ended questions that are not based on an unmentioned detail but express the expected response
Repetitions	Repetitions of a previously recommended or not recommended question
Too-long/Unclear	Questions that use a logical structure that is too complicated for the cognitive level of the child and/or are formulated in a haphazard manner and/or contains more than one concept at the time
Multiple choice	Questions that provide a predetermined list which the child is requested (explicitly or implicitly) to pick from
Time	Open-ended or closed-ended questions that require the child to provide or recollect precise time-related information
Fantasy	Open-ended or closed-ended questions that move the discussion from the reality to the fantasy level
Feelings	Open-ended or closed-ended questions that require the child to provide accounts regarding own or other's feelings

Interview simulations

The simulated avatar interviews used in all of the experiments contained in the present study were created as an alternative to theoretical training to improve the quality of investigative interviews in alleged CSA cases (Santtila et al., 2012). This kind of training, referred to as serious gaming, promotes the acquirement and improvements of complex practical skills in a cost-effective way (Wouters et al., 2013) and is potentially more accessible to practitioners.

During the interview simulations, the participants first read a background scenario explaining why sexual abuse was suspected (See Appendix A in supplementary materials). Before interviewing the avatar, participants were asked their preliminary idea (see below) and the percentage of certainty they hold for their response. After this, the participants interviewed a computerized avatar using natural spoken language. Each avatar has predefined memories associated with the alleged abuse (nine details in total), which can either reveal abuse or provide an innocent explanation. In addition, each avatar also has predefined memories that are unrelated to the suspicion (also 9 details in total), for example, the child's family circumstances, and other irrelevant details related to, for example, favourite games or food. When the interviewer asks a question, a trained human operator categorizes the question according to the scheme in Table 1. This is followed by probabilistic response algorithms associated with the question types determining the avatars' response. The avatar gives a response containing relevant

TABLE 2 Descriptive statistics of the seven data sets included in the analysis

Study	Experimental conditions	Rounds of interviews	Participant			
			Population	n_{Female}/N	M_{age}	SD_{age}
*Pompedda et al. (2017)	Control ($n = 12$) Outcome Feedback ($n = 12$) Process Feedback ($n = 12$) Feedback ($n = 12$)	4	Europe; Students	38/48	27.9	9.1
*Krause et al. (2017)	Control ($n = 19$) Feedback ($n = 19$) Feedback +Reflection ($n = 21$)	8	Europe; Students	35/59	24.4	3.7
Haginoya et al. (2020)	Control ($n = 15$) Feedback ($n = 17$)	6	Japan; Students	23/32	20.5	0.6
Pompedda et al. (2020) Study 1	Control ($n = 20$) Feedback ($n = 20$)	6	Europe; Psychologists	37/40	27.4	2.2
*Pompedda et al. (2020) Study 2	Control ($n = 32$) Feedback ($n = 32$)	6	Europe; Students	44/64	23.1	3.6
Haginoya et al. (2021)	Modelling ($n = 11$) Feedback ($n = 10$) Feedback +Modelling ($n = 11$)	5	Japan; Psychologists	22/32	35.1	8.7
*Kask et al. (2022)	Control ($n = 11$) Control+Feedback ($n = 11$)	4/8	Europe; Police	3/22	41.2	6.2
Segal & Santrila (2022)	Control ($n = 20$) Feedback ($n = 20$) Supportive ($n = 20$) Feedback +Supportive ($n = 20$)	4	Japan; Mixed	53/80	35.6	9.9

or neutral narrative details if the interviewer asks recommended (i.e. open) questions. To increase the realism of the simulation, the operator can also manually provide additional narrative details unrelated to the suspected abuse labelled side details (Pompedda, 2018).

If the interviewer asks not recommended (i.e. closed) questions about the alleged abuse (such details can be either present or absent from the avatar's memory), they could obtain wrong details, the probability of which is determined by the algorithms associated with these question types. For example, if the interviewer asks an option-posing question, and the avatar's memory contains the answer to the question (the correct answer is 'Yes'), the probability of the incorrect response 'No' would be 35%. If the option-posing question is about a detail not present in the avatar's memory (the correct answer is 'I don't know'), the probability of a wrong detail would be 80% (50% for a 'Yes' and 30% for a 'No' response). These probabilities have been taken from previous experimental studies assessing children's suggestibility and memory in interviews.

After each interview, participants provided a final decision about the presence or absence of abuse and an explanation of what happened. In the experimental conditions, participants receive feedback on the appropriateness of the questions they used (i.e. process feedback for two recommended and two not recommended questions used during the interview) and/or on the interview outcome (i.e. if they correctly determined the abuse status of the avatar) while those in the control condition did not receive any feedback. The process feedback would change for each new category used by the interviewer in each interview. For example, if the interviewer asked option-posing and multiple-choice questions in the first interview and, together with the first two categories, also ask specific and unspecific suggestive questions in the second interview, the priority was given to the latter two categories (the same applied for recommended questions). In this way, the feedback was always tailored to the questioning style of the interviewer. Each participant conducted 4–8 rounds of interviews depending on the conditions and study design (see Table 2). All of the coded questions from the interviewers and the answers from the avatars then were extracted for further study.

Data coding

Recommended and not recommended questions

In all the studies included in the present report, recommended and not recommended questions were coded as continuous variables with the value indicating the number of questions asked in an interview. For the categorization of questions, see Table 1 (Haginoya et al., 2021; Pompedda et al., 2015). Not all studies reported interrater reliability. Krause et al. (2017) reported that the percentage of agreement between raters was 80%, with $\kappa = .68$; Pompedda et al. (2020) reported 87% agreement, with $\kappa = .76$ for Study 1, and 76% agreement, with $\kappa = .70$ for Study 2; Haginoya et al. (2020), Haginoya et al. (2021) reported percentage agreement values of 74–80% and κ values of .63–.72, all showing adequate reliability.

Relevant, neutral and wrong details

Relevant, neutral and wrong details were coded as continuous variables with the value indicating the number of details elicited from the avatar in an interview. For one hundred and twenty interviews (or 15 participants) out of 2208 interviews, the number of wrong details was not recorded at the time of the interview (Krause et al., 2017) and is not included in the data set.

Preliminary assumption about the alleged CSA case

After reading the background scenario and before the interview, participants were asked the following question to probe their preliminary assumption: 'Please, tell me your preliminary idea about what

happened' and the different possible options were provided. They were also asked to indicate their confidence in their judgement. In the first study conducted by Pompedda et al. (2017), the preliminary assumption of whether there was abuse was coded ordinally (i.e. 'Yes', 'Maybe yes', 'Don't Know', 'Maybe no', 'No') and the confidence was measured from 1 to 100. Two studies conducted in Estonia used three categories (Abuse vs. I do not have enough information vs. No Abuse) to log the preliminary assumption (Study 2 from Pompedda et al., 2020; Kask et al., 2022). While two categories (Abuse vs. No Abuse) were used in the remaining studies (Krause et al., 2017; Study 1 from Pompedda et al., 2020; Haginoya et al., 2020, 2021; Haginoya & Santtila, 2022). In the current analysis, we recorded the ordinal data so that 'Yes' and 'Maybe Yes' correspond to 'Abuse' while 'No' and 'Maybe No' correspond to 'No abuse'. The 'I don't know', or 'I do not have enough information' category was recoded as 'Not Sure'. All other studies except Pompedda et al. (2017) measured confidence in judgement using a scale from 50 to 100. As the first study in Pompedda et al. (2017) used a different metric to measure confidence, we standardized confidence ratings based on the study before doing analyses.

Conclusion about the alleged CSA case

After conducting the interview, the participants had to conclude whether there had been abuse and indicate their confidence in the conclusion. In all studies but the two from Estonia, the question was asked in dichotomous format (Abuse vs. No Abuse). Two studies conducted in Estonia used three categories (Abuse vs. No Conclusion vs. No Abuse). Only half the data sets contained the measure of confidence in conclusion (marked with an asterisk in Table 2).

Conclusion correctness

Conclusion correctness refers to whether the conclusion about the presence (vs. no presence) CSA was correct, that is, corresponding to the memory of the avatar being interviewed.

Statistical analyses

All statistical analyses were conducted in R (version 4.1.0, R Core Team, 2021). We first employed correlational analysis to ensure that the coding and combination of data sets were correct. Then, we used lme4 (Bates et al., 2014) to perform a series of (generalized) linear mixed-effects analyses to examine the presence of confirmation bias in CSA interviews and the moderating factors. For random effects in the models, we had intercepts for participants nested in studies. Confidence intervals (95%) of the parameters in the linear mixed models and the generalized linear mixed models were calculated using the profile method.

RESULTS

The correlation between interview quality indicators

The correlations between quality indicators suggested that the combined data set was valid (see the correlation matrix in the supplementary materials). The number of recommended questions was positively correlated with the number of relevant details elicited ($r = .79, p < .001$), the number of neutral details elicited ($r = .79, p < .001$) and negatively correlated with the number of wrong details elicited ($r = -.25, p < .001$). The number of recommended questions was also positively correlated with conclusion correctness ($r = .20, p < .001$). The number of not recommended questions was

negatively correlated with the number of relevant details elicited ($r = -.19, p < .001$), the number of neutral details elicited ($r = -.19, p < .001$) and positively correlated with the number of wrong details elicited ($r = .57, p < .001$). The number of not recommended questions was weakly negatively correlated with conclusion correctness ($r = -.06, p = .004$). Both the number of relevant details ($r = .23, p < .001$) and neutral details elicited ($r = .19, p < .001$) were positively correlated with conclusion correctness and the number of wrong details weakly negatively predicted conclusion correctness ($r = -.07, p = .004$).

The presence of confirmation bias

First, a proportion test was conducted to examine whether people tend to form a preliminary assumption of abuse (vs. no abuse). Among the 2,006 interviews where the interviewers formulated an assumption before the interview process (i.e. cases where an assumption of abuse or no abuse was formed), 1,163 interviews reported an assumption of abuse (proportion = 0.58, 95% CI [0.55, 0.61]), $\chi^2(1) = 50.73, p < .001$.

Results showed that this preliminary assumption formulated before the interview predicted the conclusion reached after the interview, even after controlling for the ground truth in the interview scenario (i.e. whether in the scenario the avatar was abused or not) (see Table 3). As there were three responses, Abuse, No Abuse and No Conclusion, we dummy coded the conclusion in two ways (i.e. Conclusion of Abuse vs. other responses and Conclusion of No Abuse vs. other responses) and ran two models with these two dummy variables as dependent variables. Compared with those whose preliminary assumption was that there was no abuse, those who believed there was abuse were more likely to reach a conclusion of abuse and less likely to conclude no abuse. H1 was supported.

Effects of preliminary assumption correctness on question use and details elicited

To examine whether the correctness and the type of preliminary assumption (sexual abuse vs. no abuse) had an impact on question use and the details elicited in the interviews, we performed a series of random intercept models with the assumption correctness as fixed effect ($N = 1970$, with the exception for the model of wrong details, which was 1850). Cases, where participants did not reach a preliminary assumption, were excluded from the analyses. Results showed that if the preliminary assumption did not match the ground truth (vs. matched), the interviewers asked fewer recommended questions ($B = -0.06, SE = 0.01, 95\% CI [-0.08, -0.03]$) and not recommended questions ($B = -0.03, SE = 0.01, 95\% CI [-0.06, -0.01]$). They also obtained significantly fewer neutral details ($B = -0.06, SE = 0.02, 95\% CI [-0.11, -0.01]$) and fewer wrong details ($B = -0.08, SE = 0.03, 95\% CI [-0.14, -0.01]$). No significant

TABLE 3 Preliminary assumption predicted conclusion

Predictors	Conclusion of abuse (N= 2065)			Conclusion of no abuse (N= 2065)		
	B	SE	95% CI	B	SE	95% CI
(intercept)	-1.16	0.12	[-1.40, -0.92]	1.03	0.14	[0.73, 1.33]
Assumption of Abuse (vs. No Abuse)	1.39	0.10	[1.19, 1.60]	-1.42	0.11	[-1.63, -1.21]
No Clear Assumption (vs. No Abuse)	0.49	0.30	[-0.11, 1.07]	-0.59	0.30	[-1.18, -0.004]
Avatar Actually Abused (vs. Not)	0.64	0.10	[0.45, 0.83]	-0.70	0.10	[-0.89, -0.51]

Note: Avatar Actually Abused (vs. Not) refers to whether the avatar being interviewed has a memory of sexual abuse (i.e. the ground truth in the simulated interview).

All significant predictors in the model are in bold.

Only fixed effects are presented.

difference was found regarding the propensity of eliciting relevant details ($B = -0.04$, $SE = 0.02$, 95% CI $[-0.08, 0.01]$). H2.1 and H2.2 were therefore not supported.

Effects of assuming abuse on confidence of the assumption and subsequent conclusion as well as question use and details elicited

Even though a confirmation bias means that preliminary assumptions predict conclusion reached no matter what the content of the assumption is, because of the emotion-laden nature of suspected sexual abuse, we expected a stronger confirmation bias in those whose preliminary assumption was there had been sexual abuse. More specifically, we predicted that those who held a preliminary assumption/ conclusion of sexual abuse would be more confident in this preliminary assumption and in their subsequent conclusion, use a more positive hypothesis-testing technique (i.e. more not recommended questions), elicit more wrong details, and as a result, would be more likely to reach the wrong conclusion.

To test the hypotheses, we examined whether believing there was a CSA case would increase the participants' confidence in their preliminary assumption ($N = 2006$) and post-interview conclusion ($N = 999$). Two linear mixed models with preliminary assumption/post-interview conclusion as fixed effect were run to predict the confidence of the judgement. As for random effects, we had random intercepts for participant ID nested in the Study. Results showed that if the preliminary assumption was that there was abuse (vs no abuse), the participants' confidence in this assumption was higher ($B = 0.25$, $SE = 0.04$, 95% CI $[0.17, 0.33]$). Similarly, a conclusion of abuse also led to an increase in confidence compared with a conclusion of no abuse ($B = 0.35$, $SE = 0.06$, 95% CI $[0.24, 0.47]$). H3.1 was supported.

To examine whether the type of preliminary assumption (sexual abuse vs. no abuse vs. not sure) had an impact on question use and the details elicited in the interviews, we performed a series of random intercept models with the type of assumption as fixed effect. If the preliminary assumption was that sexual abuse had happened, the interviewers tended to use more not recommended questions compared with those whose preliminary assumption was there was no abuse ($B = 0.06$, $SE = 0.01$, 95% CI $[0.04, 0.08]$). Having a preliminary assumption of abuse or not having a clear assumption did not differ from having a preliminary assumption of no abuse in terms of recommended question use, as indicated by the non-significant estimates of *preliminary assumption of abuse vs. no abuse* ($B = 0.02$, $SE = 0.01$, 95% CI $[-0.01, 0.05]$) and *not sure vs. no abuse* ($B = 0.03$, $SE = 0.03$, 95% CI $[-0.04, 0.10]$). A preliminary assumption of sexual abuse (vs no abuse) also led to more wrong details being elicited ($B = 0.14$, $SE = 0.04$, 95% CI $[0.07, 0.21]$), possibly as a result of the interviewers in this situation using more not recommended questions (detailed results can be found in the supplementary materials). H3.2 and H3.3 were supported.

Preliminary assumption content moderated the utilization of information

Built on literature examining interviewers' emotional response to child sexual abuse (Segal & Santtila, 2022; Magnusson et al., 2021) and basic emotion and cognition research (Briñol & DeMarree, 2012; Briñol et al., 2007, 2018), we reasoned that those whose preliminary assumption was one of sexual abuse would also show greater confirmation bias. In terms of selective information processing, they would be less likely to fully make use of relevant details to reach a correct conclusion. Information utilization was here defined as the predictive power of the number of relevant details on reaching the correct conclusion. A generalized linear mixed model with preliminary assumption (abuse vs. no abuse), the number of relevant details elicited, and their interaction terms as fixed effects was performed. Compared with those who thought there was no abuse, interviewers who assumed that there had been sexual abuse utilized the relevant information to a lesser extent, as indicated by

the negative interaction between assumption of abuse and number of relevant details ($B = -0.09$, $SE = 0.04$, 95% CI $[-0.16, -0.02]$), supporting H4. Next, we ran separate models with the number of relevant details as fixed effect for the assumption-of-abuse and the assumption-of-no abuse samples. Results showed that in both samples, number of relevant details positively predicted the probability of reaching the correct conclusion, with the predictive power being stronger in the assumption-of-no abuse sample ($B = 0.21$, $SE = 0.03$, 95% CI $[0.16, 0.27]$, $Pseudo R^2 = .11$) compared with the assumption-of-abuse sample ($B = 0.13$, $SE = 0.02$, 95% CI $[0.09, 0.17]$, $Pseudo R^2 = .04$). We also ran robustness check models with the number of wrong details included as covariates; the results held (see supplementary materials).

What is more, on average, compared with those who presumed no abuse, those who presumed that there was sexual abuse were less likely to reach a correct conclusion (Abuse: 638 correct conclusions, 54.9%; No Abuse: 507 correct conclusions, 60.1%, $\chi^2(1) = 5.75$, $p = .016$), supporting H3.4. For cases where the interviewer did not have a clear preliminary assumption, the correct conclusion rate is 61.02% (36/59). However, no significant difference was detected between cases with a preliminary assumption of abuse and cases with no clear preliminary assumption, $\chi^2(1) = 0.46$, $p = .50$), possibly due to the limited number of the latter.

DISCUSSION

Mega-analysing data from a series of experiments, we examined the presence and effects of confirmation bias in CSA interviews. First, people tended to form an assumption of abuse when being presented with a suspected CSA case. Fifty-eight percent of interviewers who formed an assumption believed that the case was probably an abuse case when the base rate was exactly 50%. The participants may have assumed that the task is to help abused children talk about their experiences. This is an erroneous assumption for an investigative interview but one that may be held by both laypersons and professionals. Moreover, psychological research has also long discovered that negativity bias is a universal character of animals and humans (Ito et al., 1998; Vaish et al., 2008). People tend to give more weight to negative information and affect. Thus, when presented with a suspected CSA case, people could give more weight to those details that support the assumption of abuse, leading to a higher proportion of people forming such an assumption.

In addition, the results showed that the assumption formed based on case introduction predicted the conclusion after the interview (supporting H1), which is indicative of a tendency to confirm more than disconfirm previous thoughts (Tversky & Kahneman, 1974; Wright & Anderson, 1989). This result is consistent with previous research employing other types of techniques (Bruck et al., 1999, as cited in O'Donohue & Cirlugea, 2021; Powell et al., 2012; Quas et al., 2007). Despite repeated evidence suggesting that confirmation bias is present in forensic interviews, previous research has also suggested ways to successfully combat confirmation bias. For example, Powell et al. (2012) found that interviewers who adhered to open-question practices were not influenced by biasing information. Also, interviewers who are blind to the details of an allegation performed better in subsequent interviews, eliciting more accurate information from children (Rivard et al., 2016).

However, in departure from previous findings that interviewers who were exposed to biased incorrect information used more leading and suggestive questions and less open questions (O'Donohue & Cirlugea, 2021; Powell et al., 2012), we found that those who had a wrong preliminary assumption asked both fewer recommended and not recommended questions and elicited fewer neutral and wrong details (rejecting H2.1 and H2.2). One possible reason is that the scenarios used were explicit child sexual abuse scenarios while previous studies have mostly employed neutral mock events. Moreover, even though the simulated interview training gathered data from police officers and psychologists, the majority of interviews (1783 out of 2084) came from interviewers who had no previous experience of CSA cases. It is reasonable to assume that inexperienced interviewers would refrain from asking too many questions about an event of such sensitive nature when not receiving confirming evidence from the avatars.

Moreover, as predicted, we found that the assumption of abuse was associated with greater confirmation bias in terms of positive hypothesis testing and evidence interpretation. Participants that assumed an abuse happened were more confident in their judgements, used more not recommended questions, elicited more wrong details, reached less correct conclusions and used the relevant details they uncovered about the case less effectively (supporting H3.1, H3.2, H3.3, H3.4 and H4). These results are in line with empirical and theoretical development in emotion research (Briñol & DeMarree, 2012; Briñol et al., 2018) and research examining the interviewer's emotional response to suspected child sexual abuse cases (Segal & Santtila, 2022; Magnusson et al., 2021). Certain emotions frequently evoked in CSA interviews such as anger and disgust (Segal & Santtila, 2022; Magnusson et al., 2021) could induce feelings of confidence in one's thoughts (Briñol et al., 2018), which in turn may cause greater cognitive biases such as confirmation bias. These effects of assumption of abuse can have a direct impact on the quality of subsequent decision-making in those alleged CSA cases where there is no other corroborating evidence besides the interview with the child. There is a danger that a preliminary assumption of abuse can increase false-positive conclusions with deleterious effects for the child, the innocent suspects, and the family at large. Therefore, interviewers should use a neutral hypothesis-testing approach in alleged CSA cases. In the current analysis, we also observed that an assumption of abuse was associated with a lower rate of correct conclusions. We believe that this important result warrants more rigorous examination before generalization to real-world settings.

Even though current training programmes are effective in improving interview quality (e.g. question use, rapport-building and utilization of relevant information), the programmes are often interviewee-centred, that is, provide the interviewers with feedback on what types of questions they should or should not use while taking into considerations the behavioural pattern of children. By incorporating the newly discovered knowledge about interviewers' behaviours into these programmes, hopefully, interview quality can be further improved. Cognitive biases are ever-present, and it may well be impossible to prevent interviewers from having the first opinion about a case. However, knowledge of bias and the ability to make implicit biases explicit together with employing a balanced hypotheses testing approach can help interviewers in controlling potential biasing effects such as the tendency of conducting an interview in a confirmatory manner.

Limitations and future directions

Due to the nature of the data used in the current analysis, we did not have measures of specific emotions that in theory contributed to the elevated confirmation bias among those who presumed sexual abuse. However, the link between abuse and anger and disgust for the materials used in the studies included in the present analyses has been established in a recent study conducted by Segal and Santtila (2022) and the literature on emotion and metacognition is abundant (Briñol & DeMarree, 2012; Briñol et al., 2007, 2018). Still, to offer a more rigorous examination of confirmation bias in CSA interviews, future research should measure specific emotions after reading the case introduction and after the interview.

Secondly, the analysis on question use employed summary data collapsing several types of not recommended questions. Among these questions, not all of them are necessarily connected with positive hypothesis testing such as those categorized as *feeling* or *fantasy* (see Table 1). Therefore, it would be beneficial if future studies could examine the effect of the assumption of abuse on the use of specific not recommended questions. The interviews in the present study were part of a simulation. While this carries some limitations (e.g. the strength of emotions and the level of involvement could be lower compared to real cases), the simulated interviews allowed us to know the ground truth and calculate some diagnostic effects (e.g. the impact of relevant detail on the conclusion), which would not be calculable in real cases. The effect sizes were small in some analyses. However, this may be related to the magnitude of emotion experienced by the interviewer, which—as pointed out above—could be weaker compared to real cases.

Despite the above-mentioned limitations, it is important to stress that the simulative interviews are interactive with avatars behaving and responding like real children. This allows an arguably more ecologically valid examination of confirmation bias in forensic settings, compared with reading vignettes and case briefs while maintaining high levels of experimental control. Future works may also adopt such an approach to examine emotional and cognitive processes in legal proceedings.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTIONS

Shumpei Haginoya (Data curation; Writing – original draft; Writing – review & editing) **Pekka Santtila** (Supervision; Writing – original draft; Writing – review & editing) **Francesco Pompedda** (Data curation; Writing – original draft; Writing – review & editing) **Yikang Zhang**, M.Ed (Conceptualization; Formal analysis; Methodology; Writing – original draft; Writing – review & editing) **Aleksandr Segal** (Writing – original draft; Writing – review & editing).

DATA AVAILABILITY STATEMENT

The original data that support the findings of this study are available from the authors of the published articles used in current mega-analysis upon reasonable request. The R code of this study are openly available at open science framework (<https://doi.org/10.17605/OSF.IO/CZXEY>).

ORCID

Yikang Zhang  <https://orcid.org/0000-0001-5173-562X>

Aleksandr Segal  <https://orcid.org/0000-0002-6490-2498>

Francesco Pompedda  <https://orcid.org/0000-0001-9253-0049>

Shumpei Haginoya  <https://orcid.org/0000-0001-6909-674X>

Pekka Santtila  <https://orcid.org/0000-0002-0459-1309>

REFERENCES

- Segal, A., & Santtila, P. (2022). *Simulated interview avatars with child sexual abuse (vs. no abuse) scenarios elicit different emotional reactions: a pre-registered study* [manuscript submitted for publication]
- *Kask, K., Pompedda, F., Palu, A., Schiff, K., Mägi, M., & Santtila, P. (2022). *Avatar training effects transfer to investigative field interviews of children conducted by police officers* [manuscript submitted for publication].
- *Haginoya, S., & Santtila, P. (2022). *Improving supportiveness of interviewing the simulated child sexual abuse interviews* [manuscript submitted for publication].
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4*. arXiv preprint arXiv:1406.5823.
- Briñol, P., & DeMarree, K. (2012). *Social metacognition*. Psychology Press.
- Briñol, P., Petty, R. E., & Barden, J. (2007). Happiness versus sadness as a determinant of thought confidence in persuasion: A self-validation analysis. *Journal of Personality and Social Psychology*, 93(5), 711–727. <https://doi.org/10.1037/0022-3514.93.5.711>
- Briñol, P., Petty, R. E., Stavradi, M., Lamprinakos, G., Wagner, B., & Díaz, D. (2018). Affective and cognitive validation of thoughts: An appraisal perspective on anger, disgust, surprise, and awe. *Journal of Personality and Social Psychology*, 114(5), 693–718. <https://doi.org/10.1037/pspa0000118>
- Brown, D. A., Lamb, M. E., Lewis, C., Pipe, M.-E., Orbach, Y., & Wolfman, M. (2013). The NICHD Investigative Interview Protocol: An analogue study. *Journal of Experimental Psychology: Applied*, 19(4), 367–382.
- Carlson, K. A., & Russo, J. E. (2001). Biased interpretation of evidence by mock jurors. *Journal of Experimental Psychology: Applied*, 7(2), 91–103.
- Fischer, P., Jonas, E., Frey, D., & Schulz-Hardt, S. (2005). Selective exposure to information: The impact of information limits. *European Journal of Social Psychology*, 35(4), 469–492. <https://doi.org/10.1002/ejsp.264>
- Frey, D. (1986). Recent research on selective exposure to information. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, Vol. 19 (pp. 41–80). Academic Press.
- *Haginoya, S., Yamamoto, S., Pompedda, F., Naka, M., Antfolk, J., & Santtila, P. (2020). Online simulation training of child sexual abuse interviews with feedback improves interview quality in Japanese university students. *Frontiers in Psychology*, 11, 998. <https://doi.org/10.3389/fpsyg.2020.00998>

- *Haginoya, S., Yamamoto, S., & Santtila, P. (2021). The combination of feedback and modeling in online simulation training of child sexual abuse interviews improves interview quality in clinical psychologists. *Child Abuse & Neglect*, *115*, 105013. <https://doi.org/10.1016/j.chiabu.2021.105013>
- Hartnagel, T. F., & Templeton, L. J. (2012). Emotions about crime and attitudes to punishment. *Punishment & Society*, *14*(4), 452–474. <https://doi.org/10.1177/1462474512452519>
- Huntsinger, J. R., Clore, G. L., Brinol, P., & DeMarree, K. (2012). Emotion and social metacognition. In P. Brinol, & K. DeMarree (Eds.), *Social metacognition* (pp. 199–217). Psychology Press.
- Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, *75*(4), 887–900. <https://doi.org/10.1037/0022-3514.75.4.887>
- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, *80*(4), 557–571. <https://doi.org/10.1037/0022-3514.80.4.557>
- Jonas, E., Traut-Mattausch, E., Frey, D., & Greenberg, J. (2008). The path or the goal? Decision vs. information focus in biased information seeking after preliminary decisions. *Journal of Experimental Social Psychology*, *44*(4), 1180–1186.
- Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, *2*(1), 42–52. <https://doi.org/10.1016/j.jarmac.2013.01.001>
- Klayman, J. (1995). Varieties of confirmation bias. In J. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Psychology of learning and motivation* (Vol. 32, pp. 385–418). Academic Press.
- Koppelaar, L., Lange, A., & Van De Velde, J.-W. (1997). The influence of positive and negative victim credibility on the assessment of rape victims: An experimental study of expectancy—confirmation bias. *International Review of Victimology*, *5*(1), 61–85. <https://doi.org/10.1177/026975809700500104>
- *Krause, N., Pompèdda, F., Antfolk, J., Zappalà, A., & Santtila, P. (2017). The effects of feedback and reflection on the questioning style of untrained interviewers in simulated child sexual abuse interviews. *Applied Cognitive Psychology*, *31*(2), 187–198. <https://doi.org/10.1002/acp.3316>
- Kukucka, J., & Kassin, S. M. (2014). Do confessions taint perceptions of handwriting evidence? An empirical test of the forensic confirmation bias. *Law and Human Behavior*, *38*(3), 256–270. <https://doi.org/10.1037/lhb0000066>
- Magnusson, M., Joleby, M., Luke, T. J., Ask, K., & Lefsaer, S. M. (2021). Swedish and norwegian police interviewers' goals, tactics, and emotions when interviewing suspects of child sexual abuse. *Frontiers in Psychology*, *12*, 606774. <https://doi.org/10.3389/fpsyg.2021.606774>
- McCartan, K. F., Kemshall, H., & Tabachnick, J. (2015). The construction of community understandings of sexual violence: Rethinking public, practitioner and policy discourses. *Journal of Sexual Aggression*, *21*(1), 100–116. <https://doi.org/10.1080/13552600.2014.945976>
- Melinder, A., Brennen, T., Husby, M. F., & Vassend, O. (2020). Personality, confirmation bias, and forensic interviewing performance. *Applied Cognitive Psychology*, *34*(5), 961–971. <https://doi.org/10.1002/acp.3674>
- Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*(1), 85–95. <https://doi.org/10.1080/00335577743000053>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- O'Brien, B. (2009). Prime suspect: An examination of factors that aggravate and counteract confirmation bias in criminal investigations. *Psychology, Public Policy, and Law*, *15*(4), 315–334. <https://doi.org/10.1037/a0017881>
- O'Donohue, W., & Cirluega, O. (2021). Controlling for confirmation bias in child sexual abuse interviews. *The Journal of the American Academy of Psychiatry and the Law*, *49*(3), 371–380.
- Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child Abuse & Neglect*, *24*(6), 733–752. [https://doi.org/10.1016/S0145-2134\(00\)00137-X](https://doi.org/10.1016/S0145-2134(00)00137-X)
- Oxburgh, G., Williamson, T., & Ost, J. (2006). Police officers' use of emotional language during child sexual abuse investigations. *Journal of Investigative Psychology and Offender Profiling*, *3*(1), 35–45. <https://doi.org/10.1002/jip.41>
- Pompèdda, F. (2018). *Training in investigative interviews of children: Serious gaming paired with feedback improves interview quality* [Ph.D. thesis]. Åbo Akademi University, Turku. Available at <http://urn.fi/URN> [Google Scholar]
- *Pompèdda, F., Antfolk, J., Zappalà, A., & Santtila, P. (2017). A combination of outcome and process feedback enhances performance in simulations of child sexual abuse interviews using avatars. *Frontiers in Psychology*, *8*, 1474. <https://doi.org/10.3389/fpsyg.2017.01474>
- *Pompèdda, F., Palu, A., Kask, K., Schiff, K., Soveri, A., Antfolk, J., & Santtila, P. (2020). Transfer of simulated interview training effects into interviews with children exposed to a mock event. *Nordic Psychology*, *73*(1), 43–67. <https://doi.org/10.1080/19012276.2020.1788417>
- Pompèdda, F., Zappalà, A., & Santtila, P. (2015). Simulations of child sexual abuse interviews using avatars paired with feedback improves interview quality. *Psychology, Crime & Law*, *21*(1), 28–52. <https://doi.org/10.1080/1068316X.2014.915323>
- Powell, M. B., Hughes-Scholes, C. H., & Sharman, S. J. (2012). Skill in interviewing reduces confirmation bias. *Journal of Investigative Psychology and Offender Profiling*, *9*(2), 126–134. <https://doi.org/10.1002/jip.1357>

- Quas, J. A., Malloy, L. C., Melinder, A., Goodman, G. S., D'Mello, M., & Schaaf, J. (2007). Developmental differences in the effects of repeated interviews and interviewer bias on young children's event memory and false reports. *Developmental Psychology, 43*(4), 823–837. <https://doi.org/10.1037/0012-1649.43.4.823>
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org>
- Rassin, E., Eerland, A., & Kuijpers, I. (2010). Let's find the evidence: An analogue study of confirmation bias in criminal investigations. *Journal of Investigative Psychology and Offender Profiling, 7*(3), 231–246. <https://doi.org/10.1002/jip.126>
- Rivard, J. R., Pena, M. M., & Schreiber Compo, N. (2016). “Blind” interviewing: Is ignorance bliss? *Memory, 24*(9), 1256–1266.
- Santtila, P., Zappalà, A., & Pompèdda, F. (2012). *EIT® (Version 1.0)*. [Computer software]. Åbo Akademi University.
- Talluri, B. C., Urai, A. E., Tsetsos, K., Usher, M., & Donner, T. H. (2018). Confirmation bias through selective overweighting of choice-consistent evidence. *Current Biology, 28*(19), 3128–3135. <https://doi.org/10.1016/j.cub.2018.07.052>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.
- Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*(3), 383–403. <https://doi.org/10.1037/0033-2909.134.3.383>
- Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*(2), 249–265. <https://doi.org/10.1037/a0031311>
- Wright, W. F., & Anderson, U. (1989). Effects of situation familiarity and financial incentives on use of the anchoring and adjustment heuristic for probability assessment. *Organizational Behavior and Human Decision Processes, 44*(1), 68–82. [https://doi.org/10.1016/0749-5978\(89\)90035-6](https://doi.org/10.1016/0749-5978(89)90035-6)

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Zhang, Y., Segal, A., Pompèdda, F., Haginoya, S., & Santtila, P. (2022). Confirmation bias in simulated CSA interviews: How abuse assumption influences interviewing and decision-making processes? *Legal and Criminological Psychology, 00*, 1–15. <https://doi.org/10.1111/lcrp.12213>