Witnesses' verbal evaluation of certainty and uncertainty during investigative interviews: Relationship with report accuracy

To appear in Journal of Police and Criminal Psychology

Rui M. Paulo

College of Liberal Arts, Bath Spa University

Pedro B. Albuquerque

School of Psychology, University of Minho

Ray Bull

Department of Law and Criminology, University of Derby

May 2019

Witnesses' verbal evaluation of certainty and uncertainty during investigative interviews: Relationship with report accuracy

Abstract

The Enhanced Cognitive Interview (CI) is a widely studied method to gather informative and accurate testimonies. Nevertheless, witnesses still commit errors and it can be very valuable to determine which statements are more likely to be accurate or inaccurate. This study examined whether qualitative confidence judgments could be used to evaluate report accuracy in a timesaving manner. Forty-four participants watched a mock robbery video and were interviewed 48hours later with a revised CI. Participants' recall was categorized as: (1) evaluated with very high confidence (certainties); (2) recalled with low confidence utterances (uncertainties); or (3) recalled with no confidence markers (regular recall). Certainties were more accurate than uncertainties and regular recall. Uncertainties were less accurate than regular recall, thus its exclusion raised participants' report accuracy. Witnesses were capable of qualitatively distinguish between highly reliable information, fairly reliable information and less reliable information in a timesaving way. Such a distinction can be important for investigative professionals who do not know what happened during the crime and may want to estimate which information is more likely to be correct.

Keywords: Cognitive Interview; Uncertainties; Certainties; Confidence; Accuracy

1. INTRODUCTION

Interviewing witnesses is a crucial procedure that can determine the outcome of many police investigations (Fisher & Geiselman, 1992). However, memory is seldom fully accurate and several errors can occur during the encoding, storage and/or retrieval phase (Manning & Loftus, 1996). Thus, witnesses' reports rarely correspond with the witnessed event as witnesses frequently omit important information and/ or report incorrect details. Nevertheless, using appropriate interviewing techniques can increase the quality of witnesses' reports and produce informative testimonies (Paulo, Albuquerque, & Bull, 2013).

With the purpose of obtaining a more accurate and informative report, Fisher and Geiselman (1992) developed the Enhanced Cognitive Interview (CI). This interview comprises four cognitive mnemonics (report everything, mental reinstatement of context, change order and change perspective) and several social and communicative components (e.g., rapport building, witness-compatible questioning, transferring control of the interview to the witness and mental imagery) that are important for conducting good investigative interviews. For more information about these CI components as well as the theory and research underlying this interview protocol, see Paulo et al. (2013).

Several studies have shown witnesses are able to recall more accurate information when interviewed with the CI in comparison with other interview protocols, such as a structured or standard interview. This finding is frequently referred to as the CI superiority effect (Geiselman & Fisher, 1988; Geiselman & Latts, 1990; Geiselman et al., 1994; Higham & Memon, 1999; Köhnken, Milne, Memon, & Bull, 1999; Memon, Wark, Bull, & Köhnken, 2011; Rivard, Fisher,

Robertson, & Mueller, 2014). This effect has been found both in laboratory and field studies (Colomb & Ginet, 2012; Colomb, Ginet, Wright, Demarchi, & Sadler, 2013), in many countries - e.g., USA, UK, Brazil, Portugal (Paulo, Albuquerque, Saraiva, & Bull, 2015b; Stein & Memon, 2006), with different witnesses - e.g., children, adults, elderly (Brown & Geiselman, 1990; Verkampt & Ginet, 2010; Wright & Holliday, 2006); with a range of delays between the crime and the interview - minutes to months (Larsson, Granhag, & Spjut, 2002), and a variety of to-be-remembered events - e.g., crimes, traffic accidents, and phone calls (Chapman & Perry, 1995).

Recently, several studies focused on testing new procedures that could enhance the CI (Brelet et al., 2018; Brunel, Py, and Launay, 2013; Paulo, Albuquerque, and Bull, 2016). For instance, Brunel et al. (2013) found using the open depth instruction, i.e. asking participants to focus their attention on the small details while recalling the event, enhanced recall in comparison with the change perspective mnemonic. Paulo et al. (2016) used a revised CI with a new recall strategy (Category Clustering Recall) and without the change perspective mnemonic and found participants interviewed with the revised CI were able to recall a considerably higher amount of correct information in comparison with participants interviewed with the conventional CI. Other authors focused on increasing CI report accuracy and developing new procedures to evaluate which parts of the testimony were more likely to be reliable (Allwood, Ask, & Granhag, 2005; Paulo, Albuquerque, & Bull, 2015a; Roberts & Higham, 2002). One method to evaluate this could be to use witnesses' metacognitive monitoring (Evans & Fisher, 2011), as we will address below.

Metacognition refers to what we know about our own cognition and how this knowledge can be used to regulate cognition, which includes metamemory, i.e., what

we know about our own memory and mnemonic strategies and how we can use this information to improve it (Metcalfe & Shimamura, 1994). Several studies have now addressed how different metacognitive monitoring techniques (e.g., confidence judgments, adjusting report precision, report option or frequency judgments) can be used to improve or evaluate eyewitnesses' accuracy (Evans & Fisher, 2011; Higham, Luna, & Bloomfield, 2011; Koriat & Goldsmith, 1996; Roberts & Higham, 2002). For instance, asking witnesses to estimate how many questions they have answered correctly or incorrectly (frequency judgments) could be a way to evaluate report accuracy (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Liberman, 2004; Sniezek & Buckley, 1991). However, several studies have questioned frequency judgments' accuracy during investigative interviews (Granhag, Jonsson, & Allwood, 2004; Paulo et al., 2015b). As an example, Paulo et al. (2015b) asked participants to perform frequency judgments during an investigative interview and found no association between these judgments and participants' actual accuracy, suggesting frequency judgments not to be valid. Another alternative for evaluating report accuracy might be using confidence judgments (Wixted, Mickes, & Fisher, 2018). In some situations, such as selections from lineups (Brewer, Weber, Wootton, & Lindsay, 2012; Lindsay et al., 2013), cued recall (Luna & Martín-Luengo, 2012), or free recall (Allwood et al., 2005), several studies suggest a positive relationship between confidence and accuracy (C-A relationship). Nonetheless, this relationship's magnitude seems to vary due to different factors such as: (1) the statistical analysis used – even though recent studies that used more informative types of analysis like calibration found a C-A relationship, early studies where correlations were used did not find a C-A relationship (Luna & Martín-Luengo, 2012; Wixted & Wells, 2017); (2) the moment when confidence is assessed – a higher C-A relationship is found

when confidence is assessed immediately after the response is provided (Brewer & Weber, 2008); or (3) if confidence is assessed between- or within- participants.

Regarding this last factor, most studies have focused on what Smith, Kassin, and Ellsworth (1989) call a between-participants C-A relationship, for instance, comparing the accuracy of 'confident' witnesses against the accuracy of less confident witnesses in line-up identifications. However, in an investigative interview, the within-participants C-A relationship (Smith et al., 1989) could be more informative and very distinct from this. In fact, if the information a witness (sometimes the only witness) recalls with higher confidence ('I'm sure the robber had a gun') is more reliable than information recalled with less confidence ('I'm unsure the robber was wearing black') that may be very important for detectives and other legal professionals who do not know what happened during the crime and may need to estimate which information is more likely to be reliable (Brewer & Weber, 2008; Buratti, Allwood, & Johansson, 2014; Potter & Brewer, 1999). Some authors have addressed this issue (Kebbell, 2009; Perfect, 2002). Kebbell (2009) looked at the C-A relationship during a questioning phase and found a high within-participants C-A relationship, particularly for the 'absolutely sure' responses, which were usually accurate. In two studies involving the CI (Allwood et al., 2005; Roberts & Higham, 2002) witnesses were interviewed with either the CI or a Structured Interview and later asked to provide confidence judgments for their statements with numerical rating scales. Witnesses were able to distinguish between their more and less reliable recall, as the statements portions assigned with high confidence were more reliable than the full set of statements. However, these studies focused on metacognitive judgments performed after the interview was conducted. This is, after finishing the interview the witness rated parts of her report in terms of confidence. With this

procedure, Paulo et al. (2015a) identified some concerns. First, these procedures are very time demanding as the interviewer would have to select the information that will be evaluated by the interviewee, who also needs time to perform these confidence judgments using the numerical scales provided. Thus, when using these numerical scales, only a limited number of units of information, selected by the interviewer, could be rated in terms of confidence as it would take too much time to apply these scales to every unit of information the witness reported. Furthermore, Kebbell (2009) wisely stated confidence needs to be communicated to jurors or investigators in a natural and practical way and this is very unlikely to occur through the use of confidence scales which seem to be unfit for real police interviews (Paulo et al., 2015a). Thus, one could focus instead on how witnesses do this spontaneously during the interview.

Paulo et al. (2015a) suggested using witnesses' spontaneous verbal utterances of uncertainty (e.g., I think; maybe; I believe, etc.) to identify information that is less likely to be reliable. They found regardless of interview condition (Cognitive Interview or Structured Interview) when participants spontaneously used such verbal utterances to report somewhat uncertain information (e.g., 'I think the robber was armed') this information (which they named uncertainties) was more likely to be incorrect in comparison with their remaining recall (i.e. when such expressions were not used). Nonetheless, Paulo et al. (2015a) focused on information recalled with low confidence utterances (uncertainties) only, and they suggested studying more confidence levels, such as information evaluated with very high confidence (certainties), this is, when participants state they are positively sure a unit of information they reported is correct ('I am definitely sure the robber was armed'). However, Paulo et al. (2015a) found participants rarely spontaneously provided

information recalled with very high confidence utterances (certainties) and a different experimental design would be necessary to have a number of elicited certainties that would allow including this confidence level in the accuracy analysis.

1.1. Current Study

In the present study, we devolved an experimental design that allowed this assessment to be made. Participants were asked during the summary phase to retrospectively state when they were completely sure a unit of information they had previously reported was accurate. This procedure allowed us to test whether information evaluated with very high confidence ('I am definitely sure the robber was armed') would be more reliable than (a) information recalled with low confidence utterances - uncertainties (e.g., 'I think the robber was armed') and (b) information recalled with no confidence markers (e.g., 'The robber was armed'). It was hypothesized a C-A relationship would be found (Paulo et al. 2015a), i.e., information recalled with low confidence utterances (uncertainties) would be less reliable than information recalled with no confidence markers (regular recall) and information evaluated retrospectively with very high confidence. If this C-A relationship exists, it can have important implications for police officers and other legal professionals because if witnesses are capable of distinguishing more reliable recall (in the form of certainties) from their fairly reliable recall (regular recall) and less reliable recall (uncertainties), these indices may be used in a forensically-relevant and practical way (Brewer & Weber, 2008).

2. METHOD

2.1. Participants

Forty-four Portuguese psychology students, 41 females and 3 males, with an age range from 17 to 48 years old (M = 19.64, SD = 5.18) participated in this study for course credits.

2.2. Design

We used a within-subjects design with the confidence level participants qualitatively assigned to the recalled information (see coding section) as the main independent variable with three levels: (1) information recalled with low confidence utterances: uncertainties; (2) information evaluated with very high confidence: certainties; and (3) information recalled with no confidence markers: regular recall. Our main dependent variables were the amount of reported information and accuracy, measured in units of information and proportion, respectively.

2.3. Materials

Participants watched a non-violent clip of a bank robbery on a Fujitsu L7ZA LCD computer screen. The video recording was edited from the second episode of the 2004 Portuguese television drama “Inspector Max” (Riccó & Riccó, 2004) and was three minutes and eleven seconds long. This non-violent video recording shows an armed male subject walking inside a bank and taking several hostages to carry out a robbery. He verbally and physically interacts with them, with the cashier and a police officer who later approaches the robber.

2.4. Procedure

Ethics committee approval was obtained. Having signed a consent form after reading general information about the study, participants took part in two sessions. During the first session, they were shown the video recording. They were asked to pay as much attention as possible to the video recording because they would be later interviewed about it. A second session took place approximately forty-eight hours later and each participant was interviewed with a modified CI (Paulo et al., 2016). All interviews were video and audio recorded.

2.4.1. Interview protocol. The interview protocol had previously been translated and adapted for the Portuguese language (Paulo et al., 2015b). This interview protocol was very similar (see Table I) to a CI protocol (Fisher & Geiselman, 1992), but included Category Clustering Recall which proved to be effective for further enhancing recall (Paulo et al., 2016, Paulo, Albuquerque, Vitorino, & Bull, 2017) and excluded the change perspective mnemonic which has shown to be less effective (Bensi, Nori, Gambetti, & Giusberti, 2011; Davis, McMahon, & Greenwood, 2005). No control interview (e.g., a Structured Interview) was used as previous studies found witnesses' ability to monitor their accuracy to be unaffected by interview condition (Allwood et al., 2005; Paulo et al., 2015a).

The interview protocol used in this study included three of the four CI cognitive mnemonics: Report Everything, Context Reinstatement, and Change Order, as well as all the social and communicative components described in Fisher and Geiselman (1992) such as rapport building, transfer of control, appropriate questioning (e.g., witness-compatible questioning) and mental imagery. Fisher and

Geilseman's (1992) guidelines for conducting a good investigative interview were followed.

All interviews included seven main phases: (1) preliminary phase; (2) free report; (3) open-ended questioning; (4) second recall; (5) third recall; (6) summary; and (7) closure. A full description of the interview protocol according to the interview phase is included (see appendix).

Information evaluated with very high confidence (certainties) was coded retrospectively on phase 6 (summary). In this interview phase, the interviewer reported back to the witness all the relevant details she/ he previously recalled during the preceding interview phases and asked the witness to correct him if he misheard or misinterpreted any part of the statement. The interviewee could also report any new detail she/he might remember. This type of summary is frequently included at the end of real police interviews (Griffiths & Milne, 2010) and research interview protocols (Milne & Bull, 2003, Paulo et al., 2013, Paulo et al., 2015b). However, in this study, another procedure was added to the summary phase in order to evaluate certainty. Participants were asked to clearly state whenever they could undoubtedly declare a unit of information they had previously recalled and was now being reported to them was accurate. This is, the interviewer asked participants to state when they were absolutely sure a unit of information that was now being summarized corresponded fully with what happened in the event: (…) *Please tell me when you are absolutely sure a detail you previously reported and I am now telling you corresponds exactly to what happened in the video* (…) *For instance, I believe if I told you previously reported there was a robber, you could state with absolute certainty the robber was present in the video (...) This is the kind of certainty level I'm looking for (...).* When the participant told the interviewer he or she was

absolutely sure a unit of information previously reported was accurate (e.g., the presence of a weapon) this unit of information was coded as information evaluated with very high confidence (certainty). This evaluation of certainty was always performed for units of information (e.g., presence of a weapon) instead of larger statements (e.g., "there as a gun, which was a shotgun, and it was black") as participants might be certain about a specific unit of information (e.g., presence of a weapon) but uncertain about other units of information reported in the same statement (e.g., the colour and type of weapon). Participants were instructed to perform this task at the summary phase because: (1) as mentioned above, participants do not spontaneously provide a considerable number of certainties unless they are instructed to do so (Paulo et al., 2015a); (2) by using this procedure in one of the last interview phases we assured participants' recall during the previous interview phases (e.g., second recall attempt) was not being influenced by such a procedure.

2.4.2 Interviewer training. An expert in the CI who had followed several qualified courses on investigative interview techniques consisting of more than 50 lecture hours, practice, role-playing exercises, and feedback/ evaluation conducted all the interviews. To assure the interviewer performance was adequate and consistent across interviews, the interview protocol was read verbatim whenever possible (e.g., open-ended questioning and summary phase need to be adapted according to the participants' previous recall). Furthermore, a researcher who was not involved in this study randomly checked 25% of the interviews with a structured evaluation grid that included parameters such as the type of questions used, rapport, instructions clarity, interviewer's posture and behavior, amongst others. She/ he concluded all interviews were adequately conducted.

2.4.3. Coding. Interview recordings were coded with the template scoring technique from Paulo et al. (2016). A comprehensive list of details in the video event was compiled resulting in 378 units of information. Recalled information was classified as either correct, incorrect (e.g., saying the pistol was brown when it was black) or confabulation (mentioning a detail or event that was not present or did not happen). Also noted was the interview phase a unit of information was recalled. If a unit of information (correct or not) was repeated during the same or a subsequent phase, this unit of information was coded as a new detail only the first time it was mentioned (Prescott, Milne, & Clark, 2011). The confidence level participants qualitatively assigned to the recalled information was also coded as: (1) information evaluated with very high confidence (certainties) – when participants retrospectively stated during the summary phase they were definitely sure a unit of information they previously recalled was correct; (2) information recalled with low confidence (uncertainties) - when participants spontaneously showed uncertainty (e.g., I think; Maybe; I believe, I'm not sure, Possibly, etc.) to communicate to the interviewer they were unsure about the accuracy of such information (e.g., 'I believe he had a jacket'); (3) information recalled with no confidence markers (regular recall) – when participants recalled information which had neither an uncertainty adjacent expression, nor was retrospectively classified as certain during the summary phase (e.g., 'He had a gun'). Therefore, participants did not assign any qualitative confidence judgment to these last units of information. As addressed below, inter-rater reliability was calculated to assess whether different researchers agreed on how to categorize this information. Subjective statements or opinions were disregarded (e.g., 'The robber was gorgeous!').

2.4.4. Inter-rater reliability. To assess inter-rater reliability 12 (25%) interviews were selected randomly and scored independently by a researcher who was naive to the experiment aims and hypotheses but familiar with the template scoring method and had access to the crime video. Intraclass correlation coefficients (ICC) were calculated for correct information, incorrect information, and confabulations, as well as for information retrospectively evaluated with very high confidence (certainties), information recalled with low confidence utterances (uncertainties), and regular recall (when no confidence markers were present). High inter-rater reliability was found for all measures in that the ICC values ranged between .977 and 1.000, with an overall ICC of .995.

3. Results

3.1. Overall interview time, recall, and accuracy

The interviews took an average of 51 minutes (SD = 9). Overall, participants were able to recall a substantial number of units of information (M = 109, SD = 22, 95% CI [102, 115]) with high accuracy (M = .91, SD = .04, 95% CI [.90, .92]), i.e., high correct recall proportion: ratio between the number of correct units of information recalled over all units of information recalled.

We first conducted a one-way within-subjects ANOVA to see if the number of correct units of information participants newly recalled varied across interview phases. As foreseeable, there was a progressive decline in new information recall at later interview phases, F (1.981, 85.196) = 172.70, $p < .001$, $\eta_p^2 = .80$. Pairwise comparisons revealed differences between all interview phases concerning the

number of newly recalled units of information, except between phase 2 (free report) and phase 3 (questioning) as both elicited a similar (and high) number of newly recalled details, t (43) = 1.59, p = .119, d = .35 (see Table I). Phase 1 (preliminary phase) was not included in this (and subsequent) analysis because participants were not requested to recall information at this phase.

Insert Table I

We then conducted three one-way within-subjects analysis of variance (ANOVA) to see if interview phase had an effect on three different measures of report accuracy: (1) correct recall proportion (the ratio between the number of correct units of information recalled, over all recalled units of information), F (1.955, 35.195) = 1.87, p = .170, η_p^2 = .09; (2) error proportion (ratio between the number of errors produced over all produced units of information), F (1.790, 34.014) = 1.18, p = .316, η_p^2 = .06; and (3) confabulation proportion (ratio between the number of confabulated units of information over all produced units of information), F (1.236, 24.720) = 1.24, p = .522, η_p^2 = .04. In sum, report accuracy was high and similar for all interview phases (see Table I).

3.2. Confidence-accuracy relationship

A one-way within-subjects ANOVA was then conducted to see if report accuracy (correct recall proportion: ratio between the number of correct units of information recalled, over all recalled units of information) varied according to participants' confidence (very high: certainties vs. no markers: regular recall vs. low: uncertainties). We found differences in report accuracy according to participants'

confidence, F (1.203, 50.509) = 35.36, $p < .001$, $\eta_p^2 = .46$ (see table II). Pairwise comparisons revealed: (1) information evaluated with very high confidence (certainties) was more accurate than information recalled with no confidence markers (regular recall), t (43) = 5.78, $p < .001$, $d = 1$, and information spontaneously recalled with low confidence utterances (uncertainties), t (42) = 7.24, $p < .001$, $d = 1.17$; (2) information recalled with low confidence utterances (uncertainties) was less accurate than information recalled with no confidence markers (regular recall), t (42) = 4.64, $p < .001$, $d = .86$ (see table II). Thus, excluding information which was spontaneously recalled with low confidence utterances (uncertainties) raised report accuracy, t (43) = -5.44, $p < .001$, $d = .74$. In sum, participants were able to use qualitative confidence evaluations to distinguish between three types of information with different accuracy values: uncertainties ($M = .74$, $SD = .19$, 95% CI [.69, .80]), certainties ($M = .95$, $SD = .03$, 95% CI [.94, .96]) and regular recall ($M = .89$, $SD = .07$, 95% CI [.87, .91]). A Goodman and Kruskal's gamma test was then conducted to see whether there was an association between participants' confidence and report accuracy, showing a positive correlation between these two variables (G = .68, $p < .001$).

Insert Table II

3.3 Frequency of certainties and uncertainties

Next, we looked at the absolute frequency (number of newly recalled units of information) and relative frequency (ratio between the number of newly recalled units of information with a given confidence level over all newly recalled units of information) of: (1) information retrospectively evaluated with very high confidence

during the summary phase (certainties); (2) information spontaneously recalled with low confidence utterances (uncertainties); and (3) information recalled with no confidence markers (regular recall). As seen in Table II, participants reported being certain (information retrospectively evaluated with very high confidence) about a large number of recalled information (80.05 out of 118 units of information) and this represents a major fraction of their report (.67). Information recalled with no confidence markers (M = 23.09) and with low confidence utterances (M = 14.86) constituted a smaller recall fraction (.20 and .13 respectively). Only a small number of recalled units of information (M = 4.82) were first spontaneously recalled with low confidence utterances and later retrospectively rated as certainties. These were coded as both low confidence (uncertainties) and very high confidence (certainties) statements. This was infrequent and represents a very small portion of the report (.04).

To understand how information evaluated with very high confidence (certainties) and information recalled with low confidence utterances (uncertainties) was distributed among the several interview phases, two one-way within-participants analysis of variance (ANOVA) were conducted to see if interview phase (phase 2 vs. phase 3 vs. phase 4 vs. phase 5 vs. phase 6) had an effect on (1) uncertainties proportion (ratio of new information recalled with low confidence utterances at an interview phase, over all new units of information recalled at that same interview phase) and (2) certainties proportion (ratio of new information recalled at an interview phase which was later retrospectively rated as certain, over all the new units of information recalled at that same interview phase). Phase 1 (preliminary phase) was not included in this analysis for the reasons stated above. We found no

differences in uncertainties proportion across the different interview phases, F (2.076, 37.367) = 1.89, p = .164, η_p^2 = .10 (see Table III).

Insert Table III

However, we found certainties' proportion varied across interview phases, F (2.060, 35.024) = 4.64, p = .016, η_p^2 = .21. Pairwise comparisons revealed a higher proportion of new information later retrospectively rated by participants as certain was recalled during the free report (phase 2) in comparison with the questioning phase (phase 3), t (43) = 8.59, p < .001, d = 2.31, the second recall attempt (phase 4), t (43) = 8.91, p < .001, d = 2.18, and the summary phase (phase 6), t (26) = 3.21, p = .004, d = 1.22. Information evaluated with very high confidence (certainties) at phase 6 (summary phase) refers to new information participants recalled during the summary phase which was immediately after classified as certain. No differences were found between the other interview phases.

Lastly, overall recall (total number of recalled details) was not associated with the proportion of produced 'uncertainties' (proportion of uncertainties in a given report), r = .20, p = .203, or proportion of produced 'certainties' (proportion of certainties in a given report), r = .13, p = .392.

4. DISCUSSION

This study examined whether witnesses are capable of using qualitative verbal confidence judgements to evaluate which of the information they recalled is more or less likely to be accurate. Our major finding was participants were able to use these confidence judgements to differentiate three types of information that have

different accuracy values: information recalled with low confidence utterances (uncertainties), information recalled with no confidence markers (regular recall) and information retrospectively evaluated with very high confidence (certainties). Information evaluated with very high confidence (certainties) was found to be more accurate than information recalled with no confidence markers (regular recall) and information recalled with low confidence utterances (uncertainties). Additionally, information recalled with low confidence utterances was less reliable than information recalled with no confidence markers. Thus, uncertainties exclusion from the accuracy analysis raised report accuracy (Paulo et al., 2015a).

Even though other methods for assessing confidence (e.g., frequency judgments) do not seem to be valid to estimate report accuracy for an investigative interview (Paulo et al., 2015b), verbal qualitative confidence judgments may allow the interviewer to easily differentiate during the course of an investigative interview between three types of information that may have different accuracy values. This finding is consistent with previous literature that found a positive within- and between- C-A relationship (Luna & Martín-Luengo, 2012; Wixted et al., 2018) and supported by the positive correlation between participants' confidence and report accuracy found in this study. Although previous studies (Allwood et al., 2005; Roberts & Higham, 2002) already found a positive C-A relationship with the use of numerical confidence scales applied at the end of an investigative interview, this study qualitative approach to evaluate confidence may have several practical advantages. Using numerical scales for witnesses to rate their confidence for each unit of information recalled during the interview would take a substantial amount of time and could only be realistically used for limited portions of the report (Paulo et al., 2015a). In the current study, participants were able to monitor their whole report

during the course of the interview in a very natural and time-saving manner which could be easily explained and communicated to jurors or investigators (Kebbell, 2009). For information recalled with low confidence (uncertainties), participants spontaneously used verbal utterances to naturally monitor their report, adding no additional time to the interview. For certainties, participants performed this evaluation retrospectively while their report was being summarized, adding only a few more minutes to the summary phase that is already included in most investigative interviews (Griffiths & Milne, 2010; Milne & Bull, 2003; Paulo et al., 2013). Therefore, differentiating information which is recalled with low confidence utterances and information retrospectively evaluated with very high confidence is a timesaving procedure that could easily be used during real investigative interviews where time constraints are frequent.

It is also important to discuss how information recalled with low confidence utterances (uncertainties) and information retrospectively evaluated with very high confidence (certainties) was distributed across the different interview phases. For instance, it could be problematic if witnesses, when confronted with consecutive retrieval attempts as typically used during investigative interviews, would provide uncertain information during the later interview phases. This is, witnesses could be withholding uncertain information at the beginning of the interview and later choose to reveal it, assuming if the interviewer is asking for successive retrieval attempts he/she expects more information regardless of its accuracy. Our study does not support this as the number of produced uncertainties was similar for all interview phases. Accordingly, the proportion of correct recall, errors, and confabulations is also similar for all interview phases. The CI instruction for participants 'not to guess' might have contributed towards this, as although participants were being

asked to perform additional recall attempts, they were also being instructed not to guess. This might also explain, to some extent, why uncertainties frequency was low and represented a small fraction of participants' total recall. Thus, this study supports using several retrieval attempts as an appropriate interview technique (Fisher & Geiselman, 1992). Furthermore, we found no correlation between report size (total number of units of information recalled) and the proportion of produced uncertainties/ certainties. Therefore, our study does not support participants who recall more information are adopting a more liberal report criterion and consequently recalling a higher proportion of uncertainties or a lower proportion of certainties. As suggested by Paulo et al. (2015a) information recalled with low confidence utterances may be the result of metacognitive monitoring homogeneously and effectively performed throughout the interview regardless of the interview phase.

Regarding information retrospectively evaluated with very high confidence during the summary phase (certainties), participants proportionally recalled more of this type of information during the free report in comparison with the remaining interview phases, except the third recall attempt. Although further studies are necessary, it might be possible participants have a tendency to proportionally recall more information, which will later be rated with high confidence, during the beginning of the interview. This might occur because participants consider higher confidence information to be more valuable for the interviewer and choose to reveal it first or because this information is more accessible and consequently recalled and reported first (Buratti et al., 2014; Koriat & Goldsmith, 1996).

5. LIMITATIONS, CONCLUSION AND PRACTICAL IMPLICATIONS

As with the majority of laboratory mock witness research, the present study contained methodological limitations. Although the interview protocol used in this study was mostly read verbatim (see appendix) to assure consistency across participants, some interview phases had to be adapted to each participant. For instance, the witness-compatible questioning phase and the summary phase needed to be adapted to the participants' previous recall. Although this was necessary to ensure the adequacy of the interviews and increase ecological validity, it causes some degree of variability across participants that we were not able to control. This is a common limitation of most research in the field of investigative interviewing where a witness-compatible questioning or summary phase is used (Fisher, Schreiber Compo, Rivard, & Hirn, 2014). Secondly, the present study used a mock non-emotional video and mainly female students as mock eyewitnesses. Thus, future research using real eyewitnesses and police detectives as interviewers can be important to further test whether these qualitative confidence judgements can be used to evaluate which information is more likely to be reliable during real police investigations.

Despite the limitations mentioned above, this study found witnesses might be able to evaluate which of the information they provided is more or less likely to be accurate using verbal/ qualitative confidence judgements. Moreover, these judgments can be performed during the interview in a timesaving manner that could realistically be used during real investigative interviews unlike other types of confidence judgements that were used in previous studies (Allwood et al., 2005; Roberts & Higham, 2002). This can be useful for police officers, jurors, forensic and legal professionals who do not know what happened during the crime and may want to evaluate which information is more likely to be reliable (Brewer & Weber, 2008; Buratti et al., 2014; Potter & Brewer, 1999). Thus, differentiating information

recalled with spontaneous low confidence utterances (uncertainties) and information evaluated with very high confidence might be a natural and time-saving procedure professionals might want to consider as memory strength indices when analyzing witnesses' reports (Brewer & Weber, 2008). Nonetheless, these measures should not be taken as indisputable accuracy markers.

6. REFERENCES

Allwood, C., Ask, K., & Granhag, P. (2005). The cognitive interview: Effects on the realism in witnesses' confidence in their free recall. *Psychology, Crime & Law, 11*, 183-198. doi:10.1080/10683160512331329943

Bensi, L., Nori, R., Gambetti, E., & Giusberti, F. (2011). The enhanced cognitive interview: A study on the efficacy of shortened variants and single techniques. *European Journal of Cognitive Psychology, 23*, 311-321. doi:10.1080/20445911.2011.497485

Brewer, N., & Weber, N. (2008). Eyewitness confidence and latency: Indices of memory processes not just markers of accuracy. *Applied Cognitive Psychology, 22*, 827-840. doi:10.1002/acp.1486

Brewer, N., Weber, N., Wootton, D., & Lindsay, S. (2012). Identifying the bad guy in a lineup using confidence judgments under deadline pressure. *Psychological Science, 23,* 1208-1214. doi:10.1177/0956797612441217

Brelet, L., Ginet, M., Colomb, C., Jund, R., & Désert, M. (2018). Memory for cars among a female population: Is the cognitive interview beneficial in reducing stereotype threat. *Journal of Police and Criminal Psychology*, 1-12. https://doi.org/10.1007/s11896-018-9261-5

Brown, C. L., & Geiselman, R. E. (1990). Eyewitness testimony of mentally retarded: Effect of the cognitive interview. *Journal of Police and Criminal Psychology, 6*, 14-22. https://doi.org/10.1007/BF02806593

Brunel, M., Py, J., & Launay, C. (2013). Cost and benefit of a new instruction for the cognitive interview: the open depth instruction. *Psychology, Crime & Law, 19*, 845-863. https://doi.org/10.1007/s11896-018-9261-5

Buratti, S., Allwood, M. A., & Johansson, M. (2014). Stability in the metamemory realism of eyewitness confidence judgments. *Cognitive Processing – International Quarterly of Cognitive Science, 15*, 39-53. doi: 10.1007/s10339-013-0576-y

Chapman, A. J., & Perry, D. J. (1995). Applying the cognitive interview procedure to child and adult eyewitnesses of road accidents. *Applied Psychology, 44*, 283-294. doi:10.1111/j.1464-0597.1995.tb01081.x

Colomb, C., & Ginet, M. (2012). The cognitive interview for use with adults: An empirical test of an alternative mnemonic and of a partial protocol. *Applied Cognitive Psychology, 26*, 35–47. doi:10.1002/acp.1792

Colomb, C., Ginet, M., Wright, D., Demarchi, S., & Sadler, C. (2013). Back to the real: Efficacy and perception of a modified cognitive interview in the field. *Applied Cognitive Psychology, 27*, 574-583. doi:10.1002/acp.2942

Davis, M. R., McMahon, M., & Greenwood, K. M. (2005). The efficacy of mnemonic components of the cognitive interview: Towards a shortened variant for time-critical investigations. *Applied Cognitive Psychology, 19*, 75-93. doi:10.1002/acp.1048

Evans, J. R., & Fisher, R. P. (2011). Eyewitness memory: Balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. *Applied Cognitive Psychology, 25*, 501-508. doi:10.1002/acp.1722

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Charles C. Thomas.

Fisher, R. P., Schreiber Compo, N., Rivard, J., & Hirn, D. (2014). Interviewing Witnesses. In T. Perfect & S. Lindsay (Eds.), *The SAGE Handbook of Applied Memory* (pp. 559-578). Los Angeles, USA: SAGE Press.

Geiselman, R. E., & Fisher, R. P. (1988). The cognitive interview: An innovative technique for questioning witnesses of crime. *Journal of Police and Criminal Psychology, 4*, 2-5. https://doi.org/10.1007/BF02806548

Geiselman, R. E., Latts, M. G. (1991). Interviewing survivors of rape. *Journal of Police and Criminal Psychology, 7*, 8-17. https://doi.org/10.1007/BF02806601

Geiselman, R. E., Taras, L. Schaap, R., & Woodruff, N. (1994). Recall of constituent elements from multiple episodes of an event using the cognitive interview. *Journal of Police and Criminal Psychology, 10*, 1-7. https://doi.org/10.1007/BF02803662

Gigerenzer, G. Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506-528. http://dx.doi.org/10.1037/0033-295X.98.4.506

Granhag, P. A., Jonsson, A., & Allwood, C. M. (2004). The cognitive interview and its effect on witnesses' confidence. *Psychology, Crime & Law, 10*, 37-52. doi:10.1080/1068316021000030577

Griffiths, A., & Milne, R. (2010). The application of cognitive interview techniques as part of an investigation. In C. A. Ireland & J. M. Fisher (Eds.), *Consultancy and advising in forensic practice: Empirical and practical guidelines* (pp. 71–90). Chichester, UK: BPS Blackwell.

Higham, P. A., Luna, K., & Bloomfield, J. (2011). Trace-strength and source-monitoring accounts of accuracy and metacognitive resolution in the misinformation paradigm. *Applied Cognitive Psychology, 25*, 324 -335. doi:10.1002/acp.1694

Higham, P. A., & Memon, A. (1999). A review of the cognitive interview. *Psychology, Crime & Law, 5*, 177-196. doi:10.1080/10683169908415000

Kebbell, M. R. (2009). Witness confidence and accuracy: Is a positive relationship maintained for recall under interview conditions? *Journal of Investigative Psychology and Offender Profiling, 6*, 11-23. doi:10.1002/jip.89

Köhnken, G., Milne, R., Memon, A., & Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime & Law, 5*, 3-27. doi:10.1080/10683169908414991

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490-517.

Larsson, A. S., Granhag, P. A., & Spjut, E. (2002). Children's recall and the cognitive interview: do the positive effects hold over time? *Applied Cognitive Psychology, 17*, 203-214. doi:10.1002/acp.863

Liberman, V. (2004). Local and global judgments of confidence. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30,* 729-732. http://dx.doi.org/10.1037/0278-7393.30.3.729

Lindsay, R., Kalmet, N., Leung, J., Bertrand, M., Sauer, J., & Sauerland, M. (2013). Confidence and accuracy of lineups selections and rejections: Postdicting rejection accuracy with confidence. *Journal of Applied Research in Memory and Cognition*, *2*, 179-184. http://dx.doi.org/10.1016/j.jarmac.2013.06.002

Luna, K., & Martín-Luengo, B. (2012). Confidence-accuracy calibration with general knowledge and eyewitness memory cued recall questions. *Applied Cognitive Psychology, 26*, 289-295. doi:10.1002/acp.1822

Manning, C. J., & Loftus, E. F. (1996). Eyewitness testimony and memory distortion. *Japanese Psychological Research, 38*, 5-13. doi:10.1111/j.14685884.1996.tb00003.x

Memon, A., Wark, L., Bull, R., & Köhnken, G. (2011). Isolating the effects of the cognitive interview techniques. *British Journal of Psychology*, *88*, 179-197. doi:10.1111/j.2044-8295.1997.tb02629.x

Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing*. Cambridge, MA: The MIT Press.

Milne, R., & Bull, R. (2003). *Does the cognitive interview help children to resist the effects of suggestive questioning? Legal and Criminological Psychology, 8*, 21-38. doi:10.1348/135532503762871219

Paulo, R. M., Albuquerque, P. B., & Bull, R. (2013). The enhanced cognitive interview: Towards a better use and understanding of this procedure. *International Journal of Police Science & Management, 15,* 190-199. http://dx.doi.org/10.1350/ijps.2013.15.3.311

Paulo, R. M., Albuquerque, P. B., & Bull, R. (2015a). The Enhanced Cognitive Interview: Expressions of uncertainty, motivation and its relation with report

accuracy. *Psychology, Crime & Law, 22,* 366-381. doi: 10.1080/1068316X.2015.1109089

Paulo, R. M., Albuquerque, P. B., Saraiva, M., & Bull, R. (2015b). The enhanced cognitive interview: Testing appropriateness perception, memory capacity and error estimate relation with report quality. *Applied Cognitive Psychology, 29,* 536-543. doi: 10.1002/acp.3132

Paulo, R. M., Albuquerque, P.B., & Bull, R. (2016). Improving the enhanced cognitive interview with a new interview strategy: Category clustering recall. *Applied Cognitive Psychology, 30*, 775-784. doi:10.1002/acp.3253

Paulo, R. M., Albuquerque, P. B., Vitorino, F., & Bull, R. (2017). Enhancing the cognitive interview with an alternative procedure to witness-compatible questioning: Category clustering recall. *Psychology, Crime & Law, 23,* 967-982. https://doi.org/10.1080/1068316X.2017.1351966

Perfect, T. J. (2002). When does eyewitness confidence predict performance? In T. J. Perfect, & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 95–120). Cambridge: Cambridge University Press.

Potter, R., & Brewer, N. (1999). Perceptions of witness behaviour-accuracy relationships held by police, lawyers and mock-jurors. *Psychiatry, Psychology and Law, 6*, 97–103. doi:10.1080/13218719909524952

Prescott, K., Milne, R., Clark, J. (2011). How effective is the enhanced cognitive interview when aiding recall retrieval of older adults including memory for conversation? *Journal of Investigative Psychology and Offender Profiling, 8*, 257-270. doi:10.1002/jip.142

Riccó, A. (Director), & Riccó, R. (Director). (2004). O Assalto [The robbery] [Television series episode]. In V. Castelo (Producer), *Inspector Max*. Lisbon: Produções Fictícias.

Rivard, J. R., Fisher, R. P., Robertson, B., & Mueller, D. H. (2014). Testing the cognitive interview with professional interviewers: Enhancing recall of specific details of recurring events. *Applied Cognitive Psychology, 28,* 917-925. doi:10.1002/acp.3026

Roberts, W.T., & Higham, P.A. (2002). Selecting accurate statements from the cognitive interview using confidence ratings. *Journal of Experimental Psychology: Applied, 8*, 33-43. doi:10.1037/1076-898X.8.1.33

Smith, V. L., Kassin, S. M., & Ellsworth, P. C. (1989). Eyewitness accuracy and confidence: Within- versus between-subjects correlations. *Journal of Applied Psychology, 74*, 356–359.

Sniezek, J. A. & Buckley, T. (1991). Confidence depends on level of aggregation. *Journal of Behavioral Decision Making, 4,* 263-272. doi:10.1002/bdm.3960040404

Stein, L. M., & Memon, A. (2006). Testing the efficacy of the cognitive interview in a developing country. *Applied Cognitive Psychology, 20*, 597–605. doi:10.1002/acp.1211

Verkampt, F., & Ginet, M. (2010). Variations of the cognitive interview: which one is the most effective in enhancing children's testimonies? *Applied Cognitive Psychology, 24*, 1279–1296. doi:10.1002/acp.1631

Wixted, J. T., Mickes, L., & Fisher, R. P. (2018). Rethinking the reliability of eyewitness memory. *Perspectives on Psychological Science, 13* (3), 324-335. doi: 10.1177/1745691617734878.

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18(1)*, 10-65. https://doi.org/10.1177/1529100616686966

Wright, A., & Holliday, R. (2006). Enhancing the recall of young, young–old and old–old adults with cognitive interviews. *Applied Cognitive Psychology, 21*, 19–43. doi:10.1002/acp.1260

Table I

Mean and standard deviation for the number of newly recalled units of information (recall), proportion of correct information (correct), proportion of errors (error) and proportion of confabulations (confabulation), according to interview phase.

.

	Recall		**Correct**		**Error**		**Confabulation**	
	(New unit)		(Proportion)		(Proportion)		(Proportion)	
Interview Phase	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Free report	38	16	.96	.04	.04	.04	.01	.01
Questioning	43	13	.89	.06	.09	.06	.02	.03
2nd recall – Category Clustering	24	7	.90	.07	.09	.06	.01	.03
3rd recall – Reverse Order	2	3	.82	.31	.09	.20	.09	.25
Summary	1	2	.92	.17	.06	.15	.02	.10

Table II

Mean and standard deviation for the absolute frequency, relative frequency and correct recall proportion (accuracy) for information evaluated with very high confidence (certainties), information recalled with low confidence utterances (uncertainties) and information recalled with no confidence markers (regular recall).

	Absolute Frequency		**Relative Frequency**		**Accuracy**	
Confidence evaluation	***M***	***SD***	***M***	***SD***	***M***	***SD***
Certainties (very high confidence)	80.05	15.96	.67	.09	.95	.03
Regular recall (no confidence marker)	23.09	10.42	.20	.07	.89	.07
Uncertainties (low confidence)	14.86	6.87	.13	.07	.74	.19

Table III

Mean and standard deviation for the proportion of information spontaneously recalled with low confidence utterances (uncertainties) and proportion of information retrospectively evaluated with very high confidence (certainties) according to interview phase.

	Uncertainties (low confidence)		**Certainties (very high confidence)**	
Interview phase	***M***	***SD***	***M***	***SD***
Phase 2 - Free report	.05	.06	.84	.07
Phase 3 - Questioning	.17	.11	.61	.12
Phase 4 – 2nd recall (Category Clustering)	.10	.09	.56	.17
Phase 5 – 3rd recall (Reverse Order)	.16	.27	.63	.33
Phase 6 - Summary	.22	.38	.49	.40