



## OPEN Collecting initial accounts using ChatCharlie chatbot improves eyewitness memory in later investigative interviews

Coral J. Dando<sup>1</sup>✉ & Charlotte E. Adam<sup>2</sup>

Initial account interviews (IAi) offer eyewitnesses more immediate opportunities to answer a series of brief questions about their experiences prior to an in-depth, more formal investigative interview. An IAi is typically elicited in-person near/at the scene of a crime using broadly systematic questioning. Retrieval practice can improve subsequent recall in some contexts, but there is a dearth of research centred on the potential costs and benefits of a quick IAi. Furthermore, where an in-person IAi is impossible, no alternative quick provision exists. Given the systematic nature of the IAi protocol, we developed a conversational chatbot as a potential alternative. Using a mock-witness paradigm, we investigated the memory performance of adults from the general population during in-depth in-person interviews one week after having provided an IAi 10 min post event either (1) in person, (2) via the ChatCharlie chatbot, or (3) no IAi (control). IAi conditions leveraged significantly improved event recall during later investigative interviews versus the Control. Accounts were more accurate and complete, and more correct information was remembered without increased errors indicating the potential of digital agents for IAi purposes. Findings concur with predictions from theoretical understanding of episodic memory consolidation and the empirical eyewitness literature regarding the benefits of practice in some contexts.

**Keywords** Eyewitness memory, Initial accounts, Chatbots, Investigative interview, Delay, Memory consolidation

Information provided by witnesses and victims of crime (henceforth witnesses) is fundamental for guiding the investigation of crime from the offset, and remains crucial during the prosecution process, and beyond<sup>1,2</sup>. However, witness memories are not objective records of events, but representative reconstructions of personal experiences which can be likened to physical trace evidence such as DNA and fingerprints in terms of fragility<sup>3–5</sup>. Accordingly, witness memory should be treated with care since episodic long-term memory is hard to access, malleable and error-prone<sup>6,7</sup>.

How ‘best’ to manage the elicitation of information from witnesses to maximise the quality and quantity of information recalled is an enduring challenge for front-line investigators<sup>8–11</sup>. In the UK and elsewhere, witnesses typically provide information about their experiences during an investigative interview conducted by a police officer or other investigative professional. While practices may vary according to jurisdiction, the term investigative interview typically refers to a formal in-depth interview that takes place at a police station, a witness’ home or other appropriate interview facilities<sup>12</sup>. Here, interviewers seek to elicit a detailed and accurate account using techniques that draw on cognitive and social psychological theory and empirical evidence to support witnesses to recall and verbalise their experiences<sup>13,14</sup>.

Investigative interviews necessarily occur sometime after an event so delays between encoding and retrieval are common, and have long been highlighted<sup>13,15–17</sup>. Delays are variable according to numerous factors including investigator availability, access to appropriate interview facilities, operational challenges, event and individual witness considerations and can range from a few hours, to months and even years later in some instances<sup>12,15,17,18</sup>.

The extant theoretical and applied memory literature is consistent in indicating the value of timely retrieval practice (where appropriate), since episodic information is less stable and so vulnerable to interference during the period after encoding. Encoding occurs rapidly, followed by the more gradual consolidation process of transforming fragile newly encoded episodic memories into long-term more stable forms<sup>19–23</sup>. Practice can

<sup>1</sup>Department of Psychology, University of Westminster, 115 New Cavendish Street, London W1W 6UW, UK.

<sup>2</sup>Division of Psychology, University College London, London, UK. ✉email: c.dando@westminster.ac.uk

support consolidation, helping to stabilize memory, making it more robust than might otherwise be the case, typically enhancing recall memory for practised items by supporting goal-directed episodic memory<sup>20,21,23</sup>. A timely initial retrieval also has potential to confer investigative benefits by leveraging important investigation relevant information at an earlier stage than might otherwise be the case.

In recognition of the potential benefits of a quick first retrieval, the College of Policing (CoP), the professional policing public body in England and Wales, has developed and implemented an initial account interview protocol. The Initial Account interview (IAi) is a ‘quick’ initial interview that takes place at or near the scene of the event in question, towards improving the accuracy and quantity of information provided by witnesses during a later investigative interview and supporting police to take timely and appropriate action quickly after an event (College of Policing, 2024).

In some instances, witnesses may be able to provide an initial statement about their experiences by completing a Self-Administered Interview (SAI). The SAI (available in paper-based and more recently digital formats)<sup>15,24–26</sup> aims to elicit detailed accounts and so differs considerably to the IAi protocol in terms of rationale, implementation, complexity, retrieval techniques and questioning approach. The IAi elicits a very brief account of an incident to inform police of what happened and help them take immediate action at the scene. An IAi is short, context-dependent, and focuses on securing evidence, identifying offenders, safeguarding, assessing risk, and identifying offences. The SAI on the other hand gathers a more comprehensive initial statement, and so seeks to leverage detailed episodic retrieval, and thus is not countenanced for immediate action purposes. Although not the focus of the research reported here, nonetheless, the SAI literature offers significant evidence of the applied benefits of timely first retrieval.

Despite the extant theoretical and experimental research in support of the potential cognitive benefits of an IAi, as advocated by the CoP, there is a dearth of applied eyewitness memory research in this regard. The CoP approach for IAi has yet to be empirically evaluated with adults from the general population, and so the cognitive benefits are not well understood. Further, since fast initial accounts are provided in response to a series of questions posed by a police officer or other investigative professional, when this is not possible there appears little alternative but to forgo an IAi. In such instances the cost in terms of memory performance in later investigative interviews is unclear. A lack of appropriate alternatives in the face of finite human resources, or in contexts where gathering an initial account is inappropriate due to large numbers of witnesses for example, might prove costly for criminal justice processes.

The research reported here moves towards filling gaps in understanding by investigating the efficacy of a CoP style IAi on later recall during an in-depth investigative interview versus providing no initial account. Since the current CoP guidance presents a largely prescriptive approach to asking a series of straightforward initial questions, this type of interaction lends itself to some element of atomisation with potential for offering alternative provision<sup>16,27–29</sup>. Here we empirically evaluate a potential addition to the initial account ‘toolbox’ in the form of a remote contextual chatbot which we refer to as ChatCharlie. ChatCharlie has been designed to engage witnesses in a conversation style interaction in a manner that broadly maps onto the CoP initial account approach, and so we also consider the impact of a chatbot based IAi on recall one week later (see “Methods”).

ChatCharlie has potential to offer witnesses 24/7 access to a synchronous initial account digital agent to guide them in providing a timely initial account in line with current best practice for supporting episodic recall, and in a manner that is well aligned with the operational benefits of an initial account suggested by the CoP. The use of various types of chatbots (e.g., hybrid, AI, conversational) is rapidly expanding across numerous sectors, including education, financial services, and public health<sup>30–32</sup>, for example. The benefits of human–chatbot interactions include timely access to information and targeted assistance, improved self-disclosure in part resulting from the anonymous non-judgmental nature of human–chatbot interactions, and in some cases, chatbots have been found effective for providing social support.

Human–chatbot interactions are not without challenges. Poor user experiences are widely reported<sup>33,34</sup>, particularly when chatbots do not provide relevant answers to questions and so do not support users in resolving their queries or problems<sup>35,36</sup>, and when access is limited and interface design is poor<sup>37,38</sup>. User needs can be complex and so human–chatbot dialogue that is helpful and provides relevant and timely responses are key drivers to improving user experiences. For the purposes of the research reported here, in the absence of a human questioner, a suitably designed chatbot may offer a ‘next best’ option for triggering the process of extracting knowledge from personal experiences via IAi. One corollary being memory consolidation processes are supported and strengthened by ChatCharlie, resulting in quantifiable benefits and improved recall later versus the absence of an IAi practice when a human agent is unavailable.

ChatCharlie is not a customer support agent, but rather a conversational agent designed to collect and securely store textual responses to a series of contextually relevant questions. ChatCharlie has been designed to reduce perceptions of unresponsiveness by timing messages optimally<sup>39–41</sup>, ensuring witnesses are not overwhelmed by information and recall requests, but that response lag does not appear so unhuman like as to feel uncomfortable and/or frustrating. ChatCharlie offers users a human like name, which can increase perceptions of an authentic interaction<sup>42</sup> especially when paired with an informal communication style). Reassurances regarding confidentiality, and if appropriate anonymity, can also increase engagement and acceptability and as such ChatCharlie immediately and then intermittently offers some information in this regard. Finally, interface feedback was considered towards maximising the likelihood of an acceptable user experience. Alpha testing of ChatCharlie included testing buttons, icons and input fields, considering the logical flow and understandability of the questions and information, and limit testing to establish ChatCharlie responses in situations of unforeseen response scenarios. ChatCharlie was refined based on Alpha testing responses and adjusted accordingly prior to this research, which represents Beta style testing.

The novelty of this research and a dearth of literature relevant to quick IAi style interview with adults and the use of chatbots in eyewitness memory contexts supported the formulation of the following hypotheses. As

predicted by contemporary understanding of long-term memory, our first hypothesis  $H^1$  was that irrespective of method of initial account retrieval method (In-person; ChatCharlie), providing an IAi promptly would result in improved recall during a more in-depth In-person investigative interview one week later. Our second null hypothesis  $H^0$  was that there would be non-significant memory performance differences between ChatCharlie and In-person initial account retrievals, since ChatCharlie mimics the prevailing IAi protocol. Rapport building is countenanced by the CoP protocol and is present in the In-person initial account condition, but necessarily absent when initiating an initial account via ChatCharlie. Reductions in social demand inherent in interactions when humans are absent may counterbalance the absence of rapport in ChatCharlie. Further, attempts to build rapport would not be expected, and so likely viewed as ‘false’ and inappropriate when communicating with a digital agent, which can divert cognitive resources away from invoking episodic retrieval mode<sup>43–45</sup>.

## Results

To investigate  $H^1$  we conducted a series of multivariate and univariate analyses of Time 2 interview recall performance metrics (e.g., correct, errors, confabulations, accuracy and completeness), globally (that is, from start of the Time 2 investigative interview to the end) and as a function of interview phase (free recall; cued recall). We analysed self-reported confidence in the correctness of recall and errors at Time 1 as a function of IAi condition and then again at Time 2 across all conditions to understand self-reported confidence in memory performance. To test the null hypothesis ( $H^0$ ) we conducted Bayesian analyses across each of the Time 1 IAi performance metrics (correct, errors, confabulations) to understand the relative likelihood of a difference between ChatCharlie and In-person IAi memory performance at Time 1<sup>46</sup>. Additional exploratory analyses investigated self-report confidence in global correct and erroneous recall using t-tests and ANOVAs.

### Time 1 initial account

Bayesian tests (with default priors) revealed the amount of correct information recalled at Time 1 across the In-person and ChatCharlie IAi conditions did not significantly differ,  $p=0.399$ ,  $d=0.22$ ,  $BF_{01}=3.71$  (95% CI – 3.46, 8.39). Likewise, the number of errors and confabulation did not significantly differ,  $p=0.237$ ,  $d=0.30$ ,  $BF_{01}=2.70$  (95% CI – 0.56, 2.16) and  $p=0.097$ ,  $d=0.43$ ,  $BF_{01}=1.44$  (95% CI – 0.09, 0.96), respectively. See Table 1 for means, SDs and 95% CI.

### Global recall performance at time 2

Multivariate Analysis of Variance for global composite correct, erroneous and confabulated recalled during Time 2 investigative interviews (from start to finish) was significant,  $F(6, 172)=9.680$ ,  $p<0.001$ ,  $\eta_p^2=0.25$  (Pillai’s Trace). Follow-up discriminant analysis revealed two significant discriminant functions. The first explained 75.4% of the variance, conical  $R^2=0.35$ , the second explained 24.6%, conical  $R^2=0.15$ . In combination, both discriminant functions significantly differentiated the Time 2 retrieval groups,  $\Lambda=0.55$ ,  $\chi^2(6)=51.623$ ,  $p\leq 0.001$  and  $\Lambda=0.85$ ,  $\chi^2(2)=14.131$ ,  $p\leq 0.001$ . Correlations between outcomes and the discriminant functions revealed that correct recall loaded highly onto both functions ( $r=0.64$  and  $r.72$ , respectively). However, confabulations loaded more highly into the first function ( $r=0.46$ ) which explained 75.4% of the variance than in function 2 ( $r=-0.60$ ), whereas errors were loaded more highly into the second function ( $r=0.15$ ) than in the first function ( $r=-0.25$ ), which explained just 24.6% of the variance. Correct recall combined with confabulations significantly discriminated the Control from ChatCharlie and In-person Time 1 conditions. Correct recall combined with errors significantly discriminated ChatCharlie from both the Control and In-person Time 1 conditions.

Univariate  $F$ s for each of the Time 2 dependent variables (followed by post hoc Games-Howell tests (applying Bonferroni’s correction as appropriate) revealed a significant main effect of condition for correct recall at Time 2,  $F(2, 87)=13.935$ ,  $p<0.001$ ,  $\eta_p^2=0.24$ . Participants in the ChatCharlie and In-person Time 1 conditions recalled more correct information at Time 2 than those in the Control condition, all  $p$ s  $< 0.001$ . There was a non-significant difference between the former two conditions,  $p=0.256$  (see Table 2).

#### Erroneous recall

There was a non-significant main effect of condition for the amount of erroneous recall at Time 2,  $F(2, 87)=3.791$ ,  $p=0.026$ ,  $\eta_p^2=0.08$  (see Table 2).

#### Confabulations

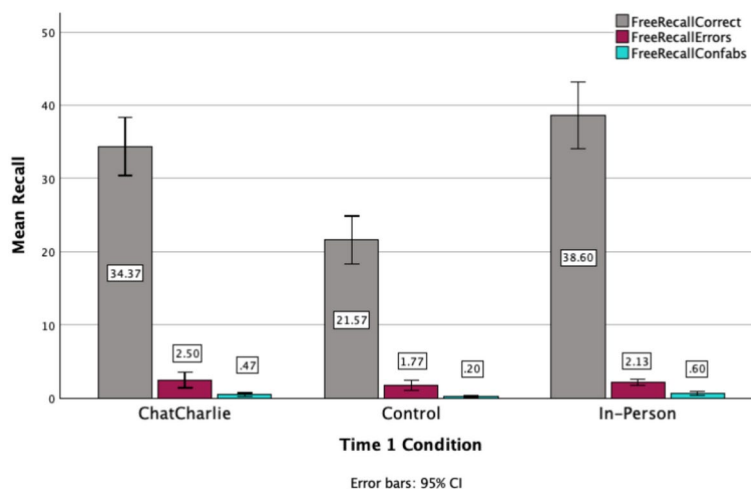
There was a significant main effect of condition for the number of confabulations at Time 2,  $F(2, 87)=6.019$ ,  $p=0.004$ ,  $\eta_p^2=0.12$ . Participants in both the In-person and ChatCharlie Time 1 conditions reported fewer confabulations at Time 2 than those in the Control,  $p=0.002$  and  $p=0.007$ , respectively. There was a non-significant difference between the former two conditions,  $p=0.635$  (see Table 2).

	Mean (SD) 95% CI		
	Correct	Error	Confabulation
In-Person	39.50 (13.08) 34.62, 44.38	4.73 (2.84) 3.67, 5.79	1.00 (1.23) 0.54, 1.46
ChatCharlie	37.03 (9.05) 33.65, 40.41	3.93 (2.32) 3.07, 4.80	0.68 (0.76) 0.31, 0.82

**Table 1.** Means, SDs and 95% CI In-person and ChatCharlie Time 1 initial account memory performance (correct, errors and confabulations).

	Means (SD) 95% CI			
	Correct	Errors	Confabulations	% Accuracy
ChatCharlie	56.67 (16.05) 51.50, 61.83	6.57 (4.01) 5.26, 7.71	0.87 (0.86) 0.47, 1.26	88.94 (4.61) 86.90, 90.97
In-Person	52.47 (14.40) 47.30, 57.63	4.40 (1.69) 3.26, 5.54	1.63 (1.16) 1.24, 2.03	89.87 (4.71) 87.84, 91.90
Control	38.17 (12.00) 33.00, 43.33	5.97 (3.28) 4.83, 7.11	0.73 (1.20) 0.34, 1.13	85.24 (7.11) 83.21, 87.27

**Table 2.** Time 2 mean global memory performance measures and percentage accuracy as a function of Time 1 initial account condition ( $N=90$ ).



**Fig. 1.** Mean recall for correct, errors and confabulations in the free recall phase of Time 2 interviews as a function of Time 1 retrieval condition ( $N=90$ ).

#### Percentage accuracy

Analysis of variance revealed a significant main effect for percentage accuracy in Time 2 interviews,  $F(2, 87) = 180.126$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.12$  (see Table 1 for means SDs and 95% CIs). Participants in the Control condition were less accurate at Time 2 than those in both the In-person and ChatCharlie Time 1 conditions,  $p = 0.006$  and  $p = 0.037$ , respectively, with a non-significant difference between the ChatCharlie and In-Person IAi conditions,  $p = 0.718$  (see Table 2).

#### Completeness

The template approach used for coding Time 2 interviews comprised 111 items of event information. Completeness was the amount of correct recalled as a percentage of 111. There was a significant difference across conditions for completeness at Time 2,  $F(2, 87) = 13.925$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.24$ . Participants in the Control condition were significantly less complete ( $M_{\text{Control \% complete}} = 34.38$ ,  $SD = 18.81$ ,  $95\% \text{ CI } 30.35, 38.42$ ) than those in the ChatCharlie and In-person conditions, all  $ps < 0.001$  ( $M_{\text{ChatCharlie \% complete}} = 51.05$ ,  $SD = 14.42$ ,  $95\% \text{ CI } 45.67, 56.44$ ;  $M_{\text{In-person \% complete}} = 47.27$ ,  $SD = 12.97$ ,  $95\% \text{ CI } 42.42, 52.11$ ), with no significant difference between the latter two conditions,  $p = 0.537$ .

#### Time 2 interview phase memory performance

##### Free recall

There was a significant main effect for correct recall in the free recall phase of Time 2 interviews,  $F(2, 87) = 20.786$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.32$  (see Fig. 1). Participants in the Control recalled significantly fewer correct information items than those in both ChatCharlie and In-person Time 1 conditions, all  $ps < 0.001$ , with a non-significant difference between the latter two conditions,  $p = 0.718$ . There were non-significant differences for errors and confabulations,  $F(2, 87) = 0.932$ ,  $p = 0.397$ ,  $\eta_p^2 = 0.02$ , and  $F(2, 87) = 3.053$ ,  $p = 0.052$ ,  $\eta_p^2 = 0.07$ , respectively.

There was a significant difference across conditions for percentage accuracy,  $F(2, 87) = 7.572$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.15$ . Participants in the ChatCharlie condition were more accurate ( $M = 92.66$ ,  $SD = 5.74$ ,  $95\% \text{ CI } 89.99, 95.34$ ) than participants in and Control ( $M = 92.09$ ,  $SD = 7.15$ ,  $95\% \text{ CI } 89.41, 94.77$ ),  $p = 0.003$  and In-person ( $M = 85.97$ ,  $SD = 8.90$ ,  $95\% \text{ CI } 83.30, 88.65$ ) conditions,  $p = 0.013$ , with non-significant differences between the former two conditions,  $p = 0.937$ .

### Cued recall

There was a significant main effect for the number of correct information items recalled in the cued recall phase at Time 2,  $F(2, 87) = 12.215$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.22$ . Participants in ChatCharlie recalled significantly more correct information than participants in both the Control and In-person conditions,  $p = 0.009$  and  $p < 0.001$ , respectively, with no difference between the latter two conditions,  $p = 0.204$  (see Fig. 2). There was a significant difference for errors,  $F(2, 87) = 7.641$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.22$ . Participants in the In-person condition recalled significantly fewer errors than Control and ChatCharlie,  $p = 0.003$  and  $p < 0.001$ , respectively, with a non-significant difference between the latter two conditions,  $p = 0.993$ .

There was a significant difference across conditions for confabulations in the cued recall phase of Time 2 interviews,  $F(2, 87) = 5.635$ ,  $p = 0.005$ ,  $\eta_p^2 = 0.11$ . Participants in the In-person Time 1 condition confabulated more during the cued recall phase at Time 2 than participants in ChatCharlie condition,  $p = 0.004$ . All other differences were non-significant, all  $ps > 0.061$  (see Fig. 2).

There was a significant difference in percentage accuracy in the cued recall as a function of Time 1 condition,  $F(2, 87) = 6.581$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.13$ . Participants in ChatCharlie ( $M = 84.12$ ,  $SD = 6.96$ , 95% CI 79.08, 89.15) were significantly more accurate than those in the In-person Time 1 condition ( $M = 71.14$ ,  $SD = 19.79$ , 95% CI 66.12, 76.18),  $p = 0.005$ . There was a non-significant difference between the control ( $M = 78.27$ ,  $SD = 11.70$ , 95% CI 73.24, 83.30) and ChatCharlie,  $p = 0.59$ .

### Memory confidence

Self-reported confidence was gathered from participants in both IAi conditions immediately post completion of an IAi (at Time 1) and from all participants immediately after the Time 2 interviews. It is not uncommon to collect self-report global confidence in this manner so as not to interfere with, and potentially truncate, the retrieval process (e.g., Odinet & Wolters, 2006; Dando et al., 2023). Interrupting retrieval to variously collect confidence data following responses to each question in turn may be of import for theoretical and confidence-accuracy calibration purposes, but here we were more concerned with mimicking the real world and thus employing witness-focused retrieval techniques when collecting confidence measures.

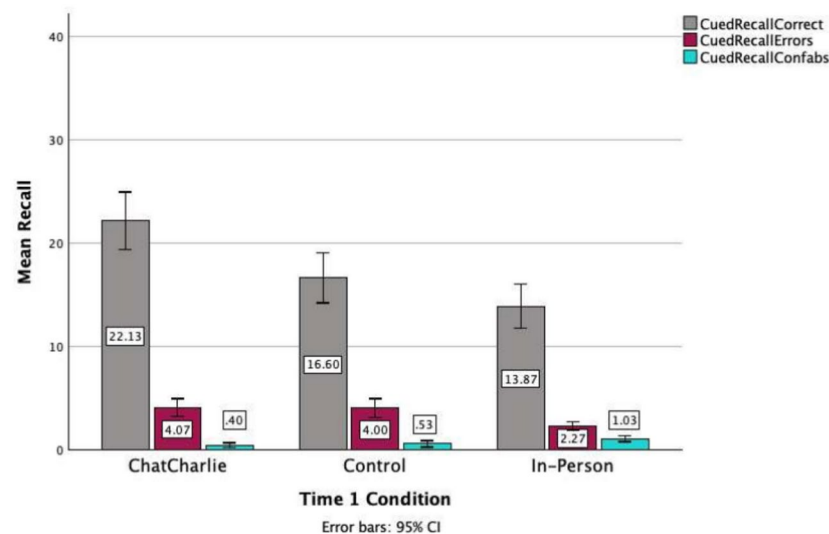
#### Confidence Time 1

There was a significant difference between ChatCharlie and In-person initial accounts conditions for self-reported confidence (ranging from 1 = not at all confident to 5 = completely confident) in the amount of correct event information recalled at Time 1,  $t(58) = 8.056$ ,  $p < 0.001$ ,  $d = 1.04$ . Participants in the In-person condition were more confident that they had remembered a lot of correct information ( $M_{\text{In-person}} = 3.47$ ,  $SD = 0.860$ ) than ChatCharlie participants ( $M_{\text{ChatCharlie}} = 2.80$ ,  $SD = 1.19$ ). There was a non-significant difference between ChatCharlie and In-person conditions for self-reported confidence in the number of errors recalled at Time 1,  $t(58) = 4.409$ ,  $p = 0.040$  ( $M_{\text{In-person}} = 3.67$ ,  $SD = 0.880$ ;  $M_{\text{ChatCharlie}} = 2.80$ ,  $SD = 1.37$ ).

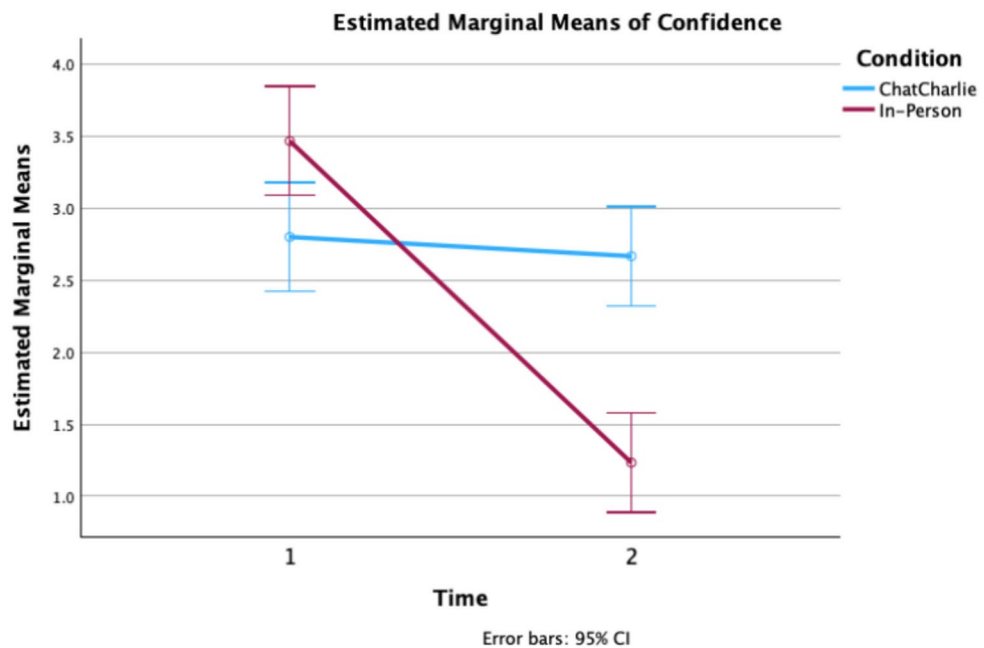
#### Confidence Time 2

There was a significant main effect for self-reported confidence in the amount of correct event information recalled at Time 2,  $F(2, 87) = 18.286$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.29$ . Participants in the In-person IAi condition were significantly less confident at Time 2 than those in both the ChatCharlie and Control conditions, all  $ps < 0.001$ , with a non-significant difference between the latter two conditions,  $p = 0.945$ .

There was also a significant main effect for how confident participants were that they had not made any errors at Time 2,  $F(2, 87) = 11.694$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.21$ . Participants in the In-person IAi condition were more confident



**Fig. 2.** Mean recall for correct, errors and confabulations in the cued recall phase of Time 2 interviews as a function of Time 1 retrieval condition ( $N = 90$ ).



**Fig. 3.** Confidence in correct recall across retrieval attempts (from Time 1 to Time 2) as a function of ChatCharlie and In-Person Time 1 conditions ( $N=60$ ).

that they had not made any errors at Time 2 than those in both the ChatCharlie and Control conditions, all  $p=0.003$  and  $p<0.001$ , respectively. There was a non-significant difference between the latter two conditions,  $p=0.458$ .

#### Confidence Time 1 to Time 2

Repeated measures analysis revealed a main effect of Time for self-reported confidence in correct information,  $F(1, 58)=66.906$ ,  $p\leq 0.001$ ,  $\eta_p^2=0.54$ , and a significant Time X Condition mixed interaction,  $F(1, 58)=52.678$ ,  $p\leq 0.001$ ,  $\eta_p^2=0.48$ . Irrespective of Time 1 condition (ChatCharlie; In-person), confidence in the amount of correct information recalled dropped from Time 1 ( $M=3.13$ ,  $SD=1.08$ , 95% CI, 2.866, 3.033) to Time 2 ( $M=1.95$ ,  $SD=0.86$ , 95% CI 2.051, 2.649) at Time 2. Confidence of participants in the In-person Time 1 condition dropped significantly from Time 1 to Time 2,  $p<0.001$ , whereas confidence of participants in ChatCharlie remained constant from Time 1 to Time 2,  $p=0.517$  (see Fig. 3).

There was a non-significant repeated measure main effect for confidence in errors from Time 1 ( $M=3.08$ ,  $SD=1.17$ ) to Time 2 ( $M=2.92$ ,  $SD=1.02$ ) and a non-significant Time X Condition mixed interaction,  $F_s<1.495$ , all  $p_s>0.226$  ( $M_{\text{ChatCharlie T1}}=2.50$ ,  $SD=1.14$ ;  $M_{\text{ChatCharlie T2}}=2.55$ ,  $SD=1.00$ ;  $M_{\text{In-person T1}}=3.67$ ,  $SD=0.88$ ;  $M_{\text{In-person T2}}=3.33$ ,  $SD=0.84$ ).

#### Confidence and accuracy

Regression analysis indicated that Time 1 and Time 2 self-reported confidence for correct and erroneous information were non-significant predictors of accuracy at Time 2,  $F(2, 59)=0.741$ ,  $p=0.481$ , and,  $F(2, 59)=1.133$ ,  $p=0.329$ , respectively accounting for only 2.5% and 7.6% in the variation of percentage accuracy at Time 2, respectively.

## Discussion

IAi interviews offer witnesses an immediate opportunity to briefly explain their experiences. They are not appropriate in all instances nor for all witnesses, and IAis do not replace in-depth investigative interviews nor SAIs but rather precede both if appropriate. Nonetheless, IAis offer the opportunity for witnesses to practice in a managed fashion, potentially stabilizing eyewitness memory, making it more robust than might otherwise be the case. Irrespective of Time 1 IAi retrieval mode (In-person or ChatCharlie), global memory performance during investigative interviews at Time 2 (one week later) was significantly improved.

Our results offer clear support for  $H^1$ , thus providing additional evidence of the benefits for eyewitnesses of providing an initial account promptly<sup>15,24</sup>. Participants who 'practised' via an IAi at Time 1 recalled 37% (In-person) and 47% (ChatCharlie) more correct information at Time 2, versus those who did not provide an IAi. Confabulations were reduced and significant improvements in episodic recall were not accompanied by an increase in errors, thus accuracy was higher indicating positive verbal and textual initial retrieval practice effects. Others have reported similar benefits following a SAI<sup>15,24,47</sup> which was designed to support comprehensive retrieval attempts akin to an in-person investigative interview (if appropriate) but without the need for an interviewer. The IAi is not a replacement for an SAI, nor vice versa, since they have been designed and developed for different purposes and the SAI does not capture a 'quick' account in line with current CoP

guidance. Nonetheless, in instances where the SAI has also been used, subsequent recall has been found to be more accurate and detailed.

ChatCharlie is a conversational question–answer digital agent to guide witnesses in providing a quick initial account (if appropriate) when human agents are unavailable. ChatCharlie is akin to the verbal in-person IAI currently advocated in terms of investigative approach, although rapport building is necessarily absent<sup>48,49</sup>. Yet, ChatCharlie appears equally as effective since Time 2 global correct recall performance was improved versus the Time 1 control condition (no IAI) with the Time 1 centred null hypothesis being more likely than the alternative hypothesis.

The pattern of results as a function of the individual retrieval phases typically present in basic investigative interviewing practices in the UK and elsewhere is more nuanced and worthy of consideration. Global interview findings are mirrored in the initial free recall phase, but performance in the second cued questioning retrieval reveals ChatCharlie participants reported almost 60% more correct information at Time 2 and confabulated significantly less than In-person IAI participants. However, ChatCharlie participants verbalised more errors (partially correct information) in this phase. The number of errors were few and the mean difference is small, and because ChatCharlie participants reported fewer confabulations they are more accurate in this phase. Nonetheless, this difference in errors was significant.

One potential, albeit partial explanation for the pattern of results in the second recall phase of Time 2 interviews emerges from in-person interviewing contexts and may have been triggered by the move from person absent interview contexts to In-person interviewing. Real and/or perceived demand<sup>50</sup> has the potential to inflate confidence and can encourage responses to questions resulting in the reporting of erroneous information in face-to-face interviews<sup>51–54</sup>. The absence of a human interviewer can support more cautious remembering, thus reducing errors and so our findings may be social in nature, rather than cognitive<sup>16,55</sup>. However, more research focusing on the potential impact of moving from an initial account interview with a digital agent to an in-person investigative interview is necessary to fully understand the locus of this effect.

Our null hypothesis was also supported with non-significant IAI memory performance differences at Time 1 across the ChatCharlie and In-person conditions. We considered the potential costs and benefits of no rapport when engaging with a digital agent<sup>16,45,55</sup>. Demand characteristics can be amplified in ‘performance’ contexts as is the case in applied witness memory research and so here, Time 1 performance across both Time 1 conditions suggests the trade-off between the presence/absence of a human agent and presence/absence of rapport appears acceptable in terms of memory performance. Further research regarding the investigative relevance of recalled information at Time 1 would improve understanding, as would more nuanced manipulation of rapport across conditions. However, given the prescriptive ‘getting down to business’ nature of an initial account interview when a human agent is unavailable our findings indicate ChatCharlie may be appropriate in some instances.

Self-report confidence did not predict memory accuracy, so despite experimental control in mock witness laboratory research resulting in ‘pristine’ conditions, predictive confidence judgments shortly after the event at Time 1 and again at Time 2 were not ‘useful’ performance indicators. We concur with others in that, at best, in applied contexts recall confidence should be treated with caution, particularly following in-person interviews where confidence may be inflated for reasons previously discussed centred on social cognition and perceived performance demand. A well-managed in-person interview is by its very nature witness-focused and aims to support witnesses to feel socially comfortable. Feeling socially comfortable may boost confidence in one’s performance<sup>56,57</sup>. Although speculative and so further research is necessary, here the In-person initial account protocol included rapport building and did leverage significantly higher levels of confidence at Time 1, which dropped significantly from Time 1 to Time 2. ChatCharlie confidence levels were lower from the offset but remained stable across time points.

Best practice and investigative training emphasise the need for caution centred on appropriate acceptance of responses to questions and considered verbal encouragement of effortful retrieval. However, witnesses can misinterpret appropriate encouragement, and over encouragement can occur, albeit this may be subtle in nature or unintended since interviewing is cognitively demanding for interviewers and so behavioural leakage can occur. Confidence has been found to decline in conditions of real stimuli and repeat confidence checking<sup>58</sup>. Although we did not repeat check confidence within interviews, for participants in the In-person and ChatCharlie Time 1 conditions, we did collect confidence data at Time 1 and again at Time 2. However, it is unclear why a repeat check might trigger a drop in confidence across the two time points for In-person Time 1 participants but not for ChatCharlie Time 1 participants, and why in the latter condition confidence remained constant. Again, there is significant scope for further applied confidence, accuracy and digital versus human agent research.

Our research is novel in terms of investigating the current guidance on gaining an initial account in England and Wales. However, findings are limited on serval counts. Our sample size was sufficient to achieve 80% power for a medium effect for  $H^1$  and is in line with sample size norms described in recent similar research but may lack power to robustly reject the null hypothesis. Further, we have reported the benefits of retrieval practice, but practice can sometimes negatively interfere with later recall, particularly recognition performance. For example, providing a verbal description of a perpetrator prior to identifying the perpetrator from a line-up can compromise recognition accuracy in some instances. Referred to as verbal overshadowing<sup>59</sup>, this retrieval-based interference appears robust for forensic recognition tasks, where describing a visual experience can interfere with subsequent visual recognition<sup>60,61</sup>, although not always<sup>62,63</sup>. However, the IAI paradigm employed here was followed by an in-depth investigative interview retrieval that did not include a recognition task. Nonetheless, future research may wish to investigate IAI in terms of verbal overshadowing.

Retrieval-induced forgetting (RIF) offers an example of when retrieval practice can trigger the forgetting of related information at a later stage that was not previously recalled from the same information category<sup>64</sup>. The applied eyewitness RIF literature is very mixed whereby repeated questioning has been found to strengthen recall of certain event details while inducing the forgetting of other details in some contexts, but not always and

is often context dependent<sup>65,66</sup>. Furthermore, different event details seem variously vulnerable to RIF in some instances<sup>67,68</sup>, and retrieval effort in the form of multiple recall attempts can predict memory accuracy over time sometimes resulting in improved recall with higher confidence in correct recall performance<sup>69</sup>. We did not investigate the possibility that an IAI might trigger RIF in adults, hence future research may also wish to consider the impact of an IAI on this phenomenon.

This research presents one study from a wider ongoing research programme concerning IAI and computer-mediated communication in investigative contexts more generally, and so the sample size and novelty limits generalisation. ChatCharlie is a prototype which will evolve, and maybe an AI version of ChatCharlie could incorporate some form of rapport which was absent here. Acceptability of ChatCharlie was not considered, yet acceptability for organisations and individuals is fundamental when considering the use of Chatbots in criminal justice contexts.

ChatCharlie may not emerge as the most desired method of providing an initial account for some groups, but in the absence of a human agent may be acceptable to many. Other witness populations may indicate a preference for this form of interaction since there is real potential for our ChatCharlie prototype to be more user-centred in terms of adjustment for neurodivergent eyewitnesses, for example. We did not consider the type of information recalled. Quantity and quality are important metrics in the first instance, but type of information and investigation relevance should be considered in the future, which would be relevant for researchers concerned with verbal overshadowing and RIF. Finally, mock witness research cannot replicate the real-world conditions experienced by witnesses but does offer an important opportunity to understand episodic memory in controlled conditions providing an indication of eyewitness cognition.

To conclude, a timely retrieval practice by way of responding verbally to initial account questions asked by a person or responding textually via a carefully managed contextual chatbot platform apparently offers benefits since in both instances later recall during an in-depth investigative interview was more detailed and more accurate. Our findings concur with the positive outcomes reported by others where witnesses are offered well-managed practice opportunities, providing empirical support for the cognitive benefits of eliciting initial accounts. We are not suggesting ChatCharlie is a replacement for professional skilled interviewers, rather we are exploring the potential of contextual conversational chatbots as an investigative and cognitive support tool in some instances.

## Methods

### Design

A between-subjects mock eyewitness paradigm with two distinct time points, namely Time 1 (IAi according to condition) and Time 2 (investigative interview) was employed. At Time 1 participants were randomly allocated to one of three Time 1 IAI conditions: ChatCharlie (conducted remotely via the ChatCharlie platform), In-person (conducted in-person) or Control (no initial account). At Time 2 (1 week later), all participants were then interviewed in-person, face-to-face using the same information gathering investigative interview protocol. Memory performance for the ChatCharlie and In-person IAI conditions were analysed at Time 1. Memory performance of all participants across conditions was analysed at Time 2. Self-reported confidence in correct and erroneous recall performance was collected at Time 1 (as appropriate) and Time 2.

### Participants

Ninety participants from the general population took part in this study: 28 male (31%), 60 female (67%), and 2 non-binary/third gender (2%). The mean age was 29.26 (SD = 8.01), ranging from 18 to 54 years. There was a non-significant difference in age across each of the three conditions,  $F(2, 87) = 2.128, p = 0.125$  ( $M_{\text{ChatCharlie}} = 30.83$ ;  $M_{\text{Control}} = 30.10$ ;  $M_{\text{In-person}} = 26.83$ ). Participants were recruited through online advertising and snowball sampling and were compensated for the time spent taking part through a £10 shopping voucher.

An *a-priori* power analysis was conducted using G\*Power 4<sup>70</sup> to determine the minimum sample size estimation to test  $H^1$ . Power analysis for MANOVA: special effects and interactions indicated the required sample size to achieve 80% power for a medium effect at a significance criterion of  $\alpha = 0.05$  was  $N = 87$ . Thus, the obtained sample size of  $N = 90$  was adequate given resource constraints and access to non-student populations and is in line with sample size norms described in recent similar empirical mock eyewitness studies<sup>71,72</sup>.

### Ethics

This research was ethically approved by the University of Westminster Research Ethics Committee, approval No ETH2021-0680. Experiments reported here were run in accordance with the British Psychological Society and UK Health Care Practitioner Council codes of ethical conduct, and the declaration of Helsinki. Written and digitally recorded verbal consent was obtained from each participant.

### Procedure

Participants consented to take part in both elements. Irrespective of Time 1 condition, all participants received a one-time link to watch a volume crime shop robbery on a laptop or desktop computer. Within 10 min of having viewed the event participants in ChatCharlie and In-person Time 1 conditions provided an initial account according to condition (see Appendices A & B) and completed an anonymous online questionnaire (see ESM Appendix C) regarding confidence in their memory performance (errors and correct recall). In-person initial accounts were elicited in-person face-to-face, and ChatCharlie initial accounts were elicited remotely via the ChatCharlie chatbot platform. Participants in the Control did not complete an initial account, nor a questionnaire at this timepoint. At Time 2 (one week later), all participants were interviewed In-person using a protocol, verbatim (ESM Appendix D). Immediately post Time 2 interviews, all participants completed an



anonymous survey concerning confidence in their memory performance (ESM Appendix E). Participants were naive to the study design, aims and hypotheses.

## Materials

The stimulus video (accessible <https://youtu.be/9MpWfQonSgk>) was 1-min in duration and depicted a non-violent volume crime shop robbery involving two perpetrators who were seen walking past a parade of shops. They then entered a small corner store and after a short delay came running out of the shop chased by the shopkeeper. The perpetrators ran away from the shop and around the corner out of sight.

### *Initial account interviews (Time 1)*

Both the In-person (see ESM Appendix A) and ChatCharlie (see ESM Appendix B) initial account Time 1 condition protocols draw on the CoP initial account guidance. Both focus on specific event details, individual and environmental factors and comprised ground rule instructions, open-ended invitations and closed and forced choice questions as appropriate. Confidence in correct recall was collected using a self report scale ranging from 1 (not at all confident that the information I have reported is correct) to 5 (completely confident that the information I have reported is correct) at Time 1 and again at Time 2.

### *Investigative interview (Time 2)*

Time 2 investigative interviews were developed guided by the current approach to interviewing witnesses in the UK and elsewhere to best manage variability without limiting witness responses. Interviews comprised a series of distinct but seamless phases (see ESM Appendix C) namely (1) engage and explain, (2) retrieval instructions, (3) free recall, (4) cued questioning, and (5) closure. Of these, two were retrieval focused, the free recall and cued questioning phases. All Time 2 interviews were conducted by the same interviewer In-person.

## Recall performance coding

Time 1 In-person initial account interviews and all Time 2 investigative interviews were digitally audio and video recorded and transcribed verbatim. ChatCharlie conversation transcripts were downloaded from the chatbot platform. All transcripts were coded for correct, erroneous (information relevant to the witnessed episode but described with error, e.g., describing a person's brown jacket, but stating that it was black instead of brown), or confabulated (reporting information that was not present in the film) information recalled. All information from the Time 1 initial account interviews and all information provided in the two retrieval phases of the Time 2 interviews, along with the position in the Time 2 interview the information was recalled was coded (i.e., whether recalled in the Free Recall or Questioning phase). Items recalled were only scored once (i.e., repetitions were not scored).

Fifteen Time 1 interviews and 15 Time 2 interviews from across the experimental conditions were randomly selected for recoding by an independent coder blind to the aims and hypotheses of the research but familiar with the method of scoring (RD). A series of Two-way mixed effects Intraclass Correlation Coefficient (ICC) analysis testing for absolute agreement between coders for the overall amount of correct, erroneous, and confabulated recall coding were conducted. Mean estimations with 95% CI reveal very good inter-rater reliability for Time 1 correct, ICC = 0.998 (95% CI 0.993, 0.999), errors, ICC = 0.922 (95% CI 0.766, 0.974) and confabulations, 0.944 (95% CI 0.835, 0.981) and Time 2 correct, ICC = 0.916 (95% CI 0.749, 0.972), errors, ICC = 0.840 (95% CI 0.524, 0.926) and confabulations, 0.921 (95% CI 0.765, 0.973).

## Interviewer adherence to protocols

### *Time 1 in-person initial account interviews*

Three interviewers (MZ; SA; CD) conducted the 30 In-person initial account interviews according to a protocol, verbatim (see ESM Appendix A). Interviewers conducted 11 (MZ), 10 (SA) and 9 (CD) first account interviews each. The first account protocol comprised a short introduction and rapport building phase followed by a series of fixed questions. A total of 6 (20%) initial account interviews (2 from each interviewer) were randomly selected and coded by two independent coders for interviewer adherence to the protocol for i) correct number of questions asked, ii) questions asked in the correct order, and iii) introductions and short rapport using a scoring sheet where 1 = absent, 2 = partially present and 3 = present, for each of the 3 protocol elements. Prior to coding, coders participated in a training session held by the first author, during which the initial account interview protocols and coding system were explained. Coders then practiced coding and discussed any disagreements/misunderstandings with the trainer to reach a consensus using example interviews. Two-way mixed effects intraclass correlation coefficient (ICC) analysis testing for absolute agreement between coders indicated good/very good inter-rater reliability for all interviewer behaviours: Kruskal–Wallis H tests for the presence/absence of each of the three interview instructions/phases revealed non-significant differences across the three interviewers, all  $H_s(2) < 2.090$ , all  $p_s > 0.598$ , hence all were similarly present/correct across interviewers.

### *Time 2*

Time 2 interviews were all conducted by the same interviewer (CA). Eighteen interviews (20%) were randomly selected for interviewer adherence to the Time 2 protocol and coded by the same two independent coders from Time 1 using the same coding approach as Time 1 for each of the Time 2 protocol elements. Two-way mixed effects intraclass correlation coefficient (ICC) analysis testing for absolute agreement between coders indicated good/very good inter-rater reliability for the presence and application of each of the Time 2 interview phases: engage and explain, ICC = 0.938 (95% CI 0.843; 0.975), retrieval instructions, 1.00 (95% CI 1.00; 1.00), free recall, ICC = 0.920 (95% CI -0.449; 0.362), questioning, ICC = 0.899 (95% CI 0.593; 0.975), and closure, ICC = 1.00 (95% CI 1.00; 1.00).

## Data availability

Raw datafiles and materials are available from the first author and have been made available via the OSF: <https://osf.io/> platform.

Received: 30 June 2024; Accepted: 4 March 2025

Published online: 19 March 2025

## References

- Lacy, J. W. & Stark, C. E. The neuroscience of memory: implications for the courtroom. *Nat. Rev. Neurosci.* **14**(9), 649–658 (2013).
- Seale-Carlisle, T. M., Quigley-McBride, A., Teitcher, J. E., Crozier, W. E., Dodson, C. S., & Garrett, B. L. (2024). New insights on expert opinion about eyewitness memory research. *Perspect. Psychol. Sci.* 17456916241234837.
- Tulving, E. (2000). Concepts of memory. in *The Oxford Handbook of Memory*, 33–43.
- Tulving, E. Episodic memory: From mind to brain. *Annu. Rev. Psychol.* **53**(1), 1–25 (2002).
- Wells, G. L. Scientific study of witness memory: Implications for public and legal policy. *Psychol. Public Policy Law* **1**(4), 726 (1995).
- Greenspan, R. L. & Loftus, E. F. Eyewitness confidence malleability: Misinformation as post-identification feedback. *Law Hum. Behav.* **44**(3), 194 (2020).
- Zaragoza, M. S., Belli, R. F., & Payment, K. E. Misinformation effects and the suggestibility of eyewitness memory. in *Do Justice and Let the Sky Fall*, 35–63. (Psychology Press, 2013).
- Dando, C. J., Wilcock, R., Behnkle, C. & Milne, R. Modifying the cognitive interview: Countenancing forensic application by enhancing practicability. *Psychol. Crime Law* **17**(6), 491–511 (2011).
- Fisher, R. P., Schreiber Compo, N., Rivard, J., & Hirn, D. (2014). Interviewing witnesses. in *The SAGE Handbook of Applied Memory*, 559–578.
- Milne, R., & Bull, R. (2016). Witness interviews and crime investigation. in *An Introduction to Applied Cognitive Psychology*, 175–196.
- Walsh, D., Milne, B. & Bull, R. One way or another? Criminal investigators' beliefs regarding the disclosure of evidence in interviews with suspects in England and Wales. *J. Police Crim. Psychol.* **31**, 127–140 (2016).
- Milne, R., & Bull, R. (2003). Interviewing by the police. in *Handbook of Psychology in Legal Contexts*, 109–125.
- Bull, R. The investigative interviewing of children and other vulnerable witnesses: Psychological research and working/professional practice. in *Investigating the Truth*, 126–146. (Routledge, 2018).
- Shawyer, A., Milne, B., & Bull, R. Investigative interviewing in the UK. in *International Developments in Investigative Interviewing*, 24–38. (Willan, 2013).
- Dando, C. J., Gabbert, F. & Hope, L. Supporting older eyewitnesses' episodic memory: The self-administered interview and sketch reinstatement of context. *Memory* **28**(6), 712–723 (2020).
- Dando, C., Taylor, D. A., Caso, A., Nahouli, Z. & Adam, C. Interviewing in virtual environments: Towards understanding the impact of rapport-building behaviours and retrieval context on eyewitness memory. *Memory Cognition* **51**(2), 404–421 (2023).
- Harvey, M. B. (2024). *Differential Impacts of Investigation and Witness Delays on Layperson Perceptions of Witness Credibility* (Doctoral dissertation, SIMON FRASER UNIVERSITY).
- Paterson, H. M., Eijkemans, H. & Kemp, R. I. Investigating the impact of delayed administration on the efficacy of the Self-Administered Interview. *Psychiatry Psychol. Law* **22**(2), 307–317 (2015).
- Squire, L. R. & Alvarez, P. Retrograde amnesia and memory-consolidation: A neurobiological perspective. *Curr. Opin. Neurobiol.* **5**(2), 169–177 (1995).
- Squire, L. R., Genzel, L., Wixted, J. T. & Morris, R. G. Memory consolidation. *Cold Spring Harbor Perspect. Biol.* **7**(8), a021766 (2015).
- Dudai, Y. The restless engram: Consolidations never end. *Annu. Rev. Neurosci.* **35**, 227–247 (2012).
- Cowan, E. T., Schapiro, A. C., Dunsmoor, J. E., & Murty, V. P. (2021). Memory consolidation as an adaptive process. *Psychonomic Bull. Rev.* 1–15.
- Racsmány, M., Conway, M. A. & Demeter, G. Consolidation of episodic memories during sleep: Long-term effects of retrieval practice. *Psychol. Sci.* **21**(1), 80–85 (2010).
- Horry, R., Hughes, C., Sharma, A., Gabbert, F. & Hope, L. A meta-analytic review of the Self-Administered Interview®: Quantity and accuracy of details reported on initial and subsequent retrieval attempts. *Appl. Cognitive Psychol.* **35**(2), 428–444 (2021).
- Hope, L., Gabbert, F. & Fisher, R. P. From laboratory to the street: Capturing witness memory using the Self-Administered Interview. *Legal Criminol. Psychol.* **16**(2), 211–226 (2011).
- Gabbert, F., Hope, L., Horry, R., Drain, T. & Hughes, C. Examining the efficacy of a digital version of the Self-Administered Interview. *Comput. Human Behav. Rep.* **5**, 100159 (2022).
- Brown, A., Kumar, A. T., Melamed, O., Ahmed, I., Wang, Y. H., Deza, A., et al. (2023). A Motivational Interviewing Chatbot With Generative Reflections for Increasing Readiness to Quit Smoking: Iterative Development Study. *JMIR Mental Health.* **10**, e49132.
- Han, X., Zhou, M., Turner, M. J., & Yeh, T. (2021). Designing effective interview chatbots: Automatic chatbot profiling and design suggestion generation for chatbot debugging. in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–15).
- Minhas, R., Elphick, C., & Shaw, J. (2022). Protecting victim and witness statement: examining the effectiveness of a chatbot that uses artificial intelligence and a cognitive interview. *AI & SOCIETY.* 1–17.
- Nadarzynski, T., Miles, O., Cowie, A. & Ridge, D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digital Health* **5**, 2055207619871808 (2019).
- Nadarzynski, T. et al. Barriers and facilitators to engagement with artificial intelligence (AI)-based chatbots for sexual and reproductive health advice: A qualitative analysis. *Sexual Health.* **18**(5), 385–393 (2021).
- Palanica, A., Flaschner, P., Thommandram, A., Li, M. & Fossat, Y. Physicians' perceptions of chatbots in health care: Cross-sectional web-based survey. *J. Med. Internet Res.* **21**(4), e12887 (2019).
- Følstad, A. & Brandtzaeg, P. B. Users' experiences with chatbots: Findings from a questionnaire study. *Qual. User Exp.* **5**(1), 3 (2020).
- Zhang, R. W., Liang, X. & Wu, S. H. When chatbots fail: Exploring user coping following a chatbots-induced service failure. *Inform. Technol. People* **37**(8), 175–195 (2024).
- Chhabra, S., Kaushal, V., & Girija, S. (2024). Determining the causes of user frustration in the case of conversational chatbots. *Behav. Inform. Technol.* 1–19.
- Følstad, A., & Skjuve, M. (2019). Chatbots for customer service: user experience and motivation. in *Proceedings of the 1st international conference on conversational user interfaces* (pp. 1–9).
- Khamaj, A. AI-enhanced chatbot for improving healthcare usability and accessibility for older adults. *Alexandria Eng. J.* **116**, 202–213 (2025).
- Kettle, L. & Lee, Y. C. User experiences of well-being Chatbots. *Human Factors* **66**(6), 1703–1723 (2024).
- Nordheim, C. B., Følstad, A. & Bjørkli, C. A. An initial model of trust in chatbots for customer service—Findings from a questionnaire study. *Interact. Computers* **31**(3), 317–335 (2019).

40. Følstad, A. & Taylor, C. Investigating the user experience of customer service chatbot interaction: A framework for qualitative analysis of chatbot dialogues. *Qual. User Exp.* **6**(1), 6 (2021).
41. Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. in *Internet Science: 4th International Conference, INSCI 2017, Thessaloniki, Greece, November 22–24, 2017, Proceedings 4* (pp. 377–392). Springer International Publishing.
42. Araujo, T. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers Human Behav.* **85**, 183–189 (2018).
43. Cuff, B. M., Brown, S. J., Taylor, L. & Howat, D. J. Empathy: A review of the concept. *Emot. Rev.* **8**(2), 144–153 (2016).
44. Dando, C. J. & Oxburgh, G. E. Empathy in the field: Towards a taxonomy of empathic communication in information gathering interviews with suspected sex offenders. *Eur. J. Psychol. Appl. Legal Context* **8**(1), 27–33 (2016).
45. Nahouli, Z., Dando, C. J., Mackenzie, J. M. & Aresti, A. Rapport building and witness memory: Actions may ‘speak’ louder than words. *PLoS One* **16**(8), e0256084 (2021).
46. Rosenfeld, P. J. & Olson, J. M. Bayesian data analysis: A fresh approach to power issues and null hypothesis interpretation. *Appl. Psychophysiol. Biofeedback* **46**, 135–140 (2021).
47. Mackay, T. L. & Paterson, H. M. How does timing of recall affect eyewitness memory and psychological distress?. *J. Police Criminal Psychol.* **30**, 242–253 (2015).
48. Fatima, J. K. et al. Rapport with a chatbot? The underlying role of anthropomorphism in socio-cognitive perceptions of rapport and e-word of mouth. *J. Retailing Consumer Services* **77**, 103666 (2024).
49. Lee, J., Lee, D. & Lee, J. G. Influence of rapport and social presence with an AI psychotherapy chatbot on users’ self-disclosure. *Int. J. Human–Computer Interaction* **40**(7), 1620–1631 (2024).
50. Bjorklund, D. F. et al. Social demand characteristics in children’s and adults’ eyewitness memory and suggestibility: The effect of different interviewers on free recall and recognition. *Appl. Cognit. Psychol.* **14**(5), 421–433 (2000).
51. Blank, H. & Launay, C. How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *J. Appl. Res. Memory Cognition* **3**(2), 77–88 (2014).
52. Caso, A., Gabbert, F. & Dando, C. J. Eyewitness confidence in the interviewing context: Understanding the impact of question type and order. *Appl. Cognit. Psychol.* **38**(3), e4197 (2024).
53. Fisher, R. P., & Schreiber, N. (2017). Interview protocols to improve eyewitness memory. in *The Handbook of Eyewitness Psychology: Volume I* (pp. 53–80). (Psychology Press).
54. Harlow, I. M. & Yonelinas, A. P. Distinguishing between the success and precision of recollection. *Memory* **24**(1), 114–127 (2016).
55. Taylor, D. A. & Dando, C. J. Eyewitness memory in face-to-face and immersive avatar-to-avatar contexts. *Front. Psychol.* **9**, 346899 (2018).
56. Newcomb, A. B. et al. Building rapport and earning the surgical patient’s trust in the era of social distancing: teaching patient-centered communication during video conference encounters to medical students. *J. Surg. Education* **78**(1), 336–341 (2021).
57. Tahan, H. A. & Sminkey, P. V. Motivational interviewing: Building rapport with clients to encourage desirable behavioral and lifestyle changes. *Professional Case Manag.* **17**(4), 164–172 (2012).
58. Jondani, J. A., Yazdkhasti, F., & Abedi, A. (2023). Memory confidence and memory accuracy deterioration following repeated checking: A systematic review and meta-analysis. *J. Behav. Therapy Exp. Psychiatry*. 101855.
59. Schooler, J. W. & Engstler-Schooler, T. Y. Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognit. Psychol.* **22**(1), 36–71 (1990).
60. Wilson, B. M., Seale-Carlisle, T. M. & Mickes, L. The effects of verbal descriptions on performance in lineups and showups. *J. Exp. Psychol. General* **147**(1), 113 (2018).
61. Smith, H. M. & Flowe, H. D. ROC analysis of the verbal overshadowing effect: Testing the effect of verbalisation on memory sensitivity. *Appl. Cognit. Psychol.* **29**(2), 159–168 (2015).
62. Meissner, C. A. & Brigham, J. C. A meta-analysis of the verbal overshadowing effect in face identification. *Appl. Cognit. Psychol.* **15**(6), 603–616 (2001).
63. Meissner, C. A., Brigham, J. C. & Kelley, C. M. The influence of retrieval processes in verbal overshadowing. *Memory Cognit.* **29**, 176–186 (2001).
64. Anderson, M. C., Bjork, E. L. & Bjork, R. A. Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychon. Bull. Rev.* **7**, 522–530 (2000).
65. Murayama, K., Miyatsu, T., Buchli, D. & Storm, B. C. Forgetting as a consequence of retrieval: A meta-analytic review of retrieval-induced forgetting. *Psychol. Bull.* **140**(5), 1383 (2014).
66. Migueles, M. & García-Bajos, E. Selective retrieval and induced forgetting in eyewitness memory. *Appl. Cognit. Psychol.* **21**(9), 1157–1172 (2007).
67. García-Bajos, E., Migueles, M. & Anderson, M. C. Script knowledge modulates retrieval-induced forgetting for eyewitness events. *Memory* **17**(1), 92–103 (2009).
68. Greer, J. et al. Effortful retrieval of semantic memories induces forgetting of related negative and neutral episodic memories. *Cognition* **251**, 105908 (2024).
69. Gustafsson, P. U., Lindholm, T. & Jönsson, F. U. Predicting accuracy in eyewitness testimonies with memory retrieval effort and confidence. *Front. Psychol.* **10**, 703 (2019).
70. Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. G\* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**(2), 175–191 (2007).
71. Chevroulet, C., Paterson, H. M., Kemp, R. I., & van Golde, C. (2022). Technology and eyewitness memory: evaluating the efficacy of a novel digital cued recall tool. *Psychol. Crime Law*. 1–18.
72. Anakwah, N., Horselenberg, R., Hope, L., Amankwah-Poku, M. & Van Koppen, P. J. Cross-cultural differences in eyewitness memory reports. *Appl. Cognit. Psychol.* **34**(2), 504–515 (2020).

## Author contributions

CD & CA made equal contributions to design and materials. Data collection was managed by CA. All Time 2 interviews were conducted by CA. All coding was equally coordinated by CD and CA. Analysis was conducted by CD with CA. CD completed further additional analysis. The manuscript was prepared and written by CD, with some assistance from CA centred on the writing of the Methods.

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-93281-1>

[0.1038/s41598-025-93281-1](https://doi.org/10.1038/s41598-025-93281-1).

**Correspondence** and requests for materials should be addressed to C.J.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025