

Running head: ELICITING COARSE-GRAIN INFORMATION

Interviewing Witnesses: Eliciting Coarse-Grain Information

Neil Brewer, Ambika Nagesh Vagadia

Flinders University

Lorraine Hope

University of Portsmouth

Fiona Gabbert

Goldsmiths, University of London

**Author Copy**

**Revision accepted for publication in Law and Human Behavior (May, 2018)**

This research was supported by Australian Research Council Linkage International Grant LX0989922 and Economic and Social Research Council Bilateral Programme Grant RES-000-22-3169. We are grateful to Nicole McCallum for assistance with data collection, coding and analyses, and to Carmen Lucas, Alessandra Caso and Serena Ventura for inter-rater reliability reliability analysis and coding.

Correspondence concerning this article should be addressed to Neil Brewer, School of Psychology, Flinders University, GPO Box 2100, Adelaide, South Australia, Australia. E-mail:

[neil.brewer@flinders.edu.au](mailto:neil.brewer@flinders.edu.au)

## Abstract

Eyewitnesses to crimes sometimes report inaccurate fine-grain details but fail to report accessible and potentially accurate coarse-grain details. We asked college students and community members (aged 17 to 62 years) who viewed a video of a simulated crime to answer interviewers' questions at coarse- and fine-grained levels of detail and measured the quantity and accuracy of their responses. Three experiments (overall  $N = 219$ ) also (a) provided comparative data for participants who were interviewed using the open-ended Self-Administered Interview (Gabbert, Hope, & Fisher, 2009) or one of two "report everything" open-ended procedures, (b) tested the efficacy of the procedure using both written and verbal interviews, and (c) examined the generality of the findings across different encoding stimuli which required variations in the types of cued recall questions asked. Coarse-grain reporting seldom occurred under the free recall interview conditions. Witnesses provided abundant coarse-grain details when required to respond to probes about specific details (i.e., cued recall forced report conditions) – without obvious cost to overall accuracy relative to accuracy of similar detail reported under free recall conditions – regardless of whether they responded on a written questionnaire or in a face-to-face individual interview. These experiments suggest that a procedure that requires cued recall forced reporting of coarse-grain detail may offer potential in certain investigative situations as an adjunct to the widely recommended open-ended forensic interviewing techniques.

**Public Significance Statement.** Witnesses to crimes sometimes volunteer inaccurate fine-grain detail (e.g., purple getaway car) rather than accurate coarse-grain or broad detail (e.g., dark-colored getaway car) despite the latter being potentially valuable in some investigative contexts (e.g., review of CCTV footage). Requiring witnesses to give coarse-grain or general answers to interviewers' questions produced a substantial amount of accurate information that was not volunteered in open-ended interviews, suggesting a possible way of supplementing currently recommended open-interviewing techniques.

**Key Words:** eyewitness memory, investigative interviewing, metamemory, cued recall

### Interviewing Witnesses: Eliciting Coarse-Grain Information

A key objective for forensic interviewers is to maximize the amount and quality of the information obtained from eyewitnesses to crimes and other events (Vrij, Hope, & Fisher, 2014). The “quality” of information (i.e., its contribution to advancing an investigation and incriminating or exonerating a suspect) is obviously related to its accuracy. But quality is also likely to be determined by what the information adds to the overall body of evidence, with what may seem like a trivial piece of information in one evidential context proving to be invaluable in another. Not surprisingly, a major focus in forensic interviewing research has been the investigation of interviewing approaches designed to increase the quantity and quality of witness reports. This emphasis contributed to the development of interviewing techniques such as the Cognitive Interview (Fisher & Geiselman, 1992) and the Self-Administered Interview (SAI<sup>®</sup>, Gabbert, Hope, & Fisher, 2009), techniques based on strong advocacy for the use of open questions that allow for an unlimited free-recall response from an interviewee (e.g., *Tell me in your own words what happened.*) and yield higher accuracy than closed questions (Fisher, Falkner, Trevisan, & McCauley, 2000). Yet, witnesses’ reports are often somewhat less detailed or complete than necessary to pursue effective investigations (Gabbert, Hope, Carter, Boon, & Fisher, 2015). Consequently, interviewers typically need to probe for further detail using further open-ended prompts and targeted questioning (Lamb, Hershkowitz, Orbach, & Esplin, 2008; Powell, Fisher, & Wright, 2005).

One pattern that has emerged from recent research is that witnesses may fail to report a coarse-grain detail (e.g., an imprecise or broad detail, such as the robber was driving a light-colored car) that is both accessible and quite likely to be accurate, opting instead to report a fine-grain detail (e.g., the robber was driving a pale blue-colored car) that may be inaccurate (e.g., McCallum, Brewer, & Weber, 2016). It is obvious that fine-grain detail can help shape the direction of a police investigation and, when provided in courtroom testimony, enhance the credibility of the witness’s testimony (Bell & Loftus, 1988, 1989). It is less obvious why

coarse-grain detail may be valuable. Although a coarse-grain detail, when considered in isolation, may seem so imprecise as to be non-helpful for investigating police officers, it could conceivably assist an investigation. For example, if the witness reported that the robber was driving a light-colored car, it might help the police to narrow their pool of suspects (perhaps discarding a suspect who always drove a dark blue car) or it might speak to the reliability of other witness statements that mentioned a white or beige car, thereby increasing the investigator's confidence in the direction the investigation should pursue. More pragmatically, obtaining coarse-grain information of this nature could directly inform the allocation of resources in real time (e.g., facilitate selective review of CCTV footage to identify a light-colored getaway car in the vicinity). Similarly, if the witness described the thief as being shortish, perhaps as short as 160 cm but certainly no taller than 175 cm, it may persuade the police to look for an alternative suspect to the 185 cm individual they had initially considered. [Just a few examples of other items that might manifest in coarse- or fine-grain reports are hair length, hair color, clothing color, the duration of some part of the event, the number of people involved, the positioning of individuals involved, and so on.] In isolation, the coarse-grain version of such items may not be compelling or incriminating. But even seemingly trivial coarse-grain information may be informative when combined with other evidence that emerges during an investigation.

Yet, despite evidence indicating that individuals are able to regulate their reporting of fine- and coarse-grain details in order to adaptively control output quantity and accuracy (e.g., Goldsmith, Koriat, & Weinberg-Eliezer, 2002; Weber & Brewer, 2008), coarse-grain details may be reported less frequently than is optimal from the perspective of maximizing the amount of accurate and forensically useful information reported. For example, in two separate experiments McCallum et al. (2016) reported that, when witnesses were required to choose between a fine- and a coarse-grain response to an interviewer's question, an incorrect fine-grain report was preferred to a correct coarse-grain report on 13-15% of occasions. McCallum

et al. (2016) examined the effect of various socially motivating conditions on reporting of fine- and coarse-grain information, with a signal detection analysis highlighting a response bias towards reporting fine- rather than coarse-grain information. They also found that this response bias was more marked when the experimental conditions were such that any perceived consequences associated with reporting were minimized. Specifically, the tendency to avoid coarse-grain reporting that could have optimized accuracy was less when social (i.e., an evaluative audience) and financial contingencies (i.e., a penalty for inaccuracy) were in place.

In a different domain of memory reporting (i.e., answering general knowledge questions), Ackerman and Goldsmith (2008) demonstrated that people sometimes failed to provide coarse-grain answers to questions, despite having high confidence in their prospective answer and the answer being accurate. They hypothesized that this response pattern reflected the individual's belief that their answer was not sufficiently informative given the social communication context in which they were operating (cf. Grice, 2002). Perhaps witnesses being interviewed about a crime might be extremely confident about some coarse grain details they recalled, but at the same time believe that the police would find such details uninformative (e.g., because they are too general or assumed to be already known to the police.) Thus, they may respond "don't know" (e.g., Scoboria & Fisico, 2013) or simply fail to volunteer the item.

Given these uncertainties about coarse-grain reporting, the focus of our studies was to evaluate what – given the current Zeitgeist in the forensic interviewing field – might at face value seem to be a counterintuitive approach, yet one that could (a) be used to increase witnesses' reporting of coarse-grain details, (b) provide an indication of the likely accuracy of those details, and (c) complement contemporary best practice approaches to interviewing witnesses. In the forensic interviewing field open-ended questioning represents the *gold standard* approach (Fisher, Milne, Bull, 2011). Closed questioning techniques have been criticized as potentially suggestive or leading (Memon, Wark, Holley, Koehnken, & Bull,

1996). Systematic and exhaustive free recall prompting should be the interviewer's initial focus. Directive or "wh" prompts or questions about already mentioned details might follow if further information is required and, subsequently, yes/no or forced choice questions might address new issues – but there is a consensus that the latter approaches increase the risk of eliciting inaccurate information (see, for example, Lamb et al., 2008; Powell et al., 2005). Nevertheless, unbiased cued questioning may provide direct, target-specific cues and information to facilitate retrieval (cf. Tulving & Thomson, 1973) and elicit accurate and detailed information (e.g., Pipe, Gee, Wilson, & Egerton, 1999).

Imagine, however, a situation where witnesses are asked to answer all cued recall questions (i.e., *don't know* was not an option): that is, the interview procedure involves *forced report*. Given that witnesses often provide quite impoverished accounts of witnessed events (Douglass, Brewer, Semmler, Bustamante, & Hiley, 2013; Gabbert et al., 2015), exhorting witnesses to report an answer to every question will obviously markedly increase the quantity of outputs, although the accuracy of some of the outputs that, but for forced reporting, would have been withheld will be questionable. This is especially likely to be the case for fine-grain responses. As noted above, however, cued recall questions can act as valuable retrieval cues. Thus, asking witnesses to answer all such questions is likely to lead to reporting of some accurate details which, under free recall conditions, they may judge as insufficiently informative to report. In line with Ackerman and Goldsmith (2008) and Yaniv and Foster (1995), it seems likely that coarse-grain information is the sort of information that witnesses are most likely to judge as unimportant and, in turn, withhold. Accordingly, we expected forced reporting to have a disproportionate effect on coarse- (cf. fine-) grain reporting. Moreover, given coarse-grain details are, on average, more likely to be accurate than fine-grain (see, for example, accuracy data for both numeric and verbal eyewitness reports in Weber & Brewer, 2008), we argue that forced reporting may be an effective, albeit unusual, way of

eliciting forensically useful information from witnesses, information that might never be elicited under open questioning.

We focused on the impact of one forced report procedure on the quantity of accurate coarse-grain details reported and their overall accuracy, guided by confidence-accuracy profiles for reported items. This procedure – referred to subsequently as the grain size procedure – required witnesses to report answers to interviewers' questions at both a coarse- and fine-grain level of detail, each of which was accompanied by a judgment of confidence in the accuracy of the response. Our other experimental conditions (a) provided comparative data for participants who were interviewed immediately after the encoding event using the Self-Administered Interview (Gabbert et al., 2009), as well as for participants using two other open-ended procedures, (b) tested the efficacy of the grain size procedure using both written and verbal interviews (both approaches are used in the field), and (c) examined the generality of the findings across different encoding stimuli which required variations in the types of cued recall questions asked.

### **Discriminating Accurate From Inaccurate Responses**

As forced reporting will give rise to a substantial number of inaccurate fine-, and perhaps also coarse-, grain details, investigators would need a way of discriminating accurate from inaccurate reports with reasonable reliability. A possible solution is suggested by the literature indicating that confidence provides useful information about accuracy of eyewitness reports when witnesses are asked to recall details of people and events observed. For example, some studies show that, at the group level, confidence for accurate responses is higher than for inaccurate responses (Luna & Martín-Luengo, 2012; Robinson & Johnson, 1996). Other studies have found that witnesses express higher confidence for those items they have reported accurately than for those that are incorrect (Luna & Martín-Luengo, 2012; Roberts & Higham, 2002). There is also evidence that confidence and accuracy are calibrated, with increases in confidence associated with progressive increases in accuracy (Luna, Higham, & Martín-

Luengo, 2011; Luna & Martín-Luengo, 2012). There are, however, some qualifications. Any overconfidence in memory reports means there will be imprecision if using confidence to diagnose the accuracy of any specific item of information. Specifically, probability of a correct response for items reported with extremely high confidence levels (e.g., 90-100%) will be lower than suggested by the confidence rating. Moreover, the degree of overconfidence typically increases with task difficulty (see Gigerenzer, Hoffrage, & Kleinbölting, 1991; Weber & Brewer, 2004), with lower memory strength likely to be associated with greater overconfidence. Note, however, that the literature also indicates that overconfidence for coarse-grain, forced eyewitness reports will be significantly less than for fine-grain forced reports (Weber & Brewer, 2008).

Here we used the profiles of item-level confidence judgments that best discriminated accurate from inaccurate responses to explore the extent to which the output gains produced by forced reporting could be achieved without significant accuracy decrements when compared with similar outputs from open-ended interviewing techniques. If confidence is to assist an interviewer to assess the likely accuracy of individual items, it would be useful to have a relatively simple cutoff or decision rule. One such rule would be an absolute confidence cutoff that empirical research has shown provides desirable accuracy levels (e.g., the likelihood that an item was accurate = 90-100%) across an array of forensically relevant conditions. We, therefore, examined accuracy levels for responses reported with 100% confidence,  $\geq 90\%$  confidence,  $\geq 80\%$  confidence, etc.

### **The General Experimental Approach**

In three experiments witnesses reported answers to every cued recall question at both a coarse- and a fine-grain level of detail, with each answer accompanied by a judgment of confidence in the accuracy of the response. Subsequent to these reports they provided their preferred (i.e., coarse- or fine-grain) report, again with an accompanying confidence judgment. Although our focus was on improving coarse-grain reporting, we obtained the fine-grain and



preferred reports from witnesses for two reasons. First, we considered the possibility that the proven effectiveness of this paradigm in generating both coarse- and fine-grain reports (cf. McCallum et al., 2016; Weber & Brewer, 2008) might hinge, to some degree, on witnesses producing these reports one immediately after the other. In other words, perhaps the effectiveness of the procedure is dependent on witnesses being able to make immediate contrasts between the two response options. Second, it provided an opportunity to examine how effectively item-level confidence judgments might discriminate accurate from inaccurate responses in the context of forced fine-grain reporting.

A fundamental issue we faced was how to assess whether forced reporting offers potential for improving coarse-grain outputs. Although the obvious answer may be to compare coarse-grain outputs obtained via forced report with those obtained using open-ended questioning techniques, such comparisons are difficult to interpret (see, for example, Fisher, Schreiber Compo, Rivard, & Hirn, 2014). The amount of information produced by cued recall questions depends on the number of questions asked, and that number is somewhat arbitrary. As Fisher et al. noted, if an interviewee is asked a large number of cued recall questions, s(he) most likely will report more items than someone asked, under an open-ended approach such as free report, to tell everything they can recall. In contrast, if asked only one cued recall question it is very likely that less will be reported than under free report. In a similar vein, accuracy will also vary depending on the questioning, with lots of cued recall questions about very fine details (i.e., very difficult questions) probably undermining accuracy relative to free report much more than a few questions about very prominent details (i.e., easy questions).

There is no neat or perfect way of resolving this dilemma. One way of developing cued recall (or forced report) questions would be to obtain a free report from each individual participant and follow up each item reported with targeted cued questions. In other words, every participant could conceivably receive a different battery of cued questions. Another possibility would be to solicit free reports from a number of individuals and follow up with

cued questions about the free report items more commonly reported by the sample. Pilot testing using such an approach quickly revealed, however, that few coarse-grain items were volunteered in free report.

Consequently, we adopted the following approach, despite its limitations. We constructed an array of questions that could be asked about the appearance, clothing and actions of the people involved (i.e., perpetrators and bystanders or potential witnesses) and the crime location. We chose this approach because, if the forced report procedure showed promise as a supplement to information gained via open-ended approaches, it would be possible in practice to quickly gather initial background information about any particular case – for example, how many offenders (which could lead to specific person, role, clothing and action questions); how many witnesses (leading to person, role and location descriptions) – and then frame a set of relatively standardized cued coarse-grain questions targeting such characteristics. However, given (a) interviewees exposed to open-ended questioning may be unlikely to ever report some of the details tapped by forced report questions, and (b) our desire to make some kind of meaningful comparative evaluation of the information obtained from different interviewing approaches, we excluded any items obtained under forced report that no one in any of the open-ended conditions reported from any comparisons across interview conditions. (See the Method section for more detailed discussion and the Discussion section for the limitations associated with this approach.)

The approach just outlined does not provide a magical solution to the problems associated with trying to make definitive output quantity and accuracy comparisons across interview conditions. But the comparisons should at least be *suggestive of the potential* of forced reporting of coarse-grain information for supplementing open-ended reports. Thus, in all experiments participants were randomly assigned to either the grain size procedure condition or to one of several different open-ended interviewing approaches, thereby providing some crude benchmarks for the extent of coarse-grain reporting under different reporting conditions.

We describe the differences between Experiments 1, 2 and 3 in the next section and then present the Method and Results for these experiments in a single Method and Results sections.

### **Experiment 1**

We required adult witnesses to a simulated crime to provide written answers to a series of cued recall interview questions with both a coarse- and fine-grain response (hereafter referred to as forced report: initial), and a confidence assessment for each response, followed by reporting of their preferred response (hereafter, forced report: preferred). We used a written interview format in Experiment 1 (and 3) for two practical reasons. First, there are two published studies demonstrating that the pencil-and-paper format has been used effectively to examine eyewitness reporting (e.g., McCallum et al., 2016; Weber & Brewer, 2008). Second, it provided an opportunity to conduct a preliminary and economical investigation of our hypotheses. Second, investigators sometimes use written protocols to elicit information initially from witnesses (e.g., SAI<sup>®</sup>, Gabbert et al., 2009; Hope, Gabbert, & Fisher, 2011). Our focus was on the quantity of coarse-grain accurate outputs, overall coarse-grain accuracy levels, and how the proportion of accurate coarse-grain responses varied in association with item-level confidence judgments.

Coarse-grain outputs from this forced report procedure were contrasted with outputs under an established open-ended interview procedure, an adapted version of the SAI<sup>®</sup> (Gabbert et al., 2009), as well as against two other “report everything” free recall conditions. The SAI<sup>®</sup> (Gabbert et al., 2009) is a protocol of instructions and questions drawing on the Cognitive Interview technique and designed to elicit a written free-recall statement from witnesses, followed by a number of open-ended prompts (see later section on Interview Protocols for detailed information). Like the Cognitive interview, the SAI<sup>®</sup> has been shown to outperform standard free recall by increasing output quantity without compromising accuracy (Gabbert et al., 2009). Thus, the SAI<sup>®</sup> provides a contrast of coarse-grain reporting under forced report with that elicited by a procedure used in investigative interviewing and considered likely to

optimize interviewee performance. The two other free recall conditions were: (i) free recall (similar to Gabbert et al.'s (2009) free recall condition) and (ii) free recall plus confidence. Unlike the SAI<sup>®</sup> these two conditions did not include systematic follow-up prompting regarding the people and events and might, therefore, be considered to be low fidelity open interview formats. In these two conditions witnesses provided a written report, with instructions to report line-by-line as many items as they could remember. In the free recall plus confidence condition, witnesses provided a confidence assessment for each item of information reported. This condition was included to check the possibility that reporting confidence item-by-item might sensitize witnesses to memory strength for each item and affect outputs, perhaps by influencing the criterion for reporting. The three free recall conditions were included to allow for the possibility that coarse-grain reporting might be sensitive to one or other approach, but we made no predictions regarding likely performance differences between these conditions. Figure 1 summarizes the key features of the different interview conditions used in Experiment 1 (and subsequent experiments).

Forced reporting of coarse-grain details obviously will increase output quantity compared with other interviewing approaches. Some decline in overall accuracy is also to be expected although the total number of accurate items reported should increase. Moreover, given the expected relationship between confidence and accuracy of reported items, we anticipated high accuracy for coarse-grain items reported with high confidence. Similar patterns were expected for fine-grain details although the accuracy decrement for fine-grain reports should be greater and, given that overconfidence is generally more marked when task difficulty is higher (Gigerenzer et al., 1991; Weber & Brewer, 2004), high confidence is unlikely to be as reliable an index of accuracy as it is for coarse-grain items.

## **Experiment 2**

We conducted all interviews individually using verbal rather than written reports, reduced the number of cued recall interview questions, and used only three interview techniques: the verbal grain size procedure, the adapted SAI<sup>®</sup>, and the free recall plus confidence interview).

### **Experiment 3**

We examined the generalizability of the findings using a different stimulus event. Use of the grain size procedure is obviously restricted to cued recall questions for which both fine- and coarse-grained responses can be generated. The grain size procedure used in Experiments 1 and 2 not only effectively encompassed numeric and color questions but also demonstrated that grain size distinctions can be elicited for characteristics such as hair length (e.g., fine-grain: crew cut, past shoulders; coarse-grain: short, long) and locations (e.g., fine-grain: at the counter, behind the back desk; coarse-grain: back of bank, front of bank). Thus, questions in the bank robbery grain size procedure could be categorized as (1) color questions relating to characters' clothing, features, and accessories, (2) color questions relating to objects, (3) numeric questions, (4) distance or length questions, and (5) location questions.

However, recall that in Experiments 1 and 2 we excluded responses to grain-size procedure questions from comparisons between conditions if they were never reported under open-ended questioning conditions. Those grain size details that were unreported in the open-ended conditions were predominantly details regarding the color of characters' accessories, the color of less central characters' clothing, and numeric details which involved making quantitative grain size distinctions on qualitative actions (e.g., responses to grain size questions about the length of time it took robbers to complete a certain action or how often characters spoke to other characters, perhaps better reported in open-ended techniques as qualitative descriptions of the sequence of events). The relevance (or probative value) of details such as these will obviously vary depending on the characteristics of the stimulus event (and, of course, the other evidence available to investigators). To examine the generality of our findings thus far, Experiment 3 replicated Experiments 1 and 2 using a different stimulus event.

Compared to the stimulus event used in Experiments 1 and 2, the new stimulus event had fewer characters and more sequences of actions and conversations, which reduced the number of details amenable to grain size classification. There were fewer questions dealing with color details about characters (27% vs. 51%) but more seeking numeric details about qualitative actions (51% vs. 32%). We were particularly interested in whether the grain size procedure's coarse-grain output quantity advantages evident in Experiments 1 and 2 would generalize to another stimulus event that allowed fewer grain size questions and the possibility of more grain size questions being excluded from comparisons due to the relevant detail not being reported under the open-ended techniques.

## Method

### Participants and Design

**Experiment 1.** Participants were 98 adults (50 male) recruited from an undergraduate population and from the general community via convenience sampling. Their ages ranged from 17 to 52 years ( $M = 26.8$ ,  $SD = 8.0$ ). Participants were randomly allocated to one of four written interview conditions: (1) grain size procedure ( $N = 24$ ), (2) the adapted SAI<sup>®</sup>, with confidence ( $N = 25$ ), (3) free recall ( $N = 25$ ), and (4) free recall plus confidence ( $N = 24$ ). The design was a 4 (interview procedure: grain size, adapted SAI<sup>®</sup>, free recall, free recall plus confidence)  $\times$  2 (interview session: immediate, follow-up) mixed design, with interview session as a within-subjects factor. (Exactly the same interview was conducted with all Experiment 1 participants after a delay of 3-5 months. As this second interview could not affect responses at the initial interview (the focus of the study), these data are not reported or discussed.)

**Experiment 2.** Participants were 60 adults (27 male) recruited from an undergraduate population and paid an honorarium. Their ages ranged from 17 to 62 years ( $M = 25.1$ ,  $SD = 9.6$ ). They were randomly assigned one of three verbal interview conditions: (1) grain size procedure ( $N = 20$ ), (2) the adapted SAI<sup>®</sup> with confidence ( $N = 20$ ), and (3) free recall-

confidence ( $N = 20$ ).

**Experiment 3.** Participants were 61 adults (31 male) recruited from an undergraduate population and the general community. Their ages ranged from 18 to 56 years ( $M = 26.1$ ,  $SD = 7.5$ ) and they were randomly allocated to one of the four written interview conditions used in Experiments 1.

Participants were predominantly Caucasian and spoke English as their first language. They were recruited from registrants on the online SONA system and via Facebook advertisement. The studies were approved by the Social and Behavioral Science Ethics Committee of Flinders University.

### **Materials**

The stimulus event for Experiments 1 and 2 was a 36-second video of a bank robbery. It showed two robbers in a bank, one at the counter in front of a female bank teller and the other near the door standing guard. Both were males and of Caucasian appearance. The teller put money in a bag as the main robber said *hurry up*. A customer stood near the counter covering her eyes, another female teller was watching from the counter and a male teller watched from the back of the bank. The teller said *that's all* and pushed the bag towards the main robber. Both robbers ran from the bank onto a road. An accomplice picked them up in a getaway car and drove off.

The stimulus event for Experiment 3 was a two minute video of a non-violent credit card theft at a restaurant. A male thief of Malaysian appearance walked from the street into a restaurant and approached a waiter (Caucasian) standing behind the cash register in the bar. At the same time a customer from the restaurant approached the waiter, asked for the bill and left his credit card on a plate on the bar top. The waiter took the plate and placed it on the desk in front of him. The thief distracted the waiter with a fake dinner reservation request and made a first attempt to take the credit card when the waiter's back was turned. After requesting another date, the thief stole the credit card and waited as the waiter answered a phone call. The

thief then left the restaurant and broke into a run once outside. The waiter eventually realized that the credit card had been stolen, went outside to look for the thief, and returned to the restaurant after there was no sign of him. The stimulus films used in Experiments 1-3 are of much shorter duration and qualitatively different to that used in Gabbert et al. (2009). Thus, although the SAI<sup>®</sup> and the other free recall condition manipulations are virtually identical to Gabbert et al.'s, reporting of fewer items is to be expected in the current experiments.

### **Interview Protocols**

All interviews for each experiment were conducted by the second author, a PhD (clinical) psychology trainee, who was not blind to condition. Written and verbal interviews for all conditions were completely scripted, with the scripts including precise instructions on how to answer queries, permissible forms of feedback (e.g., no information on accuracy or informativeness of responses) and clarification questions (e.g., what to do in order to obtain a confidence estimate for each item if multiple items were reported in one sentence). Verbal interviews were recorded for subsequent coding and reliability assessments.

**Experiment 1.** The written grain size procedure (see Supplementary Materials pp.1-5) included 64 cued recall questions, each requiring a coarse-grain (and a fine-grain) response. Questions focused on details about characters, the crime event, objects, and surroundings, eliciting grain size distinctions for colors, numbers, descriptions of length, and descriptions of locations. Questions were presented to participants in typewritten booklets with the questions and fine- and coarse-grain response lines on the left side of each page and corresponding confidence scales (0%-100%) on the right. The booklet contained (1) an explanation (with example) of the distinction between fine- and coarse-grain responses, (2) two examples of questions, responses, and confidence ratings, with each example indicating both a correct fine- and a coarse-grain response, and (3) written instructions including 'please answer ALL the following questions about the video you just saw and for each answer please circle how sure you are that the response you just wrote is correct.' In one example participants were asked



*What color were the walls in the bank?* and shown two possible responses, *Exact color – white* and *General Color (light, medium, dark) – light*. In the second example they were asked *How many people were in the bank during the robbery?* And shown the responses *Between (range) – 4 and 8* and *Exact – 6*.

In the first phase of the grain-size procedure, participants were instructed to provide (unlike the examples, unprompted by a specific cue) a coarse- or fine-grain response (order counterbalanced across questions) followed by circling their level of confidence, and then the other grain size response (i.e., fine or coarse) followed by a second confidence judgement. In Phase 2, participants were given their responses with the associated confidence levels removed and were asked to select, by circling or ticking, one response as their preferred answer for each question.

The adapted version of the SAI<sup>®</sup> used for the present experiments incorporated the context reinstatement and report everything components of the Cognitive Interview and comprised five sections<sup>1</sup>. Instructions throughout requested witnesses to provide the most complete and accurate account possible, even when details may seem trivial or incomplete. Prior to writing down details of their memories interviewees were asked to spend around three minutes picturing in their minds the physical context (e.g., where they were, what happened, who was involved) and the personal context (what they were thinking and how they were feeling) at the time they witnessed the event. They were instructed to close their eyes and concentrate on recreating the context in their mind. The reporting booklet was divided into sections. The first required a completely open-ended report about the event and the people, with interviewees asked to record every detail they could about the event and the people involved, no matter how trivial or incomplete the information might seem. Each piece of information was reported on a new line together with a rating of how sure they were that each response was correct on a scale of 0% (very very unsure) to 100% (very very sure). The second section focused on the perpetrators (even if they had already provide details about them) and provided an array of 12

prompts covering details such as ethnic appearance, complexion, age, gender, facial hair, build clothing, distinguishing marks and so on. Again each detail reported was accompanied by a confidence rating. Participants were provided with fine and coarse-grain examples, though they were not labeled thus. The third section prompted reporting about the sequence of actions and events relating to the crime (e.g., about any vehicles that may have been present and involved in the crime). Again each detail reported was accompanied by a confidence rating and fine and coarse-grain example were provided. The fourth section prompted (with an example) reporting about any other people who were present and may have been potential witnesses. The final section (with an example) prompted for any other information they had not already been asked about or could remember. In sum, this adapted version of the SAI<sup>®</sup> combined free recall with various open-ended cues or prompts and recorded confidence for each reported detail.

The free recall interview was presented in a typewritten booklet (see Supplementary Materials p.6) with the single free-recall instruction to “write down everything you can remember about the video you just saw” in the 10-page booklet provided. Additional opening instructions and examples of responses were used to be consistent with the other interview techniques, with the exception of explaining grain size. Participants were instructed to use as many pages in the booklet as required. The free recall plus confidence interview (see Supplementary Materials pp.7) was presented in the same way as the free recall interview, with additional instructions for participants to write one response detail per line and indicate their confidence in the veracity of that particular detail on the confidence scale underneath the line.

**Experiment 2.** The individual verbal interview procedure was the same as for Experiment 1, with the following three exceptions. First, participants were told: *I have a set of instructions about what questions I ask you, in what order I ask them, and how I ask them. Sometimes I might interrupt you. That doesn't mean you're wrong, it just means that I have to ask something according to my instructions and in a particular order. Does that make sense? Just*

*give it your best shot. Do you have any questions?* These instructions were included in an attempt to minimize the impact of interruption and repetition on reporting accuracy (Garven et al., 1998; Garven et al., 2000). Second, the number of questions in the grain-size procedure was reduced from 64 to 43. Questions that had not been spontaneously addressed by at least one open-ended interview respondent or for which coarse- and fine-grain responses proved to be ambiguous (e.g., skin color of actors in the video) in Experiment 1 were excluded. If participants provided a coarse-grain response when asked for fine-grain, or vice versa, one of the example questions was re-read to remind them of the distinction. Third, questions in both of the open-ended conditions were asked and answered verbally, and responses were audio-recorded. If participants reported multiple details in one sentence, they were requested to provide a confidence rating for each detail.

**Experiment 3.** The same interviewer conducted these interviews and the interviewing protocols were identical to those used in the written interviews in Experiment 1. Forty cued recall questions (three of which were skin-color questions and subsequently excluded), each allowing coarse- and fine-grain responses, were developed based on a written transcript for the credit card theft stimulus event. Questions elicited grain size distinctions for colors, numbers, descriptions of length, and descriptions of locations. The presentation, instructions and format matched those used in the previous experiments. The open-ended protocols matched those in the previous experiments, as did all other aspects of the procedure. (See Supplementary Materials pp.8-9 for forced-report questions.)

### **Procedure**

Experiments 1 and 3 interview sessions were conducted in university tutorial rooms or rooms at community libraries and involved testing of small groups of 5-15. Participants were told that they would be watching a short non-violent crime and were instructed to *please pay attention to the video as you will be answering some questions about it later*. The stimulus event was presented to participants on a projector screen or television depending on facilities.

Participants were then instructed to take out their interviews, read the instructions, and commence. Upon completion of the 30-80 minute session they were given a feedback sheet and paid an honorarium. Experiment 2 participants were interviewed in a laboratory on campus.

### **Dependent Measures**

The key measures were the total number of coarse-grain details reported, the number of correct coarse-grain details and the overall accuracy (i.e., proportion correct) of coarse-grain details, with these measures also indexed at and above each level of confidence. Similar statistics were also available for fine-grain details and preferred responses.

### **Scoring of Memory Reports**

Guided by participants' responses from previous studies (Weber & Brewer, 2008) a detailed list of accurate fine-grained responses, accurate coarse-grained responses, and inaccurate responses for each question was compiled against which to compare participants' responses. The key objective of our studies was to determine if the use of the grain size procedure could produce meaningful indications of increases in the quantity of accurate coarse-grain memory reports and, with or without the aid of supporting confidence judgments, provide a reliable indication of likely accuracy. In other words, while we were interested in comparative coarse-grain data obtained via the grain size procedure and the other open-ended approaches to interviewing, our objective was not to argue for some kind of definitive ranking of procedures based on the precise outputs of each. As indicated earlier, the precise outputs from any or all of the procedures must vary with the number and difficulty of questions asked, and also with an array of variables such as the characteristics of the stimulus event, the retention interval between encoding and initial interview, and so on. Our focus was on the extent of coarse-grain reporting and whether the grain size procedure could produce reliable coarse-grain information that might assist an investigation and yet – based on the comparisons with the various open-ended conditions – seemed unlikely to be elicited via the typical

interview protocols. As will become apparent, this motivation had important implications for the way in which we scored memory reports.

A detailed discussion of issues associated with data scoring is provided in the Supplementary Materials (pp. 10-11). Here we summarize the major points. To permit meaningful comparisons across interview conditions, we excluded from the summary descriptive statistics, and from any inferential statistical comparisons, any items that were reported under the specific cued recall prompts of the grain size procedure but never mentioned in any of the open-ended interview conditions. As we have already emphasized, any comparison of outputs across interview conditions is fraught with difficulty, so our focus was simply to provide some meaningful reference points for examining coarse-grain reporting. By not counting items reported under the forced report, grain size procedure but never mentioned in other conditions, any comparative considerations are at least not biased by counting items that most likely would never be reported by interviewees except under forced report.

We also excluded responses to questions relating to the skin color of characters which failed to elicit a coarse- versus fine-grain distinction, with responses such as pale, fair, light, white, tanned and dark reported interchangeably as both coarse- and fine-grain responses. The various exclusions resulted in the exclusion of responses to 21 questions (Experiment 1), 5 questions (Experiment 2), and 9 questions (Experiment 3). Thus, the maximum number of coarse-grain (and fine-grain) items that could be achieved under the grain size procedure was 43, 38 and 28 for Experiments 1 through 3, respectively.

**Inter-rater reliability.** For each of Experiments 1 and 2, five interviews (approximately 25% or 20 and 15 different interviews, respectively) from each interview technique were randomly selected and all responses were scored by two raters. Percentage agreement for each interview sampled was calculated for each raters' scores of the total number of details and the total number of accurate details reported. Across the various interviews using each technique, the range for percentage agreement spanned 93.3-100%. Cohen's kappa ranged from .68 to 1 for

accurate details reported in the grain size interview condition; elsewhere the number of responses was insufficient to derive a meaningful kappa.

## Results

We used an alpha level of .01 for inferential statistical comparisons; Cohen's  $f$  was used as the measure of effect size for ANOVAs, with 0.1, 0.25, and 0.4. the guidelines for small, medium and large effects, respectively. Note, however, that in a number of the comparisons of coarse-grain accuracy between conditions, inferential statistical comparisons were not conducted because individual witnesses volunteered so few coarse-grain responses that the proportion correct statistic was inherently unstable. Larger sample sizes would not have ameliorated this problem as the pivotal issue was the low base rate of coarse-grain reports, not one of low statistical power.

### Coarse-Grain Reporting

Table 1 shows the total number and the number of accurate coarse-grain items reported for each of the immediate interview conditions used in Experiments 1-3. Given the interdependence of fine and coarse grained responses, all response patterns (i.e., number coarse details, number fine details, number accurate coarse details, etc.) for the different interview conditions were compared with separate one-way ANOVAs. The preferred response outputs that followed the forced report interview's soliciting of coarse- and fine-grain responses are also obviously not independent of the forced report data and thus need to be interpreted in isolation, which we did by reference to the relevant confidence interval data. Three striking patterns emerged from the Table 1 means and confidence intervals (CIs) for both the total coarse-grain items reported, and the number of accurate coarse-grain details reported. The associated inferential statistical information highlights the extremely large effect sizes and appears in Table 1 in Supplementary Materials (p.12). First, compliance with the request to fully report under the grain size procedure was strong, with the total number of items reported almost at the maximum in each experiment. Second, hardly any coarse-grain details were

reported under any of the open-ended conditions. Third, the patterns of means and CIs indicate very similar patterns across conditions for coarse-grain responses when participants were exhorted to volunteer their preferred response from the coarse- and fine-grain options.

The comparative patterns across interview formats were also very similar for the number of accurate details reported, with the grain size procedure producing reporting of a large number of accurate coarse-grain details. Moreover, it was apparent that the negligible reporting of accurate coarse-grain items under open-ended conditions was clearly not due to very high frequencies of accurate fine-grain reporting instead. As shown in Table 2, in all experiments the overall mean number of accurate (coarse- and fine- grain combined) details was at least five times higher under forced report than under open-ended conditions. There was also no decisive evidence to suggest that accuracy (i.e., proportion correct) declined relative to accuracy of similar detail reported in the open-ended conditions (see Table 2 for descriptive statistics). Although inspection of the proportion correct data suggests possible accuracy variations across interview conditions, with so few participants reporting any coarse-grain details under open-ended conditions inferential statistical comparisons are not meaningful.

In sum, in all samples the forced report procedure enhanced accurate coarse-grain reporting without accuracy being meaningfully undermined. Also, when participants were forced to volunteer their preferred response from the coarse- and fine-grain options (i.e., Forced report: Preferred), the coarse-grain reporting patterns were similar to those under the initial forced report. Finally, the patterns of coarse-grain reporting on all measures were quite similar for the adapted SAI<sup>®</sup> and the other open-ended procedures.

### **Fine-Grain Reporting**

For fine-grain responses, the confidence interval data again clearly highlight the markedly higher outputs under the forced report procedure which produced two to three times as many accurate items as the open-ended conditions (see Supplementary Materials Tables 2 and 3,

pp.13-14). Not surprisingly, however, accuracy declined substantially relative to the open-ended conditions (see Table 3, main manuscript). As was the case for coarse-grain reporting, the data in Table 3 do not point to any obvious advantage in accuracy for any of the open-ended procedures when compared to the others.

### **Confidence and Reporting Accuracy**

To examine whether confidence in reported items assisted in assessing their likely accuracy, we examined the number of accurate details and proportion correct for coarse-grain responses at or above each level of confidence for each experiment. We describe the major patterns here, but for economy of illustration, provide a summary table for Experiment 2 only (Table 4). The corresponding tables for Experiments 1 and 3 appear as Tables 4 and 5 in Supplementary Materials (pp.15-16). We have not reported inferential statistical contrasts of coarse-grain accuracy at each level of (cumulative) confidence across the relevant immediate interview procedures because the overall number of coarse-grain observations produced under the open-ended conditions was so low.

Inspection of these tables for Experiments 1 and 2 (i.e., written and verbal reports) reveals that coarse-grain accuracy progressively increased as only those items reported with increasingly higher levels of confidence were encompassed. Second, when confidence was very high (i.e., 90-100%), not only was the ratio of number of accurate details reported for forced report versus open-ended interviews high but accuracy under forced report was also very high. These patterns were not evident in Experiment 3. Third, it is difficult to draw meaningful conclusions about the confidence-accuracy relation for the open-ended conditions because coarse-grain responses were so rarely volunteered and thus patterns for the various confidence cutoffs are unlikely to be stable.

The corresponding confidence-accuracy data for fine-grain responses for Experiments 1-3 are shown in Supplementary Tables 6-8 (pp.17-19). Although forced report accuracy rose in a systematic manner as only items reported with increasingly higher levels of confidence were



encompassed, it is clear from an inspection of these tables that the use of confidence did not lead to the identification of a substantially larger pool of fine-grain items likely to be as accurate as those produced under open-ended conditions. That is, if the mean number of accurate fine-grain items began to show an obvious advantage over that produced under open-ended interviews, generally the tradeoff was poorer accuracy. One-way ANOVAs, and importantly the relevant effect sizes, comparing forced and open-ended accuracy at the different levels of confidence confirmed that, in Experiments 1 and 3, forced fine-grain accuracy was compromised relative to at least one free recall condition at most levels of confidence (Supplementary Tables 9 and 11, pp.20 and 22). In Experiment 2 with verbal reporting, forced report accuracy only became meaningfully lower than open-ended interview accuracy when the confidence level cut-off extended down to around 60% or lower (Supplementary Table 10, p.21).

### **Discussion**

Across experiments we consistently found (a) reporting of coarse-grain details by eyewitnesses to a crime seldom happened under open-ended interview conditions, and (b) participants were able to provide abundant coarse-grain details under cued recall forced report conditions – without apparent cost to accuracy relative to the accuracy of corresponding types of detail reported in the open-ended conditions – regardless of whether they responded on a written questionnaire or in an individual face-to-face interview. Negligible reporting of accurate coarse-grain items under open-ended interview conditions was not due to very high frequencies of accurate fine-grain reporting, as the overall mean number of accurate (coarse- and fine- grain combined) details was never less than five times higher under forced report than open-ended conditions. In other words, the cued recall forced report procedure provided a means by which accessible, but likely to be unreported, coarse-grain items became available to interviewers. We note that, although the CIs and the effect sizes clearly demonstrate the increased accurate coarse-grain outputs under forced report relative to free recall, we cannot

say anything meaningful about possible differences between the free recall conditions. Not only are the CIs around the free recall means inconclusive but, more important, the reliability of any such contrasts is challenged due to the base rate of coarse-grain reports being so low.

There were two other noteworthy findings. First, in two of the three experiments item-level confidence provided a very neat diagnostic indicator of coarse-grain item accuracy, at least when obtained via an individual verbal report. The discrimination provided by confidence was not as impressive in Experiment 3, despite accuracy levels being similar, and we can offer no obvious explanation for this. Second, although forced report elicited more accurate fine-grain details than open-ended interview conditions, accuracy suffered. Nevertheless, the confidence-accuracy patterns were similar to those for coarse-grain reports.

Although participant numbers per interview condition were modest, the consistent data patterns were encouraging. Importantly, similar patterns were detected (a) when the open-ended procedure was a modified form of the SAI<sup>®</sup> procedure – that is, a realistic and established interview procedure including free recall and subsequent more systematic open-ended probing – and (b) when we varied the stimulus encoding materials and the specific questions asked. Thus, the procedure proved suitable for eliciting likely-to-be-unreported coarse-grain information when the questions were probing dimensions such as color of clothing, hair, features, objects and accessories; numeric details relating to distance, length or qualitative actions; and location.

Although – as we cautioned in the Introduction – the absolute levels of accurate details reported and overall accuracy are clearly dependent on the number and difficulty of forced report questions asked, these experiments suggest that a procedure involving cued recall forced reporting of coarse-grain detail offers potential as an adjunct to open-ended forensic interviewing techniques. It is perhaps tempting to assert that the lack of precision of coarse-grain details, when compared with fine-grain details, renders them much less important for the progress of an investigation, analogous to the distinction sometimes drawn in eyewitness

memory research between the utility of central and peripheral details (or verbatim and gist accounts). The central-peripheral distinction is often operationalized via independent raters' appraisals of which items are most appropriately described as central versus peripheral (e.g., Christianson, 1992) or probative versus non-probative (e.g., Lavis & Brewer, 2017), and such an examination could be applied to the evaluation of fine- and coarse-grain details. We argue, however, that such an approach sometimes misses a key point. Although an accurate fine-grain report of some item of information will obviously be more informative than its coarse-grain equivalent, the utility of any particular piece of information will often be context, or case, dependent. As we outlined earlier, coarse-grain details – despite perhaps appearing at face value as vague and unhelpful – may prove invaluable in police investigations for suggesting potential new directions of investigation, for corroborating the testimony of other witnesses and confirming the merits of a particular line of investigation, and for ruling out hypotheses about the identity of suspects.

Further, exactly the same coarse-grain detail that may be completely uninformative in one case may prove invaluable in another case depending on the broad array of evidence available. Knowing that a getaway car was light-colored may be very useful if there is only one witness who saw the getaway and there is other evidence pointing to a specific suspect who is known to always drive a beige (rather than a red or blue or black) car. The same item of information may be much less useful, however, if there were four other witnesses who apparently had a good view of the event and variously described the getaway car as white, beige, cream or very pale blue. It has already been noted that witnesses do relatively poorly at reporting descriptive features of perpetrators of crimes, a problem that may severely compromise the capacity of police to apprehend the perpetrator (e.g., Douglass et al., 2013). Consequently, successfully overcoming this problem would represent an important breakthrough for forensic investigators.

Critically, we are not suggesting that the approach we have used here could or should be adopted as an alternative to non-biased open-ended interviewing; rather, it might be used to

complement the information obtained from such approaches. Identifying a means of enhancing outputs of accurate fine-grain information would be even more valuable. However, the data were compelling in showing that forced reporting of fine-grain details undermined accuracy relative to open-ended free recall. Thus, although fine-grain forced reporting could be used to provide new leads, especially when an investigation appears to be at a dead-end and other open-ended questioning methods have been exhausted, the information obtained would have to be treated with extreme caution.

It is possible, of course, that if interviewees have reported incorrect information under free recall, subsequent forced report probing could elicit further distorted information. That said, any non-forced report interviewer probing of inaccurate free recall responses – responses which our data show are reliably elicited (e.g., in the SAI<sup>®</sup> condition, mean proportion correct was no better than 0.73 in any experiment) – presumably also has the potential to produce additional memory distortion, with the significance of any such distortion dependent on how pivotal any item of information turned out to be in the broader context of a particular investigation or trial. In other words, advancing the science of interviewing demands that interviewers – and those evaluating the outcomes of interviews (e.g., police, jurors, judges) – recognize that any item of information reported is potentially inaccurate, regardless of the conditions under which it is elicited or the characteristics of the report (i.e., the precision of the detail provided and the associated confidence). In a similar vein, it is well established that witness confidence is vulnerable to inflation (e.g., Semmler, Brewer, & Wells, 2004), making it crucial that the confidence associated with the initial memory report is preserved as the record of interest (Brewer & Wells, 2006).

### **Generalizability and Limitations**

Although we have been comparing coarse-grain outputs across conditions, we again emphasize that we make absolutely no quantitative claim about exactly how much more accurate coarse-grain information might come to light (e.g., 0%, 10% more, 20% more, etc.) in

any individual investigation through applying such a forced-report procedure rather than using free recall only. Recall that to enable meaningful comparisons between interview conditions we adopted some scoring rules to accommodate the fact that some items probed using the grain size procedure might never emerge under open questioning, while some items that emerged under the latter were not – or perhaps could not be – probed by the grain size procedure. An unavoidable consequence of this approach is that our data do not speak to a number of other possibilities, some of which we canvass below based on a purely descriptive examination of items not encompassed by our comparisons.

One possibility is that our approach understated the utility of the forced-reporting approach. This could happen by excluding items (that were never reported at open interview) that may have been useful in an investigation. Undoubtedly this could happen, although much of the information generated only by the forced report procedure was judged by a panel of 10 independent eyewitness interviewing and memory researchers as probably of low relevance to an investigation (though note our earlier comments regarding the importance of context in assessing relevance). Or, the grain-size protocol may have omitted to probe for details that would be amenable to grain size questions and were reported at open interview. This too could happen although our examination showed few details that would be readily amenable to a grain size distinction but not covered by a grain size question were reported in open interviews. The types of information that were captured in free recall but not amenable to grain-size questions were often details about actions and judgments about the character of those involved in the event.

A second possibility is the accuracy of coarse-grain items obtained under forced report may have been overstated by the protocols we adopted to permit some kind of comparison between interview conditions. For example, some items excluded from comparisons because they were never reported by any witnesses in free recall may have been items simply deemed to be so irrelevant as to not warrant reporting. But some items were likely not reported because

they were items that for some reason (e.g., shorter exposure duration, lack of clear vision) were less likely to be remembered and, hence, less likely to be accurate under forced report. To the extent the latter was the case, forced report accuracy may well have been lower than what we found.

A third possibility is that the accuracy of coarse-grain items obtained under forced report may be overstated because we had to deal with some items differently to others. For example, as noted we excluded skin color items because the same terms were sometimes used for both coarse and fine responses even though the individual participant clearly differentiated skin color appropriately. In other words, for some reason the meaning of individual terms did not map neatly onto our fine-coarse distinction. In practice this should not matter because a witness omitting skin color detail could easily be probed with an appropriate coarse-grain question (e.g., Was their skin dark or light or in-between?). In a similar vein, where the information available in the encoding event was ambiguous, we accepted some responses as accurate when the witness's response did not match the provided responses (e.g., 'medium' rather than 'light or dark' for color of jeans, see Supplementary Materials, p.10). Again, in an actual interview such a coarse response would just provide a lesser degree of specificity than a light or dark response.

In actual investigations any variations in detail reported under forced versus open-ended procedures are likely to be shaped by variables such as the stimulus encoding conditions, the delay between the event and the first interview and, as we have noted, the number and nature of the questions asked. In sum, any attempt to make 'formal' comparisons between conditions gives rise to an array of classification issues which need to be recognized when examining the data patterns. Having said that, it is very clear that the open-ended conditions – even the SAI<sup>®</sup> – consistently failed to elicit accurate coarse-grain detail whereas the grain size procedure used here was very effective in doing so.

Nor do we claim that the approach outlined guarantees accurate reporting. Accuracy rates in all conditions were well below ceiling. The obvious exception (see Experiment 2 patterns) was when forced verbal (rather than written) reports were required and coarse-grain item-level confidence was very high. Under those conditions the information elicited might prove a dependable aid for investigators.

### **Future Research**

Numerous follow-up investigations are suggested by our findings. First, for example, would the patterns of results be exactly the same if the grain size procedure was combined with (i.e., used immediately following) a thorough open-ended interview to probe for omitted, but potentially useful, coarse-grain information that may be available to the witness? Second, would it make any difference if witnesses were only probed with (and instructed in responding to) coarse-grain questions – or might the questioning need to incorporate both coarse- and fine-grain probes, as happened in our studies in order for witnesses' to appreciate fully the distinctions between coarse- and fine-grain responses? Third, would the effectiveness of the procedure be particularly suited to certain types of items, an issue that would need to be examined systematically using pools of questions that spanned particular item categories (e.g., color, numeric, a series of physical actions)? Our descriptive examination of witness responses indicated that all open interview techniques elicited 'action and sequence' details not readily amenable to grain size probes and hence not picked up by the grain-size procedure. Fourth, what might be the impact of a delay between the event and the initial interview? And fifth how might the procedure work with child witnesses given their proneness to error in response to closed questions? Answers to questions like these are obviously needed in order to ensure when and how a forced report approach such as used in these studies could be best integrated with open questioning.

There are also important issues that might arise should such a procedure be implemented by investigators. For example, appropriate training and performance monitoring for

interviewers would obviously be critical for a number of reasons, not the least being to ensure that the procedure is not misused to confirm investigators' hypotheses about the culprit. It would also be of interest to examine how lawyers and jurors react to testimonial details elicited using such an approach, and what would be required to ensure the testimony received their thoughtful consideration.

The importance of investigating just how such a procedure might be best integrated with open-ended questioning can be highlighted by considering an example of how misleading reports might be elicited. First, in an actual case where the ground truth of a certain detail is not known, a question such as *what was the color of the beanie worn by the robber near the door?* may lead to witnesses reporting something they had never observed. Of course, as already noted, open-ended interviews are also not immune to the production of confabulations. It has been well documented that cognitive schema can shape memory reports (e.g., Tuckey & Brewer, 2003), something that was illustrated in Experiments 1 and 2 by witnesses reporting the presence of a (non-existent) weapon. That aside, in a forensic setting any forced-choice probes for coarse grain detail would have to follow a preliminary scoping of information about the event and the people involved using the recommended open-ended interviewing approach. Thus, a witness who volunteered under open-ended questioning or prompting that a perpetrator was wearing some form of headwear, but failed to provide any further detail about the headwear, could be probed as to whether the headwear was hard-soft, large-small, and dark-light-colorful. Or a witness who referred to a getaway car, but provided no further detail, might be probed as to whether it was large-small, dark-light, old-new, etc. The number of forced-choice probes used by an investigator would be dictated by the specific items of information elicited by open-ended questioning, the level of detail provided in those items, and whether a coarse-grain probe was compatible with the nature of the item. Thus, an investigator might use one probe or several.



Despite the many unresolved issues these studies show that there may well be a considerable amount of potentially very useful coarse-grain information tucked away in witnesses' memories that may be accessible but unlikely to be reported spontaneously. Perhaps some of this information might be accessed using specially tailored instructions or payoffs (though the latter seems impractical in applied situations). Though, given the relatively modest changes in coarse-grain reporting seen in studies such as McCallum et al. (2016), it seems unlikely that such manipulations could produce the gains in coarse-grain reporting observed here under forced reporting. One obstacle to identifying alternative manipulations for improving coarse-grain reporting is our limited understanding of why this information is not reported spontaneously. It has been suggested in research on memory reporting in other domains that coarse-grain information may be perceived as uninformative and hence not worth reporting (Ackerman & Goldsmith, 2008). It has also been reported that witnesses to a mock-crime sometimes show a bias towards reporting fine-grain information, and under-reporting coarse-grain, with this bias appearing stronger when financial and social consequences contingent on memory reporting were minimal (McCallum et al., 2016). We need research that refines our understanding of how witnesses' perceptions of informativeness are shaped and, in turn, points to ways of encouraging reporting that ensures that potentially informative items of information – whether coarse- or fine-grain – are brought to the fore.

## **Conclusion**

We believe that the science of investigative interviewing is likely to be advanced by devising and evaluating novel approaches. Some of these approaches might prove to be generally applicable. Others might be most appropriately used given the particular circumstances of an investigation. Or procedures such as we have trialed here might be useful for certain individuals. For example, some individuals (e.g., those with ASD) have difficulty with open-ended procedures such as the CI, and benefit when support for retrieval is provided at test by providing more cues to the to-be-remembered information (Bowler, Gardiner, &

Berthollier, 2004; Maras & Bowler, 2010). It seems most unlikely that any approach will guarantee error-free performance. Thus, regardless of the interviewing approach(es) adopted, a critical focus for the field should be finding ways of ensuring that practitioners clearly appreciate the limitations of any memory report data they manage to elicit.

## References

- Ackerman, R., & Goldsmith, M. (2008). Control over grain size in memory reporting: With and without satisficing knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1224-1245. doi: 10.1037/a0012938
- Bell, B. E., & Loftus, E. F. (1988). Degree of detail of eyewitness testimony and mock juror judgments. *Journal of Applied Social Psychology*, *18*, 1171-1192. doi: 10.1111/j.1559-1816.1988.tb01200.x
- Bell, B. E., & Loftus, E. F. (1989). Trivial persuasion in the courtroom: The power of (a few) minor details. *Journal of Personality and Social Psychology*, *56*, 669-679. doi: 10.1037/0022-3514.56.5.669
- Bowler, D. M., Gardiner, J. M., & Berthollier, N. (2004). Source memory in adolescents and adults with Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *34*, 533-542. doi: 10.1007/s10803-004-2548-7
- Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*, 11-30.
- Christianson, S-A. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin*, *112*, 284-309. doi: 10.1037/0033-2909.112.2.284
- Douglass, A. B., Brewer, N., Semmler, C., Bustamante, L., & Hiley, A. (2013). The dynamic interaction between eyewitnesses and interviewers: The impact of differences in perspective on memory reports and interviewer behavior. *Law and Human Behavior*, *37*, 290-301. doi: 10.1037/lhb0000034
- Fisher, R.P., Falkner, K. L., Trevisan, M., & McCauley, M. R. (2000). Adapting the cognitive interview to enhance long-term (35 years) recall of physical activities. *Journal of Applied Psychology*, *85*, 180-189. doi: 10.1037/0021-9010.85.2.180
- Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative*

*interviewing: The cognitive interview*. Springfield, Illinois: Charles C. Thomas.

Fisher, R. P., Milne, R., & Bull, R. (2011). Interviewing cooperative witnesses. *Current Directions in Psychological Science*, 20, 16-19. doi: 10.1177/0963721410396826

Fisher, R. P., Schreiber Compo, N. S., Rivard, J., & Hirn, D. (2014). Interviewing witnesses. In T. J. Perfect & D. S Lindsay (eds.), *The Sage handbook of applied memory* (pp.559-578). Los Angeles, CA: Sage.

Gabbert, F., Hope, L., Carter E., Boon, R., & Fisher, R. (2015). The role of initial witness accounts within the investigative process. In G. Oxburgh, T. Myklebust, T. Grant, & R. Milne (Eds.), *Communication in investigative and legal contexts: Integrated approaches from forensic psychology, linguistics and law enforcement*. Chichester: Wiley-Blackwell.

Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and Human Behavior*, 33, 298-307. doi: 10.1007/s10979-008-9146-8

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

Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General*, 131, 73-95. doi: 10.1037/0096-3445.131.1.73

Grice, H. P. (2002). Logic and conversation. In D. J. Levitin (Ed.), *Foundations of cognitive Psychology: Core readings* (pp. 719-732). London: MIT Press.

Hope, L., Gabbert, F., & Fisher, R. P. (2011). From laboratory to the street: Capturing witness memory using the Self-Administered Interview. *Legal and Criminological Psychology*, 16, 211-226. doi: 10.1111/j.2044-8333.2011.02015.x

Lamb, M. E., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2008). *Tell me what happened: Structured investigative interviews of child victims and witnesses*. West Sussex, England:

John Wiley & Sons Ltd.

Lavis, T., & Brewer, N. (2017). Effects of a proven error on evaluations of witness testimony.

*Law and Human Behavior*. doi: 10.1037/lhb0000217

Luna, K., Higham, P. A., & Martín-Luengo, B. (2011). Regulation of memory accuracy with multiple answers: The plurality option. *Journal of Experimental Psychology: Applied*, 17,

148-158. doi: 10.1037/a0023276

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

Maras, K. L., & Bowler, D. M. (2010). The Cognitive Interview for eyewitnesses with Autism

Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 40, 1350-1360. doi:

10.1007/s10803-010-0997-8

McCallum, N., Brewer, N., & Weber, N. (2016). Memorial monitoring and control: How

confidence and social and financial consequences affect eyewitnesses' reporting of fine-grain information. *Applied Cognitive Psychology*, 30, 375-386. doi: 10.1002/acp.3212

Memon, A., Wark, L., Holley, A., Bull, R., & Koehnken, G. (1996). Interviewer behaviour in

investigative interviews. *Psychology, Crime and Law*, 3, 135-155. doi:

10.1080/10683169608409800

Pipe, M-E., Gee, S., Wilson, J. C., Egerton, J. M. (1999). Children's recall 1 or 2 years after an

event. *Developmental Psychology*, 35, 781-789. doi: 10.1037/0012-1649.35.3.781

Powell, M., Fisher, R., & Wright, R. (2005). Investigative interviewing. In N. Brewer & K. D.

Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 11-42). New York: Guilford Press.

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

- Robinson, M. D., & Johnson, J. T. (1996). Recall memory, recognition memory, and the eyewitness confidence–accuracy correlation. *Journal of Applied Psychology, 81*, 587-594. doi: 10.1037/0021-9010.81.5.587
- Scoboria, A., & Fisico, S. (2013). Encouraging and clarifying “don’t know” responses enhances interview quality. *Journal of Experimental Psychology: Applied, 19*, 72-82. doi: 10.1037/a0032067
- Semmler, C., Brewer, N., & Wells, G. L. (2004). Effects of postidentification feedback on eyewitness identification and nonidentification confidence. *Journal of Applied Psychology, 89*, 334-346.
- Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied, 9*, 101-118. doi: 10.1037/1076-898X.9.2.101
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373. doi: 10.1037/h0020071
- Vrij, A., Hope, L., & Fisher, R. P. (2014). Eliciting reliable information in investigative interviews. *Policy Insights from Behavioral and Brain Sciences (PIBBS), 1*, 129-136. doi: 10.1177/2372732214548592
- Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology: Applied, 10*, 156-172. doi: 10.1037/1076-898X.10.3.156
- Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of confidence. *Journal of Experimental Psychology: Applied, 14*, 50-60. doi: 10.1037/1076-898X.14.1.50
- Yaniv, I., & Foster, D. P. (1995). Graininess of judgement under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General, 124*, 424-432. doi: 10.1037/0096-3445.124.4.424

Footnote

<sup>1</sup>The adapted version of the SAI<sup>®</sup> is a copyrighted instrument and is not included in Supplementary Materials, but the authors of the instrument have advised that a copy can be obtained by contacting them directly.

Table 1

*Mean (SD) and 95% Confidence Intervals for total number and the number of accurate coarse-grain items for the various interview conditions in Experiments 1-3.*

Measure	Interview Format				
	Forced Report: Initial	Forced Report: Preferred	Adapted SAI <sup>©</sup>	Free Recall Plus Confidence	Free Recall
Number Coarse-Grain Items					
Experiment 1	42.17	26.17	1.16	1.58	1.12
	(1.24)	(8.19)	(1.41)	(2.34)	(1.59)
	41.64-42.69	22.71-29.63	0.58-1.74	0.60-2.57	0.46-1.78
Experiment 2	37.20	21.80	1.90	0.60	-
	(1.40)	(4.47)	(1.74)	(0.88)	
	36.55-37.85	19.71-23.89	1.08-2.72	0.19-1.01	
Experiment 3	26.60	14.53	1.07	0.40	1.38
	(2.41)	(2.88)	(1.10)	(0.63)	(1.31)
	25.26-27.94	12.94-16.13	0.46-1.68	0.05-0.75	0.68-2.07
Number Accurate Coarse-Grain Items					
Experiment 1	27.38	16.29	0.60	1.08	0.68
	(3.89)	(5.77)	(0.82)	(1.82)	(1.07)
	25.73-29.02	13.86-18.73	0.26-0.94	0.32-1.85	0.24-1.12
Experiment 2	23.75	12.80	1.25	0.20	-
	(3.92)	(4.28)	(1.62)	(0.52)	
	21.92-25.58	10.80-14.80	0.49-2.01	-0.04-0.44	
Experiment 3	16.33	8.80	0.73	0.33	0.50
	(2.82)	(3.08)	(0.80)	(0.62)	(0.89)
	14.77-17.89	7.10-10.50	0.29-1.18	-0.01-0.68	0.02-0.98

Note. Maximum numbers of coarse-grain responses achievable were 43, 38 and 28 for

Experiments 1, 2 and 3, respectively.



Table 2

*Mean (SD) and 95% Confidence Intervals for (a) number of accurate coarse- and fine-grain items combined and (b) proportion correct coarse-grain items for the various interview conditions in Experiments 1-3.*

Measure	Interview Format				
	Forced Report: Initial	Forced Report: Preferred	Adapted SAI <sup>©</sup>	Free Recall Plus Confidence	Free Recall
Number Accurate Coarse- and Fine-Grain Items					
Experiment 1	46.63 (6.21) 44.00-49.25	25.29 (4.19) 23.52-27.06	8.40 (2.97) 7.17-9.63	7.08 (3.19) 5.74-8.43	5.72 (3.37) 4.33-7.11
Experiment 2	44.05 (6.66) 40.93-47.17	23.70 (4.60) 21.55-25.85	8.50 (4.45) 6.42-10.58	5.50 (2.35) 4.40-6.60	-
Experiment 3	27.80 (3.82) 25.68-29.92	16.33 (2.55) 14.92-17.75	4.87 (2.26) 3.61-6.12	3.27 (2.09) 2.11-4.42	3.56 (2.34) 2.32-4.81
Proportion Correct Coarse-Grain Items					
Experiment 1 <sup>a</sup>	0.65 (0.09) 0.61-0.69	0.62 (0.11) 0.58-0.67	0.60 (0.45) 0.36-0.84	0.67 (0.42) 0.42-0.93	0.62 (0.43) 0.37-0.86
Experiment 2 <sup>b</sup>	0.58 (0.08) 0.55-0.62	0.63 (0.12) 0.58-0.69	0.68 (0.24) 0.57-0.79	0.73 (0.25) 0.62-0.85	-
Experiment 3 <sup>c</sup>	0.61 (0.09) 0.57-0.66	0.60 (0.15) 0.52-0.68	0.71 (0.25) 0.50-0.92	0.80 (0.45) 0.24-1.36	0.30 (0.40) 0.04-0.56

<sup>a</sup>SAI  $n = 16$ ; Free Recall Plus Confidence  $n=13$ ; Free Recall  $n=14$ ; <sup>b</sup>SAI  $n=15$ ; Free Recall Plus Confidence  $n=8$ ; <sup>c</sup>SAI  $n=8$ ; Free Recall Plus Confidence  $n=5$ ; Free Recall  $n=12$ .

Table 3

*Mean (SD) and 95% Confidence Intervals for proportion correct fine-grain items reported for the various immediate interview conditions in Experiments 1-3.*

Proportion Correct Fine-Grain Items	Interview Format				
	Forced Report: Initial	Forced Report: Preferred	Adapted SAI <sup>®</sup>	Free Recall Plus Confidence	Free Recall
Experiment 1	0.46 (0.08) 0.43-0.50	0.59 (0.19) 0.51-0.67	0.73 (0.13) 0.68-0.79	0.78 (0.17) 0.71-0.85	0.71 (0.22) 0.62-0.80
Experiment 2	0.54 (0.10) 0.49-0.59	0.70 (0.17) 0.62-0.78	0.69 (0.25) 0.58-0.81	0.79 (0.24) 0.67-0.90	-
Experiment 3	0.41 (0.08) 0.37-0.46	0.57 (0.13) 0.50-0.64	0.54 (0.14) 0.46-0.62	0.66 (0.29) 0.50-0.82	0.78 (0.21) 0.65-0.90

Experiment 3: Free Recall  $n = 14$

Table 4

*Experiment 2. Mean number of accurate coarse-grain items, and mean (& SD) proportion correct for all responses at and above each level of confidence for the three immediate interview conditions which recorded confidence.*

	Forced Report:		Adapted		Free Recall Plus	
	Initial		SAI <sup>®</sup>		Confidence	
	N <sub>A</sub>	P <sub>A</sub>	N <sub>A</sub>	P <sub>A</sub>	N <sub>A</sub>	P <sub>A</sub>
100	4.75	0.94	0.35	0.86	0.10	0.40
		(0.12)		(0.38)		(0.55)
≥ 90	7.70	0.87	0.45	0.72	0.20	0.50
		(0.18)		(0.44)		(0.55)
≥ 80	11.40	0.80	0.60	0.62	0.20	0.38
		(0.17)		(0.46)		(0.52)
≥ 70	14.10	0.76	0.90	0.67	0.20	0.38
		(0.14)		(0.41)		(0.52)
≥ 60	16.70	0.74	1.10	0.64	0.20	0.38
		(0.14)		(0.43)		(0.52)
≥ 50	18.45	0.71	1.15	0.62	0.20	0.38
		(0.12)		(0.42)		(0.52)
≥ 40	19.50	0.67	1.25	0.62	0.20	0.38
		(0.11)		(0.42)		(0.52)
≥ 30	20.45	0.66	1.25	0.60	0.20	0.38
		(0.10)		(0.41)		(0.52)
≥ 20	21.45	0.65	1.25	0.60	0.20	0.38
		(0.10)		(0.41)		(0.52)
≥ 10	22.65	0.64	1.25	0.60	0.20	0.38
		(0.10)		(0.41)		(0.52)
≥ 0	23.75	0.64	1.25	0.60	0.20	0.38
		(0.10)		(0.41)		(0.52)

N<sub>A</sub> = Number accurate (Mean), P<sub>A</sub> = Proportion accurate (Mean and SD)

*Figure 1.* Key elements of the different interview report modes and formats used in Experiments 1-3.

### **Interview Report Mode**

#### **Experiment 1:**

- Written report

#### **Experiment 2:**

- Verbal report

#### **Experiment 3:**

- Written report

### **Interview Conditions**

#### **Grain Size procedure (Experiments 1, 2 & 3):**

- Step 1: Forced report coarse-grain item 1 (or fine-grain)
- Step 2: Forced report fine-grain item 1 (or coarse-grain)
- Step 3: Forced report preferred grain size item 1
- Step 4: Forced report coarse-grain item 2 (or fine-grain)  
etc.

OR

#### **Free Recall (Experiments 1, 2 & 3):**

- Self Administered Interview (SAI<sup>®</sup>) – includes follow-up prompts

OR

#### **Free Recall (Experiments 1 & 3):**

- Report everything – no follow-up prompts

OR

#### **Free Recall (Experiments 1, 2 & 3):**

- Report everything plus confidence – no follow-up prompts

Supplementary Materials

The Grain Size Interview Instructions, Examples, and Questions

**PLEASE READ THE FOLLOWING INSTRUCTIONS CAREFULLY.**

**The video you have just watched showed a bank robbery**

**Please answer ALL the following questions about the video you just saw**

**For each question you will be asked to give 2 responses**

One answer will be a very specific detail – for example “sky blue” as a response to “what colour?”

The other answer will be a broader detail – for example “light” as a response to “what colour?”

**Make sure you read each question carefully because the two types of responses won’t be in the same order for each question**

**For each answer there is a confidence scale going from 0% (very very unsure) to 100% (very very sure)**

**For each answer please circle how sure you are that the response you just wrote is correct**

1

1

Here are some examples:

*What colour were the walls in the bank?*

Exact colour white

How sure are you that your response is correct? (Please circle)										
0%	10%	20%	<u>50%</u>	40%	50%	60%	70%	80%	90%	100%
Very	Very				Half					Very
Very	Very				Sure					Very
Unsure										Sure

General colour light  
(light, medium, dark)

How sure are you that your response is correct? (Please circle)										
0%	10%	20%	30%	40%	50%	60%	<u>70%</u>	80%	90%	100%
Very	Very				Half					Very
Very	Very				Sure					Very
Unsure										Sure

*How many people were in the bank during the robbery?*

Between 4 and 8  
(range)

How sure are you that your response is correct? (Please circle)										
0%	10%	20%	30%	40%	50%	60%	70%	<u>80%</u>	90%	100%
Very	Very				Half					Very
Very	Very				Sure					Very
Unsure										Sure

Exact 6

How sure are you that your response is correct? (Please circle)										
0%	<u>10%</u>	20%	30%	40%	50%	60%	70%	80%	90%	100%
Very	Very				Half					Very
Very	Very				Sure					Very
Unsure										Sure

2

2

There are 64 questions

After you finish answering ALL the questions put your hand up  
The researcher will come to your desk and explain what you need to do next

Just remember to answer ALL the questions about the video and  
circle how sure you are that your responses are correct

If you have any questions put your hand up and the researcher will come to your desk

When you finish please stay at your desk until everyone has finished

You have now finished the first part of this interview

Put your hand up and the researcher will come to your desk and explain what you need to do next

For the second part of the interview you will be asked to pick ONLY ONE out the two responses you gave for  
each question. The responses you pick will be your final responses to the questions

If you have any questions put your hand up and the researcher will come to your desk

When you finish please stay at your desk until everyone has finished

1. What shade was the skin of the robber near the door?
2. What was the color of the beanie worn by the robber near the door?
3. What was the color of the jacket worn by the robber near the door?
4. What was the color of the top worn by the robber near the door?
5. What was the color of the pants worn by the robber near the door?
6. What was the main color of the bag carried by the robber near the door?
7. What was the hair color of the robber near the door?
8. What was the hair color of the customer?
9. What shade was the skin of the customer?
10. What was the color of the jumper worn by the customer?
11. What shade was the skin of the female bank teller watching the robbery?
12. What was the hair color of the female bank teller watching the robbery?
13. What was the color of the jacket worn by the female bank teller watching the robbery?
14. What was the main color of the scarf worn by the female bank teller watching the robbery?
15. What shade was the skin of the male bank teller?
16. What was the hair color of the male bank teller?
17. What was the color of the shirt worn by the male bank teller?
18. What was the color of the tie worn by the male bank teller?
19. What shade was the skin of the female bank teller giving the money?
20. What was the hair color of the female bank teller giving the money?
21. How long was the female bank teller giving the money's hair?
22. What was the color of the jacket worn by the female bank teller giving the money?
23. What was the main color of the scarf worn by the female bank teller giving the money?
24. How many rings was the female bank teller giving the money wearing?

25. What shade was the skin of the robber at the counter?
26. What was the main color of the hat worn by the robber at the counter?
27. What was the main color of the jacket worn by the robber at the counter?
28. What was the eye color of the robber at the counter?
29. What was the hair color of the robber at the counter?
30. What was the color of the t-shirt worn by the robber at the counter?
31. What was the main color of the shirt worn by the robber at the counter?
32. What was the main color of the bag carried by the robber at the counter?
33. What was the color of the belt worn by the robber at the counter?
34. What shade was the skin of the accomplice in the car?
35. What was the hair color of the robber in the car?
36. What was the color of the jacket worn by the witness sitting outside?
37. What was the color of the t-shirt worn by the witness sitting outside?
38. What was the color of the pants worn by the witness sitting outside?
39. How long did the robbery take to occur from the time you saw the robbers until they drove away?
40. How many robbers were in the video?
41. How many staff members were in the bank?
42. How many customers were in the bank?
43. How many witnesses were in the bank?
44. How many witnesses were outside?
45. How many women were in the video?
46. How many men were in the video?
47. How many times did the robber at the counter talk?
48. How many times did the female bank teller giving the money talk?
49. How many times did the female bank teller giving the money put money from the



drawer into the bag?

50. How many times did the robber at the counter get violent/physical with the female bank teller giving the money?
51. What section of the bank was the customer standing in?
52. What section of the bank was the female bank teller watching the robbery standing in?
53. What section of the bank was the male bank teller standing in?
54. What section of the bank was the robber near the door standing in?
55. What section of the bank was the robber at the counter standing in?
56. What section of the bank was the female bank teller giving the money standing in?
57. How long did it take the robbers to run from the bank to the car?
58. What color was the car?
59. How many doors did the car have?
60. How old was the car?
61. How many times did you see the money in the drawer?
62. How many compartments was the money drawer separated into?
63. What color was the carpet in the bank?
64. What color were the chairs at the back of the bank?

**The Free Recall Interview Instructions and Examples**

**PLEASE READ THE FOLLOWING INSTRUCTIONS CAREFULLY**

**The video you have just watched showed a bank robbery**

**Please write down EVERYTHING you can remember  
about the video you just saw**

Here are some examples:

*The walls in the bank were white*

*There were 6 people in the bank during the robbery*

There are 10 pages here for you to write on

If you need more pages let the researcher know

You might not need all 10 pages

Just remember to write down EVERYTHING you can remember  
about the video you just saw

If you have any questions put your hand up  
and the researcher will come to your desk

When you finish please stay at your desk until everyone has finished

**The Free Recall plus Confidence Interview Instructions and Examples**

**PLEASE READ THE FOLLOWING INSTRUCTIONS CAREFULLY.**

**The video you have just watched showed a bank robbery**

**Please write down EVERYTHING you can remember about the video you just saw**

**After each line there is a confidence scale going from  
0% (very very unsure) to 100% (very very sure)**

**Please circle how sure you are that the response you just wrote is correct**

Here are some examples:

*The walls in the bank were white*

**How sure are you that your response is correct? (Please circle)**

*There were 6 people in the bank during the robbery*

**How sure are you that your response is correct? (Please circle)**

There are 20 pages here for you to write on

If you need more pages let the researcher know

You might not need all 20 pages

Just remember to write down EVERYTHING you can remember about the video and circle  
how sure you are that your responses are correct

If you have any questions put your hand up and the researcher will come to your desk  
When you finish please stay at your desk until everyone has finished

### **Credit Card Theft Grain Size Interview Questions**

1. What shade was the skin of the robber?
2. What was the hair color of the robber?
3. What was the main color of the jacket worn by the robber?
4. What was the color of the jeans worn by the robber?
5. What was the color of the t-shirt worn by the robber?
6. What was the color of the necklace and pendant worn by the robber?
7. What was the hair color of the waiter?
8. What shade was the skin of the waiter?
9. What was the color of the top worn by the waiter?
10. What was the color of the jeans worn by the waiter?
11. What shade was the skin of the customer?
12. What was the color of the top worn by the customer?
13. What was the color of the pants worn by the customer?
14. How long did the robbery take to occur from the time the robber entered the restaurant until he ran away?
15. How many men were in the video?
16. How many women were in the video?
17. How many witnesses were in the video?
18. How many times did the customer talk to the waiter?
19. How many times did the waiter talk to the customer?
20. How many times did the waiter talk to the robber?
21. How many times did the robber talk to the waiter?

22. How many times did the robber reach for the credit card?
23. How long did the robber wait at the counter after stealing the credit card?
24. How long did it take the waiter to realise the card was missing?
25. How many times did the phone ring before the waiter answered it?
26. How many times did the waiter talk to the customer on the phone?
27. How many times did the robber look out the restaurant door after stealing the
28. How long did the waiter spend looking for the robber outside?
29. What colour was the car parked right outside the restaurant?
30. How many doors did the car have?
31. How many people were walking down the street when the robber entered the restaurant?
32. How many people were walking down the street when the robber left the restaurant?
33. How many cars were parked on the street?
34. What color was the customer's credit card?
35. What was the color of the plate that the credit card was on?
36. What color was the telephone?
37. What was the main color of the furniture in the restaurant?
38. What was the main color of the bricks on the outsideof the restaurant?
39. What section of the restaurant was the robber standing in for most of the time?
40. What section of the restaurant was the waiter standing in for most of the time?

## Data scoring

*Grain size procedure.* A list of accurate coarse-grained responses, accurate fine-grained responses, and inaccurate responses for each question was compiled. Responses were coded as either coarse- or fine-grain, scored as accurate, inaccurate or missing, and the corresponding confidence ratings were recorded. Missing details were coded as missing due to (1) no distinction between coarse- and fine-grain, (2) an “I don’t know” response, (3) a blank response, and (4) a discrepancy between the question and response (e.g. a numerical answer for a question asking about color or location).

In the course of scoring interviews it was noted that some participants reported “medium” as a coarse-grain response to certain questions pertaining to color or length. Where “medium” had been scored as inaccurate, it became clear that this led to cases where the fine-grain response was accurate, but the corresponding coarse-grain response was scored as being inaccurate, a situation which is unusual given that coarse-grained details are presumed to be extensions of the fine-grained response. Given the subjectivity associated with “medium” as a response, scoring the medium coarse-grain responses as being inaccurate would have underestimated coarse-grain accuracy. For example, while it was apparent that one of the robbers was wearing a black (dark) jacket, it was not clear whether his blue jeans were light or a medium shade, and quite likely to be either depending on the individual. Consequently, the scoring protocol was reviewed by two raters who agreed that out of the 33 color and length items in the grain size procedure, 19 were ambiguous enough for “medium” to be considered an accurate coarse-grained response. Interviews were rescored according to these modified guidelines.

*Open-ended techniques.* A list of free recall sentences corresponding to the coarse- and fine-grain details was compiled. Details from the adapted SAI<sup>®</sup>, free recall interview and free recall confidence interview that matched grain size questions were coded as either coarse- or fine-grain and scored as either accurate or inaccurate. For example, “the robber near the door

wore a dark jacket” and “the robber near the door wore a black jacket” were considered to be accurate coarse- and fine-grain details, respectively, to the grain size procedure question “What was the color of the jacket worn by the robber near the door”. Qualitative statements such as “around”, “a few”, “some”, or “roughly” were scored as being inaccurate coarse-grain details given that they did not allude to a particular numerical range. If participants failed to comply with the confidence rating instructions and only one confidence rating was provided for multiple details, that confidence level was assigned to each accompanying detail. For example, for the response “the main robber had a white hat and wore a brown jacket” reported with 80% confidence, both “white hat” and “brown jacket” were recorded as being reported with 80% confidence.

## Experiments 1-3

**Inferential Statistics Comparing Total Number of Coarse-Grain Items Reported and  
Number of Accurate Coarse-Grain Items Reported**

Table 1

*One-Way ANOVAs on Total Number and Number of Accurate Coarse-Grain Items Reported*

Number Coarse- Grain Items	Statistic*			
	<i>F</i>	<i>df</i>	<i>p</i>	<i>f</i>
Experiment 1	3526.40	3, 94	<.001	10.61
Experiment 2	4477.18	2, 57	<.001	12.53
Experiment 3	1089.23	3, 57	<.001	7.57
Number Accurate Coarse-Grain Items				
Experiment 1	860.18	3, 94	<.001	5.24
Experiment 2	581.98	2, 57	<.001	4.52
Experiment 3	390.65	3, 57	<.001	4.53

\*In Experiments 1 and 3, forced report compared with free recall, free recall plus confidence and adapted SAI.<sup>©</sup> In Experiment 2, forced report compared with free recall plus confidence and adapted SAI.<sup>©</sup>



Table 2

*Mean (SD) and 95% Confidence Intervals for total number and number of accurate fine-grain items reported for the various immediate interview conditions in Experiments 1-3.*

Number Fine-Grain Items	Interview Format				
	Forced Report: Initial	Forced Report: Preferred	Adapted SAI <sup>®</sup>	Free Recall (confidence)	Free Recall
Experiment 1	41.79	16.13	10.64	7.79	7.20
	(1.64)	(8.17)	(3.26)	(3.11)	(4.15)
	41.10-42.48	12.68-19.57	9.29-11.99	6.48-9.10	5.49-8.91
Experiment 2	37.80	15.55	9.95	6.80	-
	(0.52)	(4.59)	(4.02)	(2.90)	
	37.56-38.04	13.40-17.70	8.07-11.83	5.45-8.15	
Experiment 3	27.73	13.20	7.40	4.53	4.19
	(1.03)	(3.10)	(2.82)	(2.48)	(2.76)
	27.16-28.31	11.48-14.92	5.84-8.96	3.16-5.90	2.72-5.66
Number Accurate Fine-Grain Items					
Experiment 1	19.25	9.00	7.80	6.00	5.04
	(3.47)	(4.20)	(2.68)	(2.72)	(3.35)
	17.79-20.71	7.23-10.77	6.69-8.91	4.85-7.15	3.66-6.42
Experiment 2	20.30	10.90	7.25	5.30	-
	(4.04)	(3.78)	(4.01)	(2.39)	
	18.41-22.19	9.13-12.67	5.37-9.13	4.18-6.42	
Experiment 3:	11.47	7.53	4.13	2.93	3.06
	(2.10)	(2.50)	(2.13)	(1.71)	(2.11)
	10.30-12.63	6.15-8.92	2.95-5.31	1.99-3.88	1.94-4.19

Table 3

*One-Way ANOVAs on Proportion of Accurate Fine-Grain Items Reported in Experiments 1-3*

Experiment*	Statistic*			
	<i>F</i>	<i>df</i>	<i>p</i>	<i>f</i>
1	19.08	3, 94	<.001	0.78
2	7.25	2, 57	.002	0.50
3	9.14	3, 55	<.001	0.71

\*In Experiments 1 and 3, forced report compared with free recall, free recall plus confidence and adapted SAI.© In Experiment 2, forced report compared with free recall plus confidence and adapted SAI.©

Table 4

*Experiment 1. Mean number of accurate coarse-grain items, and mean (& SD) proportion correct for all responses at and above each level of confidence for the three immediate interview conditions which recorded confidence.*

	Forced Report:		Adapted SAI <sup>©</sup>		Free Recall (Confidence)	
	Initial					
	N <sub>A</sub>	P <sub>A</sub>	N <sub>A</sub>	P <sub>A</sub>	N <sub>A</sub>	P <sub>A</sub>
100	5.08	0.90	0.08	0.40	0.75	0.77
		(0.14)		(0.55)		(0.36)
≥ 90	8.79	0.86	0.12	0.43	0.88	0.80
		(0.13)		(0.53)		(0.34)
≥ 80	12.75	0.78	0.36	0.64	0.96	0.78
		(0.14)		(0.50)		(0.33)
≥ 70	15.08	0.75	0.56	0.70	1.04	0.72
		(0.11)		(0.46)		(0.38)
≥ 60	17.13	0.70	0.56	0.70	1.08	0.67
		(0.11)		(0.46)		(0.42)
≥ 50	20.21	0.69	0.60	0.67	1.08	0.67
		(0.10)		(0.45)		(0.42)
≥ 40	21.08	0.68	0.60	0.65	1.08	0.67
		(0.10)		(0.44)		(0.42)
≥ 30	22.79	0.68	0.60	0.61	1.08	0.67
		(0.09)		(0.45)		(0.42)
≥ 20	23.96	0.67	0.60	0.60	1.08	0.67
		(0.09)		(0.45)		(0.42)
≥ 10	25.42	0.66	0.60	0.60	1.08	0.67
		(0.10)		(0.45)		(0.42)
≥ 0	27.29	0.65	0.60	0.60	1.08	0.67
		(0.09)		(0.45)		(0.42)

N<sub>A</sub> = Number accurate (Mean), P<sub>A</sub> = Proportion accurate (Mean and SD)

Table 5

*Experiment 3. Mean number of accurate coarse-grain items, and mean (& SD) proportion correct for all responses at and above each level of confidence for the three immediate interview conditions which recorded confidence.*

	Forced Report:		Adapted SAI <sup>©</sup>		Free Recall (Confidence)	
	Initial		N <sub>A</sub>	P <sub>A</sub>	N <sub>A</sub>	P <sub>A</sub>
100	5.80	0.71	0.40	0.80	0.33	0.80
		(0.16)		(0.27)		(0.45)
≥ 90	8.20	0.59	0.60	0.79	0.33	0.80
		(0.21)		(0.27)		(0.45)
≥ 80	10.47	0.60	0.67	0.69	0.33	0.80
		(0.09)		(0.26)		(0.45)
≥ 70	11.67	0.61	0.67	0.69	0.33	0.80
		(0.12)		(0.26)		(0.45)
≥ 60	12.60	0.61	0.73	0.71	0.33	0.80
		(0.10)		(0.25)		(0.45)
≥ 50	14.13	0.62	0.73	0.71	0.33	0.80
		(0.10)		(0.25)		(0.45)
≥ 40	14.53	0.62	0.73	0.71	0.33	0.80
		(0.10)		(0.25)		(0.45)
≥ 30	14.87	0.61	0.73	0.71	0.33	0.80
		(0.11)		(0.25)		(0.45)
≥ 20	15.53	0.61	0.73	0.71	0.33	0.80
		(0.10)		(0.25)		(0.45)
≥ 10	15.67	0.60	0.73	0.71	0.33	0.80
		(0.10)		(0.25)		(0.45)
≥ 0	16.27	0.61	0.73	0.71	0.33	0.80
		(0.09)		(0.25)		(0.45)

N<sub>A</sub> = Number accurate (Mean), P<sub>A</sub> = Proportion accurate (Mean and SD)

Table 6

*Experiment 1. Mean number of accurate fine-grain items, and mean (& SD) proportion correct for all responses at and above each level of confidence for the three immediate interview conditions which recorded confidence.*

	Forced Report:		Adapted SAI <sup>©</sup>		Free Recall (Confidence)	
	Initial		N <sub>A</sub>	P <sub>A</sub>	N <sub>A</sub>	P <sub>A</sub>
100	2.75	0.81	4.28	0.81	3.17	0.93
		(0.28)		(0.20)		(0.18)
≥ 90	5.13	0.75	5.28	0.77	4.46	0.90
		(0.26)		(0.19)		(0.17)
≥ 80	7.08	0.65	6.20	0.78	5.13	0.85
		(0.20)		(0.17)		(0.18)
≥ 70	8.79	0.60	6.80	0.76	5.54	0.82
		(0.14)		(0.16)		(0.18)
≥ 60	10.50	0.59	7.28	0.75	5.88	0.80
		(0.12)		(0.14)		(0.18)
≥ 50	13.13	0.53	7.44	0.74	5.96	0.79
		(0.11)		(0.14)		(0.17)
≥ 40	14.38	0.52	7.56	0.74	5.96	0.79
		(0.10)		(0.14)		(0.17)
≥ 30	15.38	0.50	7.64	0.74	5.96	0.79
		(0.10)		(0.13)		(0.17)
≥ 20	16.58	0.48	7.68	0.73	5.96	0.79
		(0.10)		(0.13)		(0.17)
≥ 10	17.54	0.47	7.68	0.73	5.96	0.79
		(0.09)		(0.13)		(0.17)
≥ 0	19.25	0.46	7.68	0.73	5.96	0.79
		(0.08)		(0.13)		(0.17)

N<sub>A</sub> = Number accurate (Mean), P<sub>A</sub> = Proportion accurate (Mean and SD)

Table 7

*Experiment 2. Mean number of accurate fine-grain items, and mean (& SD) proportion correct for all responses at and above each level of confidence for the three immediate interview conditions which recorded confidence.*

	Forced Report:		Adapted SAI <sup>©</sup>		Free Recall (Confidence)	
	Initial		N <sub>A</sub>	P <sub>A</sub>	N <sub>A</sub>	P <sub>A</sub>
100	3.35	0.94	3.55	0.89	2.65	0.91
		(0.11)		(0.15)		(0.25)
≥ 90	5.10	0.85	4.85	0.83	3.20	0.85
		(0.14)		(0.26)		(0.26)
≥ 80	7.80	0.81	5.55	0.82	4.10	0.80
		(0.14)		(0.24)		(0.25)
≥ 70	10.40	0.77	6.30	0.75	4.80	0.80
		(0.16)		(0.24)		(0.25)
≥ 60	12.85	0.71	6.65	0.73	5.00	0.80
		(0.11)		(0.25)		(0.24)
≥ 50	14.65	0.67	6.85	0.71	5.10	0.79
		(0.10)		(0.26)		(0.24)
≥ 40	15.90	0.62	7.15	0.70	5.25	0.79
		(0.11)		(0.25)		(0.24)
≥ 30	16.80	0.58	7.15	0.70	5.25	0.79
		(0.12)		(0.25)		(0.24)
≥ 20	17.75	0.57	7.25	0.69	5.30	0.79
		(0.11)		(0.25)		(0.24)
≥ 10	19.05	0.55	7.25	0.69	5.30	0.79
		(0.10)		(0.25)		(0.24)
≥ 0	20.30	0.54	7.25	0.69	5.30	0.79
		(0.10)		(0.25)		(0.24)

N<sub>A</sub> = Number accurate (Mean), P<sub>A</sub> = Proportion accurate (Mean and SD)

Table 8

*Experiment 3. Mean number of accurate fine-grain items, and mean (& SD) proportion correct for all responses at and above each level of confidence for the three immediate interview conditions which recorded confidence.*

	Forced Report:		Adapted SAI <sup>©</sup>		Free Recall (Confidence)	
	Initial					
	N <sub>A</sub>	P <sub>A</sub>	N <sub>A</sub>	P <sub>A</sub>	N <sub>A</sub>	P <sub>A</sub>
100	4.40	0.55	2.53	0.56	2.40	0.78
		(0.29)		(0.36)		(0.21)
≥ 90	6.53	0.56	3.40	0.57	2.53	0.76
		(0.20)		(0.17)		(0.21)
≥ 80	7.53	0.52	3.73	0.53	2.80	0.72
		(0.14)		(0.16)		(0.22)
≥ 70	8.87	0.52	3.80	0.53	2.93	0.67
		(0.11)		(0.15)		(0.28)
≥ 60	9.53	0.49	4.07	0.54	2.93	0.67
		(0.09)		(0.14)		(0.29)
≥ 50	10.40	0.46	4.07	0.54	2.93	0.67
		(0.10)		(0.14)		(0.29)
≥ 40	10.67	0.45	4.07	0.54	2.93	0.66
		(0.09)		(0.14)		(0.29)
≥ 30	10.93	0.44	4.07	0.54	2.93	0.66
		(0.08)		(0.14)		(0.29)
≥ 20	11.13	0.44	4.07	0.54	2.93	0.66
		(0.08)		(0.14)		(0.29)
≥ 10	11.20	0.42	4.13	0.55	2.93	0.66
		(0.08)		(0.14)		(0.29)
≥ 0	11.47	0.41	4.13	0.54	2.93	0.66
		(0.08)		(0.14)		(0.29)

N<sub>A</sub> = Number accurate (Mean), P<sub>A</sub> = Proportion accurate (Mean and SD)

Table 9

*One-Way ANOVAs on Proportion of Accurate Fine-Grain Items Reported at Each Level of Cumulative Confidence in Experiment 1.*

Confidence	Statistic*			
	<i>F</i>	<i>df</i>	<i>p</i>	<i>f</i>
100	2.18	2, 63	.121	0.26
≥ 90	3.37	2, 69	.040	0.31
≥ 80	7.29	2, 70	.001	0.46
≥ 70	11.79	2, 70	<.001	0.58
≥ 60	13.24	2, 70	<.001	0.62
≥ 50	21.48	2, 70	<.001	0.78
≥ 40	25.49	2, 70	<.001	0.85
≥ 30	31.11	2, 70	<.001	0.94
≥ 20	33.60	2, 70	<.001	0.98
≥ 10	36.77	2, 70	<.001	1.02
≥ 0	40.85	2, 70	<.001	1.08

\*In Experiment 1, forced report compared with free recall, free recall plus confidence and adapted SAI.©



Table 10

*One-Way ANOVAs on Proportion of Accurate Fine-Grain Items Reported at Each Level of Cumulative Confidence in Experiment 2.*

Confidence	Statistic*			
	<i>F</i>	<i>df</i>	<i>p</i>	<i>f</i>
100	0.36	2, 50	.702	0.12
≥ 90	0.08	2, 56	.923	0.05
≥ 80	0.05	2, 57	.954	0.04
≥ 70	0.26	2, 57	.772	0.10
≥ 60	1.13	2, 57	.331	0.20
≥ 50	1.83	2, 57	.169	0.25
≥ 40	3.20	2, 57	.048	0.34
≥ 30	4.55	2, 57	.015	0.40
≥ 20	5.55	2, 57	.006	0.44
≥ 10	6.76	2, 57	.002	0.49
≥ 0	7.25	2, 57	.002	0.50

\*In Experiment 2, forced report compared with free recall plus confidence and adapted SAI.©

Table 11

*One-Way ANOVAs on Proportion of Accurate Fine-Grain Items Reported at Each Level of Cumulative Confidence in Experiment 3.*

Confidence	Statistic*			
	<i>F</i>	<i>df</i>	<i>p</i>	<i>f</i>
100	2.39	2, 38	.105	0.35
≥ 90	4.66	2, 40	.015	0.48
≥ 80	5.92	2, 41	.006	0.54
≥ 70	3.10	2, 42	.056	0.38
≥ 60	3.28	2, 42	.048	0.40
≥ 50	4.50	2, 42	.017	0.46
≥ 40	4.09	2, 42	.024	0.44
≥ 30	4.67	2, 42	.015	0.47
≥ 20	5.31	2, 42	.009	0.50
≥ 10	5.72	2, 42	.006	0.52
≥ 0	6.06	2, 42	.005	0.54

\*In Experiment 3 , forced report compared with free recall, free recall plus confidence and adapted SAI.©