Running Head: THE DO NOT GUESS INSTRUCTION

# The Do Not Guess Instruction and the Quantity-Accuracy Tradeoff

Franziska Wolf

Student number: i6048213

Master's Programme Psychology & Law

Faculty of Psychology & Neuroscience
Maastricht University

Supervisors:

Dr. Robert Horselenberg
Prof. dr. Marko Jelicic

Word count: 11,029

**Abstract**

The current study sought to fill a gap in the investigative interviewing research literature, namely that of the *do not guess* (DNG) warning often included in cognitive interviews. This is the only instruction addressing the quality of reporting, and makes an appeal to witnesses for employing a minimum threshold for volunteering details in their testimony. To test the impact of this instruction on the quantity-accuracy tradeoff interviewees make in their reports, three variants were compared in an online self-administered interview (SAI) format: a control group without a DNG instruction, and two experimental conditions with a simple- and an explicit DNG instruction, respectively. It was hypothesized that (1) quantity would decrease- and quality would increase with explicitness of the DNG instruction, (2) that memory trust would be positively related to quantity and quality, and (3) that confidence-ratings about witnesses' own testimonies would be greatest for the explicit DNG condition. The data did not lend support for the first two hypotheses, and only partial support for the third – the explicit DNG group was more confident of their testimonies' completeness than the control group. Overall, witness accounts were very accurate irrespective of presence- and explicitness of DNG warnings, and despite self-reported memory distrust. The inclusion of the DNG instruction is not fully defensible on the basis of the current study, but future research is crucial in ascertaining these preliminary findings.

*Keywords*: cognitive interview, witness testimony, do not guess, quantity-accuracy tradeoff

## Table of Contents

# The Do Not Guess Instruction and the Quantity-Accuracy Tradeoff

Throughout the past decades, the field of investigative interviewing has undergone major changes in attitudes toward interviewing in general, and specifically in evidence-based techniques of effective witness-, victim- and suspect interviewing (Snook, Eastwood, Stinson, Tedeschini, & House, 2010; Vrij, 2003). The field has moved from interrogation to forensic interviewing (Boon, Odinot, Horselenberg, & Geijsen, 2016). However, this process of change is not yet completed, as far from all international jurisdictions have successfully integrated the new scientific knowledge into their daily practice. A key development is the use of evidence-based techniques compiled into the so-named cognitive interview (CI) (Fisher & Geiselman, 1992; Memon, Wark, Bull, & Koehnken, 1997), which encompasses social- as well as psychological elements of communicative interaction and witness- or suspect memory. It is a set of mnemonics used in the information-gathering framework called PEACE, entailing the stages *Planning and preparation*, *Engage and explain*, *Account*, *Closure*, and *Evaluation*, for interviews with cooperative victims, witnesses, and perpetrators (Boon et al., 2016; Dando, Wilcock, & Milne, 2008; Snook et al., 2010; Soukara, Bull, Vrij, Turner, & Cherryman, 2009). In addition, a self-administered interview (SAI, e.g. Gabbert, Hope, & Fisher, 2009; Gabbert, Hope, Fisher, & Jamieson, 2012; Hope, Gabbert, & Fisher, 2011) with the aim of facilitating the registering of testimonies when police resources are insufficient, as well as an enhanced and modified version of the CI have been developed (Colomb, Ginet, Wright, Demarchi, & Sadler, 2013).

These described forms derived from the original PEACE framework have received ample attention from the investigative interviewing research community. However, not every jurisdiction has (completely and successfully) adopted these techniques – and in order to optimally highlight the advantages of the new approaches and facilitate the change, it needs to be maximally effective. Further advancements in research and feedback from practitioners enable the continued refinement and condensing of interviewing manuals. The current research seeks to fill a gap in the research literature on the specific effectiveness of the *do not guess* warning in the *free recall* instructions used in the CI.

**The Cognitive Interview (CI)**

In 1992, Fisher and Geiselman published the manual for the cognitive interview (CI) in which they combined a number of memory enhancing techniques and social components for effective interviewing of witnesses and victims. Their work was a response to an analysis that the authors conducted of real police interviews with witnesses and the shortcomings identified within the reports (Fisher, Geiselman, & Raymond, 1987). Issues mainly revolved around interview structure, facilitating recall for the witness, type of questioning and active listening skills. Based on thorough research, multidisciplinary collaborations and extensive work in the field, Fisher and Geiselman (1992) constructed a first set of evidence-based interviewing methods in order to improve police practices. The CI advises an equal-leveled working alliance between interviewer and interviewee, in which rapport is established and the aims and objectives of the interview are clarified (Boon et al., 2016; Collins, Lincoln, & Frank, 2002; Fisher & Geiselman, 1992). As a basis for a good working collaboration, a comfortable atmosphere and trust have to be established in the opening stage of the interview. This includes mutual attention and respect, irrespective of the person or situation. A central feature of the cognitive interview (CI) is the witness-centered approach. That is, there is a transfer of control to the interviewee, and they are given the time to make their account of the incident (Huff, Meade, & Hutchison, 2011; Lipton, 1977; Milne & Bull, 1999). This also entails active listening, and refraining from interrupting the retrieval process.

In order to prime the initial free recall, context reinstatement is used. During this part of the preparation, the scene, lighting, perceptions and feelings surrounding the witnessed event are brought to mind as vividly as possible. This technique has been shown to facilitate the recall process and increase the number of correctly reported details (Hammond, Wagstaff, & Cole, 2006). Moreover, a variety of instructions and established mnemonics are used to maximize retrieval of potentially case-breaking information. These begin with the interviewer asking the witness to report everything, in as much detail as possible, and to also report details that are not recalled with absolute certainty. Giving the witness time and space to give their initial account guided purely by their own recollection – in addition to the techniques for subsequent questioning – reduces the potential social influences that the interviewer may – unknowingly – assert on the interviewee (Brunel, Py, & Launay, 2013). Furthermore, it increases accuracy relative to formats with reduced *report options* (Goldsmith & Koriat, 1999). This entails that the witness the freedom to

choose what they report or withhold, in contrast to forced-answer questions or those that already contain answer options (Paulo, Albuquerque, & Bull, 2016a).

After the initial account is completed, the interviewer may use a variety of strategies to further cue the witness. *Change perspective* instructions encourage the interviewee to imagine what they would have seen, heard and felt if they were in the position of the victim or another witness, for example. This is based on the idea that recollections can be accessed in numerous different ways, by means of associative networks (Milne & Bull, 2002). Similarly, the *change temporal order* instruction stimulates the witness to recall the event in reverse chronological order, or starting at specific moments of the event. Moreover, the interviewer is advised to make use of self-generated cues when questioning the interviewee, meaning that the witnesses' own words are used as much as possible. This, in connection with refraining from interruptions and letting the interview be guided by the retrieval process of the witness, is referred to as *witness-compatible questioning* (Dando, Wilcock, Behnkle, & Milne, 2011).

These interviewing models have been quite successful in their translation into standard police practice especially in England and Wales (Dando et al., 2008; Dando, Wilcock, Milne, & Henry, 2009). Other jurisdictions, however – such as those of the United States (Fisher & Geiselman, 2010) and the Netherlands (Boon et al., 2016) – are still in the process of dissemination and implementation. Furthermore, there have been indications that certain aspects of the CI are perceived by investigating officers as being too time consuming or not particularly effective (Dando et al., 2008). The line of work that addressed these issues revealed that the *change perspective* and *change temporal order* instructions – in which the witness is asked to recall the event from a variety of different perspectives and in reverse sequence – failed to increase recall and were reportedly used least frequently by police officers (Dando et al., 2008; Memon et al., 1997). Thus, while law enforcement has begun to adopt evidence-based interviewing techniques, the effectiveness and utility in the real-world setting are still being explored.

### The Self-Administered Interview (SAI)

Upon reporting of an incident to the police, it is crucial to obtain information from available witnesses as soon as possible. For one, this is due to the heavy reliance of police investigations on information from those involved in order to establish such factors as criminal intent, responsibility, psychological state of the perpetrator, provocation and the

exact sequence of events (Gabbert et al., 2009). However, police resources are not always sufficient to ensure effective and immediate use of available information resources. An initial early recall opportunity offers valuable protection against forgetting and misinformation effects among witnesses (Gabbert et al., 2009; Hope et al., 2011), and against skewing of witness memory through media exposure (Davis & Loftus, 2007). Thus, in order to safeguard valuable information from witnesses, it is crucial not to delay such an early recall opportunity. With the aim of providing police with a witness-reporting tool that could be used when resources are insufficient or not immediately available, a self-administered interview was created (SAI; Gabbert et al., 2009).

Self-administration of a witness interview offers the benefit of using one's own words in giving an account of events, rather than the usual practice of police writing up the statement in their words and with a possibly selective focus (Hope et al., 2011). Although the SAI is a more impersonal form of obtaining statements from witnesses and victims – and thus likely less suitable for vulnerable or perhaps traumatized individuals (Gawrylowicz, Memon, & Scoboria, 2014) – written and online versions of the modified CI (MCI) could offer the opportunity to overcome language barriers stemming from differing nationalities or illiteracy. In the event that a witness in question has limited or no local language skills, an initial account can still be obtained when in written or even recorded form. Non-readers/writers could also be given the chance to give their account of an incident by developing an online version of the SAI that is accessible to analphabets or those with visual impairments. An audio-format online SAI could be developed in which such witnesses could record their responses.

As mentioned above, the SAI is rather impersonal and lacks one of the key components from which the CI and its modified versions derive their effectiveness: building rapport (Gabbert et al., 2012; Gawrylowicz et al., 2014). Although it has been demonstrated that the SAI is superior to only a *free recall*, and that the amount of correctly reported details can level with the classic CI, not everyone may find the paper-and-pencil interview or an online version appropriate. Personal contact serves to 'catch' the witness or victim emotionally after the incident, make them feel comfortable and attempt to take away any fears or anxiety. The interviewer may ask how they are feeling, show empathy, and offer the witness something to drink – helping to establish a safe working relationship. When the established rapport is being maintained throughout the interview by showing each other respect through attention and active listening skills, trust is established. This serves the cooperation and open communication between the two parties (Collins et al.,

2002). Nevertheless, even if the SAI removes this component of social interaction, quantity and accuracy outcomes are comparable to the CI. Thus, with the aim of isolating the *do not guess* instruction under standardized conditions, the SAI was chosen for the current study. It will also provide an insight into the level of reporting that can be elicited in such an online testing environment.

**The Do Not Guess (DNG) Instruction**

Though ample research attention has been directed toward the newer versions of the CI and its effectiveness in police practice, there are a number of questions that remain. Recent studies in the field have asked diverse question, including what the influence of motivation is on memory monitoring (Paulo et al., 2016a), whether interview ground rules should be repeated (Brubacher, Poole, & Dickinson, 2015), and how a new mnemonic called category clustering recall compares to current techniques (Paulo, Albuquerque, & Bull, 2016b). Yet, there is an instruction that is being used in such research that has not undergone closer scrutiny: the *do not guess* (DNG) instruction (Brubacher et al., 2015; Colomb et al., 2013; Dando et al., 2008; Fisher, Brewer, & Mitchell, 2009). In order to claim that CI interviewing manuals are evidence-based, it is necessary to underpin each element of the instructions with research that validates its effectiveness (Boon et al., 2016). To our knowledge, there exist no studies that directly address the effect of this instruction.

Research that does mention the DNG component of the CI is largely focused on children, targeting their tendency to want to provide an answer regardless of whether they properly understood the question or have actual recollections (Brubacher et al., 2015). Considering that the DNG instruction is the only one pertaining to the quality of the testimony rather than the quantity (Paulo et al., 2016a), it is essential that we acquire an understanding of how the DNG warning influences accuracy- and error rates. In addition, it is relevant to measure whether increases in accuracy go hand-in-hand with a tradeoff in the quantity of details reported.

In the absence of ill intentions, one would expect that guessing is likely to occur when witnesses are unsure of their recollections, or simply do not know an answer. Scoboria and Fisico (2013) studied the effect of encouraging, discouraging or not mentioning don’t know (DK) responses by witnesses on the quantity-accuracy tradeoff in testimonies. Underlying this instruction is a similar notion to that of the DNG warning, the interviewer aims to discourage the witness from speculating, thereby safeguarding

accuracy. It was found that those participants who had been encouraged to answer 'I don't know' to questions and clarify these statements, answered fewer questions and thereby made fewer errors. Thus, the quantity of answers was traded for higher accuracy. These results suggest that it is possible to influence mock witnesses' perceptions of how to safeguard their reports from slip-ups and mistakes. Besides, they also helped to unveil the implicit assumption among witnesses that there is an expectation for informative responses. The control- and discouraged groups, namely, performed at comparable levels for quantity and accuracy, but below the DK-encouraged group.

Fisher and Geiselman (1992) started out with no warning of the like, and even told witnesses not to edit their thoughts – while also recognizing that this may be interpreted as a "license to fabricate" (p.148). Evidently, the difficulty lies in striking the right balance in order not to tilt the reporting criterion too much and thereby decrease the number of potentially correct and valuable details reported. The quantity-accuracy tradeoff captures the two basic elements of memory, and reflects the metacognitive- or metamnemonic control that the witness is exercising (Goldsmith & Koriat, 1999), as well as what they truly recall. Particularly within the CI and related interviewing techniques, the witness is given *report options* thanks to the *free recall* and open questions. This *witness-centered* approach brings with it the increased responsibility of editing, meaning that the individual has to choose some kind of strategy for volunteering information.

Depending on their own motivations and standards, witnesses may choose to remain rather vague in an effort to avoid making mistakes, or they may choose to only report what they are very certain of. Including a DNG warning acknowledges that a police interviewing setting is, at its worst, an unnerving and uncomfortable situation which can impose great pressure on the victim or witness. In a panicky effort to provide as much detail as possible, and with the likely motivation to want to help in catching the perpetrator, witnesses may be very compliant and engage less in self-regulatory processes when volunteering information. In the current study, the DNG instruction acted as a reminder to exercise metamnemonic control amidst efforts to provide as much information as possible. In addition, the current study is a contribution in the ongoing attempt to establishing an evidence-based manual that is as compact and effective as possible in order to facilitate the use of the CI in practice.

**Memory Distrust and the Do Not Guess Instruction**

The Oxford dictionary defines guessing as "estimat[ing] or conclud[ing] (something) without sufficient information to be sure of being correct", whereas confabulation concerns "fabricate[d] imaginary experiences as compensation for memory" ("Oxford Dictionary of English," 2010). Thus, guessing occurs in a context of uncertainty, whereas confabulation occurs when there is a lack of memory and involves an element of purpose – essentially one is 'making things up' to fill gaps in one's recall. Thus, on a conceptual level, the DNG instruction could be thought of as working along a continuum. The clarity of recollections and the level of confidence we have in them range from a complete memory gap – which one may be tempted to fill by confabulating – to a flashbulb memory that is remembered extremely vividly and with great confidence, though necessarily accurately (Hirst et al., 2015; Talarico & Rubin, 2003). When performing a memory search for what we do remember – in-between the two ends of this hypothetical spectrum – we engage in some form of metacognitive evaluation to judge how certain we are of any particular item of information that comes up. Guessing tends to be associated with a mid-point confidence level of 50% (Granhag, Jonsson, & Allwood, 2004).

It has been indicated by Van Bergen, Jelicic, and Merckelbach (2009) that memory distrust comes with increased compliance. If this is true for accepting external misinformation, there is possible that one may observe a similar deficit in the judgment of self-generated, false mental images that may be constructed based on schemata. Along a similar line of thought, memory efficacy was found to be inversely related to suggestibility, thus that increased memory efficacy lowers effects of misinformation and social influence (Liebman et al., 2002). These findings about the role of memory distrust and memory efficacy in accepting or rejecting pieces of information are also reflective of the kind of recall threshold that interviewees are setting. If those with low memory distrust and -efficacy are more vulnerable to accepting misinformation, they are likely to be setting a lower confidence criterion for volunteering information. Other sources have suggested that the threshold set may operate at an earlier stage as well, namely in what comes to mind in the first place (Rhodes & Jacoby, 2007). This is referred to as early selection, whereas the adjustment of the criterion used to report a recollection is called late correction (Goldsmith & 2016). It is argued that stimulating the witness to engage in such metacognitive monitoring of their recall process could work to increase the accuracy of statements given. What is important to consider, however, is the quantity-accuracy tradeoff

(Evans & Fisher, 2011; Hollins & Weber, 2016) that may occur as a result of these selection processes. Increasing the criterion for volunteering an answer would naturally lead to a decrease in details reported, while accuracy would be expected to increase.

By instructing witnesses not to guess, the aim is to decrease errors that can arise through source misattribution, reliance on the availability heuristic, and the use of general knowledge in the form of schemas to fill gaps in memory (Lipton, 1977; Rhodes & Jacoby, 2007). Furthermore, this instruction is likely to limit early selection somewhat, since it encourages the witness to think only about what was actually present, discouraging from too much 'fishing' in memory when something is not remembered. It is possible that the interviewee tries to recall a certain part of the incident, and may be confused about the in- or external source of an image that comes to mind. The associative nature of our memory creates the opportunity for schema-related images to appear as though they are part of the recollection, resulting in confabulation (Hyman & Pentland, 1996; Lipton, 1977; Rhodes & Jacoby, 2007). With the explicit DNG instruction, the current study sets out to make a first attempt at establishing the unique contribution of this part of the *free recall* instruction of the SAI.

**Report Errors and Self-Judgments**

Underlying investigative interviewing, one might say, is the opposite standard to the one held in our judicial system – innocent until proven guilty – because we seek to elicit as much information as possible, with the idea that all details are potentially suspicious. They may not prove useful once the lead is followed up – but ideally one aims to obtain all possible leads in order to solve the case. At the same time, the police are keen to avoid following leads that were based on incorrect details. This sets interviewers up for a difficult task: engaging and directing the witness in such a way that maximum quantity and maximum accuracy is achieved (Paulo, Albuquerque, Saraiva, & Bull, 2015). But where is it that interviewees need to be guided or even limited in their reporting?

When we look to our own memory experience, we can observe that recollections vary in terms of vividness. On the one end of the continuum lies what we call flashbulb memory, while at the other extreme we remember nothing – perhaps due to inattention and thus a lack of encoding, incomplete consolidation or mere forgetting (Tourangeau, 2000). In the grey area lie vague, incomplete recollections that are prone to being completed by general knowledge, beliefs about what one witnessed, and expectations based on schemata.

It is in the middle of this spectrum where the DNG instruction aims to influence the criterion for reporting, as this is where witnesses tend to make errors in their testimonies. Paulo and colleagues (2016a) describe this criterion-targeted instruction as allowing the witness *report options* and encouraging them to adjust their *report precision*. This means that the witness is asked to monitor their recollections and refrain from editing them or filling in gaps – and instead, say '*I don't know/remember'*. While this may indeed lower the amount of guessed-at information in testimonies, the authors argue that such metacognitive monitoring operates along a continuum between a full guess and a completely confident response. This is illustrated by the frequent use of indicators of decreased confidence, such as '*I think*', '*I believe*', and '*maybe'*. These are in themselves reflections of the metamnemonic control being exercised - but to what extent does witness confidence in their report performance indicate accurate recall?

Research on the confidence-accuracy relationship have established a modest correlation at best (Odinot, Wolters, & Van Koppen, 2008; Van Koppen, 2007). However, such studies are mainly based on recognition memory and closed questions, as the CI interviewing setting does not allow for interruptions to assess confidence levels (Granhag et al., 2004). Self-ratings usually follow a recall moment in order not to interrupt this process, and are thus retrospective in nature. In addition, specific units of information from the testimony need to be selected for these confidence-ratings. Such a procedure is likely to result in relatively high confidence-ratings due to reiteration effects (i.e. an increased belief in an item of information that is repeated) and previous metacognitive output-control by the witnesses. The latter was supported in a study by Allwood, Ask, and Granhag (2005), who investigated the realism of witnesses' confidence judgments two weeks after a CI or SI. High accuracy and low overconfidence was attributed to the *report options* allowed by said interviewing techniques. This means that with the CI instructions, witnesses tend to be very accurate and confident in their reports already, so such retrospective confidence-ratings may only give us limited insight into the upper end of the confidence spectrum. Granhag, Jonsson and Allwood (2004) used forced-choice questions about a stimulus event and respective confidence ratings, and found that mock witnesses' significantly underestimate their performance. Similarly, error estimates were found not predictive of actual error rates, with witnesses overestimating their own testimony's error content (Paulo et al., 2015).

These findings suggest that our intuitive feelings about the accuracy of our reports are poorly calibrated with actual error rates – demonstrating that metamemory judgment is

a difficult task. Other studies, however, have found well-calibrated self-evaluations of memory (Allwood et al., 2005; Van Bergen et al., 2009). Thus, this is yet another question that warrants further examination, with the aim of providing a potential additional measure of witness report accuracy. In real-world settings the ground truth of a case is unknown, making indications of witness reliability vital for efficient investigation by police.

## The Current Study

The aim of the current study was to establish whether the DNG instruction impacts the quantity and accuracy of details retrieved by mock witnesses in the SAI. The following research question was addressed: What effect does the *do not guess* instruction of the cognitive interview have on the quality and quantity of eyewitness memory in an online SAI? In addition to this focal topic, we explored the role of memory distrust between-subjects using the Dutch version of the Squire Subjective Memory Questionnaire (SSMQ, translated and validated by Van Bergen, 2008), analyzed confidence-ratings for misleading and non-misleading questions on the stimulus event, and asked participants for self-reflections on their statements.

Towards an isolation of the DNG instruction and its effects on the quantity-accuracy tradeoff, three conditions were used: a control group without a DNG warning, and two experimental manipulations. The status quo of police practice was mirrored by the experimental condition with a simple DNG instruction (that is: '*Do not leave out any details, but* do not guess *details you cannot remember.*'). A second manipulation with an explicit DNG component was included to investigate whether additional explanation and emphasis on this instruction would produce any further effects on quantity and accuracy in the statements ('*Do not leave out any details, but* do not guess *details you cannot remember. If you cannot truly recall something, do not fill in the gaps or make up any details – write down solely what you yourself have perceived.*'). This was based on concerns expressed in previous research that such an instruction may need to be very clearly stated (Gabbert et al., 2009).

In order to tempt participants to guess in a more direct fashion, a list of confidence-rating items about the presence of details within the stimulus event was constructed. The questions were either leading, e.g. '*Did you see the parking machine?*' or non-leading, e.g. '*Did you see a red car?*'. Both sets of questions contained true items to act as an additional measure of accuracy, and false items in order to tempt guessing behaviour. Moreover, self-

rating scales on qualities of the testimony given and on participants' own reporting behaviour. These items aim to give an additional insight into the relationship between subjective evaluations of testimonies and self-professed guessing, confabulating, and withholding. The set of measures described are put together to make a first attempt at a controlled online study isolating the DNG instruction. The mixed methods approach allows an objective- as well a subjective angle toward the mechanisms and effects at work. The following hypotheses were made within the current study:

Hypothesis 1a) The quantity of reported details will decrease with the explicitness of the DNG instruction.

1b) The quality (accuracy) of reports will increase with explicitness of the DNG instruction.

Hypothesis 2a) Participants with higher SSMQ scores (high memory trust) will report a higher quantity of details than those with lower SSMQ scores (low memory trust, i.e. distrust).

2b) Participants with higher SSMQ scores will give more accurate testimonies than those with lower SSMQ scores.

Hypothesis 3) Retrospective confidence-ratings should be higher for the explicit DNG condition than for the control and simple DNG groups.

In addition, answers to the leading questions will be analyzed in an exploratory fashion, due to a lack of literature on the DNG instruction in general, and in relation to suggestibility and compliance specifically. In terms of memory distrust, it would be expected that individuals with low SSMQ total scores (i.e. high memory distrust) would display a tendency to avoid answering the items (through selection of the *don't know* option) and to be more compliant with the questions. The latter is based on the finding that memory distrust is related to increased compliance (Van Bergen et al., 2009). By investigating the variables above, the current study contributes a first look at the utility of the DNG component in a SAI, in pursuit of teaching a maximally effective interviewing manual to police officers and all other information-gathering interviewers.

## Method

### Design

The study employed a between-subjects quantitative experimental design with a three-leveled manipulation of the independent variable: the DNG instruction is left out in the control condition, while the manipulations will reflect the status quo (Simple DNG), and a more elaborate version of the instruction (Explicit DNG). The first experimental condition is thus a simple request for the witness to never guess, while the second experimental condition entails additional instructions to avoid filling in memory gaps, to refrain from confabulation and report solely what was actually perceived.

Comparisons between-subjects were made, using a 3 X 3 design, (High/Medium/Low SSMQ score) X (Control/Simple DNG/Explicit DNG). The dependent variable of testimony quality was coded based on accuracy (proportion of correct details over total details), quantity (total number of details), and errors (proportion of incorrect details over total details). Assignment to experimental groups was randomized before SAI instructions were given, controlling for Dutch- and German speakers to ensure relatively even distribution across the conditions.

### Participants and Sampling

Given the scope of the current study, and based on similar previous experiments, we aimed for 20 participants per condition (Gabbert et al., 2009). Recruitment was carried out online in June 2017 via Maastricht University's online research study participation platform, in addition to in-person recruitment in the city centre, posters in supermarkets and social media advertisements. This strategy was aimed at collecting a more diverse, representative sample of the general population. Inclusion criteria were an age of 18 years and above, and being a native speaker of Dutch or German.

A total of 98 people began the survey. Due to some limitations with Qualtrics (malfunctioning with certain Internet browsers and when opened on mobile devices) and the external Pac-man task, 17 participants did not reach part 2 of the survey, while having completed part one. With additional drop-out mainly at the very start of the survey (either no entered responses or only basic demographics), a final sample of 66 participants completed the entire study. Thanks to response validation within the Qualtrics tool, no missing values were encountered. This resulted in the following sample numbers per

condition: control group ($N$ = 24), simple DNG ($N$ = 24), and explicit DNG ($N$ = 18). The majority of participants were female (78.79%) and currently studying (89.4%), with ages ranging from 18 to 55 and a mean of 22.41 ($SD$ = 4.67) years old. The sample consisted of 39 (59.01%) native Dutch- and 27 (40.91%) native German speakers. The conditions did not significantly differ with respect to gender [$\chi^2(2) = 0.51$, $p = .773$], age [$F(2,63) = 0.85$, $p = .433$], education [$\chi^2(4) = 4.09$, $p = .394$] or occupation [$\chi^2(6) = 4.21$, $p = .649$].

The total required time investment required of each participant was approximately 45 minutes (as mentioned in the advertisements and validated by actual average response times). There was no time limit for the survey, although a request was included to complete part 2 in one session. Participants were compensated by either one research participation credit (only applicable to Maastricht University's bachelor students of psychology), or a € 7.50 voucher. Approval of the ethics committee of the Faculty of Psychology and Neuroscience at Maastricht University was obtained on April 20th 2017, valid until the end of the academic year, August 31st 2017.

## Materials

The survey tool Qualtrics was used to put together the online study. It contained an information letter, declaration of consent, basic demographic information items, the Squire Subjective Memory Questionnaire, instructions for viewing the video, and the stimulus event itself.

**Squire subjective memory questionnaire (SSMQ).** A validated Dutch version of the SSMQ was used (Van Bergen, 2008). The 18-item questionnaire aims to measure how much trust individuals have in their own memory functioning. Open-ended statements such as '*My ability to recall things that happened a long time ago is…*' and '*My ability to pay attention to what goes on around me is…*' are rated on a scale of -4 (*disastrous*) to 4 (*perfect*), and are summed up for a total score ranging between -72 to 72. Van Bergen (2008) established satisfactory internal consistency with Cronbach's α ranging between .87 - .96 between three studies with student and normal population samples ($N_1$ = 293, $N_2$ = 70, $N_3$ = 128). The questionnaire discriminates well between individuals who are distrusting of their memory and those with good memory trust. Moreover, the scale correlates well with similar measures, such as the Cognitive Failures Questionnaire by Broadbent, Cooper, FitzGerald, and Parkes (1982). A test-retest stability of $r$ = 0.86 and

0.90 was established across two studies ($N_1$ = 38 & $N_2$ = 70), and concurrent construct validity was found to be satisfactory. The excellent internal consistency was confirmed in the current study, with Cronbach's α = .91.

The questionnaire was translated into German by the author, and reverse-translated back to Dutch by an independent bilingual speaker. The versions were compared in order to check for equal meaning, and the validity of German items was confirmed. Internal consistency of the German scale was excellent, with Cronbach's α = .92. This version can be found in Appendix 1.

**Stimulus event.** In order to contextualize the video acting as the stimulus event, participants were given the scenario of watching a movie at home while waiting for a friend to arrive, and suddenly hearing a loud noise from outside. The video was framed to be the scene unfolding when opening of the front door to have a look (duration 2:30 minutes). Although there was no direct criminal act visible, participants observed three men walking across a parking lot, peeking into cars and attempting to open them. A number of passers-by crossed the screen, and after some unsuccessful attempts by the suspects, the owners of the car they focused their break-in attempts on return and the three men run away. Participants were told in part 2 that a car theft did occur briefly after this scene, and were asked to give their witness statement to help solving the crime.

**Filler task.** To build in a small buffer period between witnessing and providing a statement, a five minute Pac-man game was embedded as a Qualtrics-external link before part 2 of the survey. Participants could play the game or wait while the count-down timer kept going, after which an automatic link-through to part 2 occurred.

**SAI instructions.** In the second part, participants were asked to imagine that a car theft had occurred soon after the scene they had observed in the video clip in part 1. In order to add a personal element and simulate some of the pressures of acting as a witness, we included a video of the author with a brief personal introduction, expression of thanks for their participation and an emphasis on that it is very important for them to read the following instructions carefully and to do their best when filling in the form. A motivation check on a percentage rating-scale was included between the video and SAI instructions. Instructions were sourced and slightly adapted from the Dutch police's SAI ('Zelfrapportage voor Getuigen'), and from studies with the (modified) CI (Hammond et al., 2006; Milne & Bull, 2003; Paulo et al., 2016a). They were translated into German, and checked by an independent trilingual speaker. The full set of instructions can be found in Appendix 2. The manipulated DNG component varied as follows:

*Control condition*: No DNG instruction in the control condition,

*Simple DNG condition*: 'Do not leave out any details, but **do not guess** details you cannot remember.'

*Explicit DNG condition*: 'Do not leave out any details, but **do not guess** details you cannot remember. If you cannot truly recall something, **do not fill in the gaps or make up any details** – write down **solely what you yourself have perceived**.'

Key components were displayed in bold to ensure these were the most salient in the text, and the answering field for witness reports was displayed directly beneath the instructions.

**Leading questions.** After giving their statement, participants were probed with 16 leading questions on the presence of specific details in the stimulus event, some of which using the direct article '*the'* and others the milder version '*a/an*'. Ten questions indicated correct details, and six suggested non-present, false details. Examples include: '*Did you see the parking machine?*' (false) and '*Did you see an open trunk?*' (true).

**Motivation and manipulation check.** Motivation was measured through two items asking participants to rate the level of current motivation they felt before-, and the motivation they felt retrospectively during their testimony. A visual analog scale (VAS) ranging from 0 to 100% was used. A manipulation check following the second motivation item involved an open-ended question on the suspected purpose of the study.

**Self-evaluation items.** To contrast the observed performance of mock witnesses with their own subjective perceptions, a number of self-rating items were included, again using 0 to 100% VAS. These covered judgments of completeness and accuracy of their own testimony, in addition to items on estimated error content, self-professed confabulation, guessing and withholding. The latter three served to obtain an insight into how mock witnesses regulated their own reporting behaviour. A note was added asking participants to answer truthfully, and that no consequences were attached to their responses.

## Procedure

**Part 1.** Respondents to the recruitment rounds were directed to the Qualtrics survey, and first selected their native language (German or Dutch), in order to address the likelihood that through speaking a foreign language the extent of the recollections will not

be accurately reflected (Dockrell, 2004). After the information letter on the first page followed a declaration of consent retrieved from Maastricht University, with an e-signature field. Basic demographics of gender, age, education and employment were requested. Next, the SSMQ was presented. By positioning this questionnaire at the start of the experiment, we were aiming to minimize the influence of the given answer on the witness report, since it may bias people towards their own performance – and to avoid that perceptions about the given testimony would skew the answers. Then followed the instructions for the stimulus event, in which participants were asked to imagine that they are waiting at home for a friend, and the video clip as representing the scene unfolding as they step outside following a loud noise. They were not explicitly warning them about a memory test (Evans & Fisher, 2011), but told that there would be questions on the content later. Lastly, participants were asked not to discuss the experiment or video material (e.g. when speaking to friends who may have taken part as well) in order to prevent additional consolidation through rehearsal and potential co-witness effects. A link-through to a five-minute Pac-man game with a timer followed, and participants were automatically redirected to the second part once the time was up.

**Part 2.**For the second part of the survey, participants read an introduction, asking them to imagine a car theft had taken place briefly after the witnessed event. They were now asked to act as a key witness in filling in an online witness self-report form, and provide a signature for their cooperation for added realism and perceived legitimacy. The Qualtrics randomization tool allocated the mock witnesses to one of three conditions, in which they each first received a personal introduction message from the author in video-form. Next, the condition-specific instructions followed and participants gave their testimonies. No time limit was provided, and participants then moved on to the next page with case-specific questions (leading and non-leading). After the SAI, a manipulation check asking participants to guess at the underlying aims of the study followed. The open-answer responses were coded according to whether they identified elements such as the relevance of instructions or testing of witness memory. Then a second motivation check followed, in order to compare initial motivation with retrospective motivation after participants had actually written their accounts. After that, a number of self-reflection questions with percentage rating scales were to be filled in.

Lastly, participants were given the option to comment on the study and to select their type of compensation. A full debriefing and researcher contact details for any

questions were provided on the last page.

**Scoring of Witness Testimonies**

For the evaluation of witness reports, scoring was carried out by the author based on the commonly used categories of setting, people, actions, and objects (e.g. Dando et al., 2011; Milne & Bull, 2002). Herein, the total numbers of these details were of interest – completeness based on an exhaustive list of all details of the stimulus event was not addressed within the current study. Each detail was counted only once when summing the scores, and any errors were noted. Confabulations were not coded separately, as these are not considered necessarily representative of guessing behaviour, and the focus of this preliminary study on the DNG instruction lay on the relationship between quantity and quality more generally. In addition, an initial review of the witness reports revealed that references to the various actors were often unclear and thus this category would be very subjective in coding. Any statements reflecting opinions, for instance '*They looked suspicious'*, were disregarded.

In order to provide a control for objectivity of the coding procedure, an independent, trilingual scorer who was unfamiliar with the study's aims and objectives was provided with the stimulus event video and explained the coding procedure. She then coded a random sample of ten witness reports (15.15%) and any discrepancies were resolved by discussion between the scorers before the first author scored the remainder of reports. This approach was based on similar techniques used by other researchers in this area (Dando et al., 2011; Evans & Fisher, 2011). Inter-rater reliability for all four detail categories, in addition to errors, was computed through intra-class correlation coefficients (ICC) in the Statistics Package for Social Sciences (SPSS) software, version 20. These ICCs were satisfactory, ranging between .963 and .995 ($M$ = .984).

The numbers of correct- and incorrect units of information reported were computed as proportions of the total number of units. An independent samples $t$-test was conducted to compare the experimental and control groups on these measures. Furthermore, the answers to the manipulation-check were evaluated by exploring if any of the participants guessed at the correct underlying aims. Confidence ratings for overall quality,

completeness and accuracy (errors made) of the testimonies were entered and calibrated with actual total accuracy rates to provide an additional test of the accuracy-confidence relationship (Odinot & Wolters, 2006; Van Koppen, 2007). Moreover, SSMQ scores were categorized into three levels (low, medium, high) and were compared within-subjects in order to explore how memory distrust and the quantity-accuracy tradeoff within reports relate.

## Results

The study set out to investigate how the DNG instruction impacts the quantity-accuracy tradeoff in witness testimonies in an online SAI. Before any analyses will be made to test the hypotheses, the manipulation- motivation checks will be addressed. Then follows the investigation of any quantity-accuracy tradeoffs, after which the role of memory distrust will be scoped out in further comparative testing. The correspondence between the self-evaluations and actual witness report parameters will be explored, and compliance to leading- and non-leading questions assessed. All of the following sections' analyses were carried out in SPSS, using an alpha level of .05 to determine significance.

### Manipulation- and Motivation Check

Out of 66 participants, only four had the suspicion that the aims had something to do with the reporting format (online/type of instructions/way of reporting). Around half ($n$ = 27) of participants addressed the memory test component, and 15 recognized that accuracy was a variable of interest. Furthermore, eleven individuals indicated that they believed the Pac-man task to have been merely a distractor task or time buffer. Six people mentioned the measuring of subjective views of one's memory, and five showed awareness that the specific questions were partially meant to be misleading. Based on these findings, no one was excluded from analysis as the specific aim of the study was not guessed at by any participant.

The motivation of participants just before giving their testimonies and immediately after completion of the full SAI was measured to rule out this variable as a confounding factor in the quality and quantity of witness reports. The ratings ranged between 28% and 100% for the initial motivation check, and between 31% and 100% for the second one. There was a no significant difference in mean motivation levels at either measurement

time, with $F(2, 63) = 2.08$, $p = .134$ for initial-, and $F(2, 63) = 2.71$, $p = .074$ for post-report motivation. Thus, any comparative results cannot be attributed to differing levels of motivation between the conditions.

**Quantity-Accuracy Tradeoff**

It was hypothesized that the quantity of reported details would decrease- and that the quality (accuracy) would increase with the explicitness of the DNG instruction (Hypothesis 1a & 1b). In terms of a simple word count comparison of quantity, the three conditions did not significantly differ from one another, with an overall range from 30 to 369 words and average of 196.79 words. The total number of details reported varied between 11 and 82, with each condition displaying a high accuracy – reflected by proportions of correct details above 92%, displayed in Table 1. One error score of 15 was considered an outlier as it was more than three standard deviations removed. Witness accounts did not differ in terms of reported details per category; all testimonies were mainly person- and action-focused, and contained a relatively small number of object details, and fewest for the surroundings. Irrespective of the presence and explicitness of the DNG instruction, mock witnesses made few errors in their testimonies, with a maximum of 15, and overall average of just 2.80 errors. One-way ANOVAs were run to compare the various measures between groups (all assumptions of equal variances were met), but none yielded significant results. Thus, there is no evidence for an effect of the DNG instruction.

To analyze the testimony data in terms of the quantity-accuracy tradeoff, two-tailed Pearson correlations were run for each group using the variables total details and proportion correct. No significant correlation was found for the control group [$r(24) = .321$, $p = .210$] or the explicit DNG group [$r(18) = -.179$, $p = .478$]. For the simple DNG condition, a significant, negative correlation was found [$r(24) = -.480$, $p = .018$]. Thus, for this condition only, a tradeoff was observed: as the quantity of reported details increased, the proportion of correctly reported details decreased.

**Table 1**

*Witness Report Descriptives per Condition and for Total Sample*

| | SAI Conditions | | | | |
|---|---|---|---|---|---|
| | Control | Simple DNG | Explicit DNG | One-way ANOVA | |
| Variable | *M (SD)* | *M (SD)* | *M (SD)* | *F* (2, 63) | *p* |
| Word count | 183.71 (89.82) | 216.46 (78.65) | 188.00 (85.55) | 1.01 | .370 |
| Details | | | | | |
| *Surroundings* | 1.33 (1.09) | 1.88 (2.15) | 2.06 (3.33) | 0.60 | .550 |
| *People* | 16.79 (10.58) | 18.50 (7.85) | 20.94 (12.33) | 0.85 | .432 |
| *Actions* | 15.12 (8.19) | 16.96 (6.98) | 13.56 (5.58) | 1.20 | .309 |
| *Objects* | 5.33 (2.24) | 5.42 (2.70) | 6.00 (2.74) | 0.30 | .673 |
| ***Total*** | 41.71 (18.53) | 47.38 (16.49) | 46.56 (17.98) | 0.70 | .499 |
| Proportion (%) | | | | | |
| *correct* | 92.23 (6.51) | 94.50 (5.21) | 93.98 (4.60) | 1.08 | .346 |
| *errors* | 7.77 (6.51) | 5.50 (5.21) | 6.02 (4.60) | 1.08 | .346 |

**Memory Distrust**

The SSMQ total scores were summed and compared between conditions. The current sample, comprised mainly of students and with $M = 22.91$ ($SD = 15.57$) overall, was very close to the student norm established by Van Bergen (2008), with $M = 22.59$ ($SD = 14.79$). The control group had the lowest average SSMQ, with $M = 19.75$ ($SD = 16.23$). The simple DNG followed with $M = 23.54$ ($SD = 16.16$), and the explicit DNG group scored the highest, with $M = 26.28$ ($SD = 13.80$). A one-way ANOVA was run to compare these values, but no significant difference was found ($F(2,63) = 0.93$, $p = .399$).

For further analyses taking into account the level of memory distrust, the sample was split into low-, medium- and high-scoring groups, based on the 25$^{th}$ and 75$^{th}$ percentiles. In the current data, these cut-offs lay at the scores of 12.0 and 33.5, respectively. The sample as a whole displayed a normal distribution of these levels, with 25.76% falling into the low memory trust (i.e. high memory distrust) group, exactly half into the medium category, and 24.24% into the high memory trust group (i.e. low memory

distrust). The distribution of SSMQ levels across conditions is displayed in Table 2.

It was expected that more pronounced memory distrust (i.e. indicated by low SSMQ scores) would go along with decreased quantity of reported details and quality of testimonies (Hypotheses 2a and 2b). Two one-way ANOVAs were run to test this hypothesis, with the assumption of homogeneity of variances being met in both cases. The first ANOVA was set up with the three memory distrust levels as the dependent variables, and quantity of reported details as the independent variable. No evidence was found for a relationship between increased memory distrust and quantity of details in the reports, with $F(2, 63) = 0.00$, $p = .998$. For the second ANOVA, the proportion of correct details was entered as the independent variable. The expected trend between SSMQ level and quality was not supported by the data either, with $F(2, 63) = 2.01$, $p = .142$. Thus, the hypothesis that there is a positive relationship between increasing memory trust and the quantity and quality of reported details was not supported by the current data.

**Table 2**

*Percentages of SSMQ Levels per SAI Condition*

| | SAI conditions | | |
|---|---|---|---|
| Variable | Control (%) | Simple DNG (%) | Explicit DNG (%) |
| SSMQ levels | | | |
| *Low* | 29.17 | 33.33 | 11.11 |
| *Medium* | 50.00 | 41.67 | 61.11 |
| *High* | 20.83 | 25.00 | 27.78 |

**Leading Questions**

The responses to the leading questions were coded as 1 for correct (true) or 2 incorrect (false). *Don't know* (DK) responses were coded as missing values. Table 3 gives an overview of the cell counts per condition. To test for any differences in response behaviour between the three conditions, a cross-tabulation was set up for running Chi square tests. The only significant result was for the item on the perpetrators wearing sunglasses, but due to the violation of the assumption of minimum cell counts of five (44.4% of cells did not meet this criterion), we cannot regard this as a reliable indication. Thus, there were no significant differences in true, false, or *don't know* responses across the conditions.

Carrying out the same procedure with SSMQ levels, all but one item were nonsignificant (with an acceptable 22.2% of cells with a count less than five). This was the question on the peripheral detail of having seen *'the black man'* walking behind the cars [$X^2(4, N = 66) = 10.01$, $p = .040$]. Thus, overall, no clear or consistent trends were observed on the leading questions.

**Table 3**

*Number of True, False and Don't Know (DK) Responses to Leading Questions per SAI Condition*

| | SAI conditions | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Control | | | Simple DNG | | | Explicit DNG | | |
| Item category & question | T | F | DK | T | F | DK | T | F | DK |
| True items | | | | | | | | | |
| *Did you see a bicycle come by?* | 23 | 1 | 0 | 20 | 4 | 0 | 17 | 1 | 6 |
| *Did you hear a woman shouting?* | 19 | 4 | 1 | 21 | 1 | 2 | 14 | 2 | 2 |
| *Did you see a small, silver car?* | 20 | 3 | 1 | 20 | 3 | 1 | 17 | 0 | 1 |
| *Did you see an open trunk?* | 13 | 10 | 1 | 11 | 11 | 2 | 10 | 6 | 2 |
| *Did the suspects wear sunglasses?* | 16 | 5 | 3 | 11 | 7 | 6 | 16 | 1 | 1 |
| *Did you see a red car?* | 19 | 4 | 1 | 19 | 4 | 1 | 15 | 3 | 0 |
| *Did any of the suspects wear a leather jacket?* | 21 | 0 | 3 | 20 | 2 | 2 | 15 | 2 | 1 |
| *Did you see that someone was trying to open a car door?* | 22 | 2 | 0 | 24 | 0 | 0 | 17 | 1 | 0 |
| *Did you see the child?* | 22 | 2 | 0 | 22 | 1 | 1 | 18 | 0 | 0 |
| *Did you see the break-in?* | 10 | 14 | 0 | 11 | 10 | 3 | 11 | 6 | 1 |
| False items | | | | | | | | | |
| *Did you see a suitcase?* | 15 | 3 | 6 | 16 | 1 | 7 | 12 | 2 | 4 |
| *Did you see someone drive by on a motorcycle?* | 18 | 1 | 5 | 19 | 2 | 3 | 17 | 0 | 1 |
| *Did you see the black Ford?* | 11 | 5 | 8 | 14 | 4 | 6 | 10 | 3 | 5 |
| *Did you see the parking machine?* | 13 | 3 | 8 | 18 | 1 | 5 | 16 | 0 | 2 |
| *Did you see the black man walking by behind the cars?* | 12 | 8 | 4 | 15 | 5 | 4 | 8 | 8 | 2 |
| *Did you see the metal rod?* | 16 | 5 | 3 | 18 | 2 | 4 | 14 | 2 | 2 |

**Self-evaluations**

A screening for outliers was carried out based on *z*-score conversions of the raw data for the self-evaluation items, and three extreme scores were recoded as missing values based on a cut-off at three standard deviations. One participant had indicated 0% confidence in their testimony's quality, which went along with low motivation (10% prior to- and 20% post-testimony) and below-average report accuracy of 78%. Two other participants rated their self-professed confabulation at high levels of 56% and 65%, which went along with similarly high levels of self-professed guessing (64% and 69%, respectively). In addition to these outliers identified by *z*-score standardization of the data, there were a few notable observations that could reflect differing interpretations of certain self-reflection items. Three respondents replied to the question of '*Did you withhold any details that came to mind?*' with '100%', though these individuals had correct detail totals of 11, 34 and even 76. Table 4 presents an overview of average responses, standard deviations and ranges after correction for outliers.

**Table 4**

*Self-Evaluation Descriptives per SAI Condition*

| | SAI conditions | | | | | |
|---|---|---|---|---|---|---|
| | Control | | Simple DNG | | Explicit DNG | |
| Variable (%) | *M (SD)* | Range | *M (SD)* | Range | *M (SD)* | Range |
| Self-rated | | | | | | |
| *completeness* | 58.88 (18.43) | 13 – 98 | 65.92 (17.20) | 20 – 93 | 73.50 (13.77) | 39 – 94 |
| *accuracy* | 66.58 (18.57)* | 35 – 98 | 71.39 (20.03) | 29 – 93 | 75.39 (9.41) | 51 – 94 |
| *error content* | 36.88 (21.84) | 3 – 71 | 29.96 (13.67) | 8 – 60 | 30.39 (16.75) | 9 – 71 |
| *confabulation* | 6.23 (12.16)* | 0 – 42 | 2.50 (6.32) | 0 – 24 | 5.28 (8.15) | 0 – 26 |
| *guessing* | 21.46 (24.14) | 0 – 72 | 14.71 (19.85) | 0 – 67 | 16.17 (20.89) | 0 – 70 |
| *withholding* | 21.92 (30.95) | 0 – 100 | 19.04 (27.50) | 0 – 100 | 13.50 (25.98) | 0 – 100 |

* = outliers above +/- three standard deviations present, and removed from analysis.

One-way ANOVA's were run to test for differences between the conditions. Since the variables of self-rated accuracy, error content and confabulation violated the assumption of equal variances, the Welch's *F* statistic was used. After this adjustment, the only significant difference found was for confidence in completeness of own testimonies by the control- and explicit DNG condition. The control group had a mean confidence of

58.88%, whereas the explicit DNG group had a mean of 73.50%, see Table 4. This partially confirms hypothesis 3, reflecting that the explicit DNG instruction raised the confidence criterion set by witnesses in their reports.

Moreover, participants drastically overestimated their error percentages, with a total sample mean estimate of 32.59%, as compared to the actual mean error rate of 6.45% across conditions. Pearson's *r* confirmed the lack of a correlation between these measures, with $r(64) = -.146$, $p = .246$. For the quality (accuracy) item, the correlation between self-ratings and observed accuracy was $r(64) = .175$, $p = .164$, two-tailed.

To explore whether memory distrust affected the confidence ratings of mock witnesses, further correlational tests were carried out. It would have been intuitively logical for memory distrust scores to correlate negatively with confidence-ratings, but this was neither the case for the quantity-, nor the quality-related measures. Confabulation, guessing and withholding items also returned non-significant correlations with SSMQ scores.

## Discussion

With the aim of contributing to the improvement of the CI, this study set out to investigate the only instruction specifically targeting the self-monitored quality of witness reports, namely the *do not guess* instruction. Its impact on the quantity-accuracy tradeoff in mock witness testimonies was tested within an online SAI. A mixed methods approach and multitude of statistical comparisons provide insight into the workings of the instruction, in addition to the relevance of memory distrust, self-evaluations and the nature of accounts obtained with an online SAI.

It was hypothesized that the quantity of reported details would decrease with the explicitness of the DNG instruction, while the quality (accuracy) of reports would increase (Hypotheses 1a and 1b). The data obtained in the current online study did not support such trends. No significant differences in the quantity-accuracy tradeoff were found between groups, irrespective of the presence of the DNG instruction and its level of explicitness. When looking at the measures for report quantity and quality, it seems that the simple DNG instruction yielded the most favorable results on average (highest quantity and highest accuracy), though not at statistically significant levels. It may be, then, that the DNG instruction does not have a major impact on the self-monitoring that witnesses already engage in, and that a more elaborate warning does not add to the effects of a simple one.

In terms of the types of accounts elicited by the online SAI, the current data reflected a general focus by witnesses on person and action details – the latter being in line with previous research (Fisher & Geiselman, 1992). The focus on person details is not surprising, given the relatively large number of actors in the stimulus event. Notable, however, was the generally minimal reporting on the surroundings. Although the setting was indeed a simple car park, mentioning this fact alone seemed to suffice for most participants. The cars themselves were described in greater detail than the surroundings, but these were coded as object details. Depending on the actual importance of surrounding-related details to the police, further emphasis should be placed on this component in the instructions.

Furthermore, participants overestimated their error rates significantly, with an average gap of 25% between self-rated and observed error content. On average, participants across all conditions made only 2.80 mistakes. This finding confirms what was observed in previous research (Paulo et al., 2015), and could reflect some level of pressure-induced uncertainty, given the rather elaborate instructions to the witnesses. Moreover, this doubt displayed by participants may have off-set the correlation between confidence in their own testimonies and actual accuracy rates. In terms of participants' other self-reflections, a relative decrease in self-professed guessing behaviour was observed for both DNG conditions – though again, not significant. Self-reported withholding of details that came to mind showed a relative decrease from the control to the simple- and explicit DNG conditions, which is counter to the expectation that quantity would decrease with explicitness of the DNG instruction. It is also counter to the idea that the reporting threshold increased as a product of the explicit DNG instruction – but any conclusions are very tentative at best. This is due to nonsignificant results and the observation that there may have been an issue with the interpretation of this item, based on three respondents who submitted answers of '100%'. Though under the intended meaning of the item this would have meant not volunteering any information at all (which was far from the case for these respondents), alternative interpretations are conceivable. Rather than speculating on such alternatives, we refer to the section on directions for future research.

Memory trust was expected to show a positive relationship with quantity and quality of witness reports (Hypotheses 2a and 2b). The current sample complied with a previously established norm for student populations (Van Bergen, 2008), but counter to previous research (Van Bergen, Horselenberg, Merckelbach, Jelicic, & Beckers, 2010; Van Bergen et al., 2009), showed no relationship to the quality of testimonies. Furthermore, the exploratory analyses on the leading question items did not seem to confirm the finding that individuals

with memory distrust are more compliant (Van Bergen et al., 2009). Taken together, these results relating to the SSMQ are rather counter-intuitive, but may suggest that the workings of the reporting criteria set by witnesses can overrule underlying self-perceptions of memory. Perhaps there are a number of relatively universal conceptions about what is expected from a witness in their reports, which marginalizes the impact of individual characteristics. As observed from the quantitative measures of the obtained SAI testimonies, all mock witnesses in the current study had high levels of accuracy and made few mistakes. This was also reflected in the counts of correct and incorrect responses to the leading questions that served to tempt witnesses to guess and comply by agreeing to have observed the suggested details. These findings lend support to the notion that irrespective of the presence and explicitness of a DNG instruction, and despite negative self-conceptions about memory, witness reporting operates along the two simple core components of reproducing the events in as much detail as possible, and to be as accurate as possible. While these are part of any CI instruction manual, it is very likely that these components coincide perfectly with what witnesses already believe is expected of them.

Lastly, it was expected that the retrospective confidence-ratings would be higher for the explicit DNG condition than for the control and simple DNG groups (Hypothesis 3), due to the raised recall criterion (Allwood et al., 2005; Roberts & Higham, 2002). This hypothesis was partially supported by our findings, reflected in the significantly greater confidence in testimony completeness exhibited by the explicit DNG group in comparison to the control- but not the simple DNG condition. Such a trend may be indicative of the witnesses in this group indeed having raised their reporting threshold even more than the other two groups, meaning that they only reported what they were virtually certain of. However, this raised reporting threshold was not evident in the observed quantity and quality measures: In practice, the conditions did not differ from one another. This could indicate that participants in the explicit DNG condition *believed* that they were especially thorough, but that in actual fact they still used comparable thresholds to other witnesses. Thus, the current research added to the mixed findings on the confidence-accuracy relationship.

**Limitations**

To next address the shortcomings of the current study, we turn first to the online SAI format. While this has potentially valuable benefits outlined in the introduction, it cannot

yet be considered ecologically valid equivalent of the established paper-and-pencil SAI. Since police authorities have not yet reached the stage of using online, decentralized tools for obtaining witness reports, it is questionable whether this new format had an impact on the results. One problem lies in the lack of control over the surroundings in which witnesses fill out the form, ranging from media device used, screen size and image quality to distractions in the room and temporal features of the testimony process. Although the total required time was as expected, there was substantial variability in the answering times that may be reflective of speeding through the survey or being interrupted in the process of filling it in. It is also possible that the introductory film used to recreate some kind of rapport and enhance motivation was not sufficient to stimulate best-possible performance from witnesses. The average motivation levels, however, suggest that participants were well-motivated to do their best, thus we do not expect this to have been a major factor in the nonsignificance of results.

Another shortcoming is the limited sample size due to the scope of the current study – while sufficient to fill the cells for most statistical analyses, the amount of variation in report-related variables would warrant a much larger sample to find any effects. In addition, some analyses were affected by low cell counts (such as the cross-tabulations for the leading questions) and thus did not yield reliable outcomes. The variability of participant demographics was much lower than hoped for as well. This was due to a common problem in social science research, namely that the student population is easiest to access and most likely to be motivated by the provided incentives (course credit as an external motivator and financial remuneration as an internal/external one).

Lastly, the current study did not address how participants interpreted the instruction in question: What is the common reading of *do not guess*? If – as described above – even an item on withholding can be so differentially interpreted, it is conceivable that there was variability in what *guessing* meant to the participants. This brings us to the recommendations for subsequent research based on the current study.

## Directions for Future Research

In an effort to remedy the shortcomings of the current study and continue to add to the evidence base of the *do not guess* instruction, future research should aim to validate the use of online SAIs with larger sample sizes. In order to scope out the underlying workings of the DNG instruction, it is invaluable to bring to light the interpretations of the different

variants. An interesting addition to such a study would be to address what conceptions individuals from different demographic backgrounds have about what is expected by police from a witness. On a related note, knowing whether or not someone has previous experience with witness reporting is of increased relevance when aiming for a larger age demographic. This would also take into account the finding that experience with witness reporting using the SAI can equip individuals with transferrable skills (Gawrylowicz et al., 2014).

Lastly, it would be advised to attempt replication of the current findings with similar designs but in differing formats, such as a comparison with the original SAI and CI tools. This would shed light on the impact that the online format may have had in the current study, and strengthen the evidence base for including the DNG warning.

## Conclusion

To summarize, the current study seeking to investigate the justifiability of including a *do not guess* warning in the instructions for witness testimonies, three main hypotheses were tested. The first pertained to the expected decreasing quantity and increasing quality of testimonies under increasingly elaborate DNG instructions. This was not confirmed by statistical analyses, while some trends in the data suggested that the most favorable outcome measures are obtained with a simple DNG warning. Second, it was hypothesized that memory distrust would be related to the measures of testimony quantity and quality, in that individuals with low memory trust would report less details with lower accuracy, and those with high memory trust would report more and with increased accuracy. The data did not lend support for this hypothesis – no relations were found between the described measures. Thirdly, it was expected that confidence ratings about own testimony characteristics would be highest for the explicit DNG condition, which was partially supported by the current results. The explicit DNG group showed significantly more confidence in the completeness of their reports than the control condition. This observation was not evident in actual quantity and accuracy rates, however.

Based on these findings, and keeping in mind the limitations of the current study, the suggestion is put forth to seek further insight into the DNG instruction before including it in standard interviewing manuals of the CI kind. The data does not present reason to believe that omitting this instruction until it is satisfactorily supported by research would have any detrimental effect. To the contrary – the current findings seem to suggest that the

DNG instruction does little to alter the high performance of mock witnesses as per status quo without such a warning. We make an appeal to the investigative interviewing research community to complete the evidence-base for all components of the CI, in order to continue the dissemination of interviewing tools that are fully justifiable and defensible.

## References

Allwood, C. M., Ask, K., & Granhag, P. A. (2005). The Cognitive Interview: Effects on the realism in witnesses' confidence in their free recall. *Psychology, Crime & Law, 11*(2), 183-198. doi:10.1080/10683160512331329943

Boon, R., Odinot, G., Horselenberg, R., & Geijsen, K. (2016). Naar een effectieve interviewstandaard voor de politie. *Het Tijdschrift voor de Politie, 4*(16), 20-25.

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British journal of clinical psychology, 21*(1), 1-16.

Brubacher, S. P., Poole, D. A., & Dickinson, J. J. (2015). The use of ground rules in investigative interviews with children: A synthesis and call for research. *Developmental Review, 36*, 15-33. doi:10.1016/j.dr.2015.01.001

Brunel, M., Py, J., & Launay, C. (2013). Cost and benefit of a new instruction for the cognitive interview: the open depth instruction. *Psychology, Crime & Law, 19*(10), 845-863. doi:10.1080/1068316x.2012.684058

Collins, R., Lincoln, R., & Frank, M. G. (2002). The effect of rapport in forensic interviewing. *Psychiatry, Psychology and Law, 9*(1), 69-78. doi:10.1375/pplt.2002.9.1.69

Colomb, C., Ginet, M., Wright, D., Demarchi, S., & Sadler, C. (2013). Back to the Real: Efficacy and Perception of a Modified Cognitive Interview in the Field. *Applied Cognitive Psychology, 27*(5), 574-583. doi:10.1002/acp.2942

Dando, C. J., Wilcock, R., Behnkle, C., & Milne, R. (2011). Modifying the cognitive interview: countenancing forensic application by enhancing practicability. *Psychology, Crime & Law, 17*(6), 491-511. doi:10.1080/10683160903334212

Dando, C. J., Wilcock, R., & Milne, R. (2008). The cognitive interview: Inexperienced police officers' perceptions of their witness/victim interviewing practices. *Legal and Criminological Psychology, 13*, 59-70. doi:10.1348/135532506X162498

Dando, C. J., Wilcock, R., Milne, R., & Henry, L. (2009). A modified cognitive interview procedure for frontline police investigators. *Applied Cognitive Psychology, 23*(5), 698-716. doi:10.1002/acp.1501

Davis, D., & Loftus, E. F. (2007). Internal and external sources of misinformation in adult witness memory. *The handbook of eyewitness psychology, Vol I: Memory for events,* 195-237. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Dockrell, J. E. (2004). How can studies of memory and language enhance the authenticity, validity and reliability of interviews? *British Journal of Learning Disabilities, 32*(4), 161-165. doi:10.1111/j.1468-3156.2004.00315.x

Evans, J. R., & Fisher, R. P. (2011). Eyewitness memory: Balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. *Applied Cognitive Psychology, 25*(3), 501-508. doi:10.1002/acp.1722

Fisher, R. P., Brewer, N., & Mitchell, G. (2009). The relation between consistency and accuracy of eyewitness testimony: Legal versus cognitive explanations. In T. Williamson, R. Bull, & T. Valentine (Eds.), *Handbook of psychology of investigative interviewing: Current developments and future directions* (pp. 121-136). West Sussex: John Wiley & Sons, Ltd.

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

Fisher, R. P., & Geiselman, R. E. (2010). The cognitive interview method of conducting police interviews: Eliciting extensive information and promoting therapeutic jurisprudence. *International journal of law and psychiatry, 33*(5), 321-328.

Fisher, R. P., Geiselman, R. E., & Raymond, D. S. (1987). Critical analysis of police interview techniques. *Journal of Police Science and Administration, 15*(3), 177-185.

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*(4), 298-307. doi:10.1007/s10979-008-9146-8

Gabbert, F., Hope, L., Fisher, R. P., & Jamieson, K. (2012). Protecting against misleading post-event information with a self-administered interview. *Applied Cognitive Psychology, 26*(4), 568-575. doi:10.1002/acp.2828

Gawrylowicz, J., Memon, A., & Scoboria, A. (2014). Equipping witnesses with transferable skills: the Self-Administered Interview©. *Psychology, Crime & Law, 20*(4), 315-325. doi:10.1080/1068316X.2013.777961

Goldsmith, M., & (2016). Metacognitive Quality-Control Processes in Memory Retrieval and Reporting. In L. L. Jacoby, C. M. Kelley, & B. D. McElree (Eds.), *The Oxford Handbook of Memory* (pp. 357-385). New York, NY: Oxford University Press.

Goldsmith, M., & Koriat, A. (1999). The strategic regulation of memory reporting: Mechanisms and performance consequences. *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application*, 373-400.

Granhag, P. A., Jonsson, A.-c., & Allwood, C. M. (2004). The Cogntivie Interview and its Effects on Witnesses' Confidence Psychology, *Crime & Law, 10*(1), 37-52. doi:10.1080/1068316021000030577

Hammond, L., Wagstaff, G. F., & Cole, J. (2006). Facilitating eyewitness memory in adults and children with context reinstatement and focused meditation. *Journal of Investigative Psychology and Offender Profiling, 3*(2), 117-130. doi:10.1002/jip.47

Hirst, W., Phelps, E. A., Meksin, R., Vaidya, C. J., Johnson, M. K., Mitchell, K. J., Lustig, C. (2015). A ten-year follow-up of a study of memory for the attack of September 11, 2001: Flashbulb memories and memories for flashbulb events. *Journal of Experimental Psychology: General, 144*(3), 604.

Hollins, T. J., & Weber, N. (2016). Monitoring and regulation of accuracy in eyewitness memory: Time to get some control. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 171).

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*(2), 211-226.

Huff, M. J., Meade, M. L., & Hutchison, K. A. (2011). Age-related differences in guessing on free and forced recall tests. *Memory, 19*(4), 317-330. doi:10.1080/09658211.2011.568494

Hyman, J. I. E., & Pentland, J. (1996). The Role of Mental Imagery in the Creation of False Childhood Memories. *Journal of Memory and Language, 35*(2), 101-117. doi:10.1006/jmla.1996.0006

Liebman, J. I., McKinley-Pace, M. J., Leonard, A. M., Sheesley, L. A., Gallant, C. L., Renkey, M. E., & Lehman, E. B. (2002). Cognitive and psychosocial correlates of adults' eyewitness accuracy and suggestibility. *Personality and Individual Differences, 33*(1), 49-66. doi:10.1016/S0191-8869(01)00135-0

Lipton, J. P. (1977). On the psychology of eyewitness testimony. *Journal of Applied Psychology, 62*(1), 90.

Memon, A., Wark, L., Bull, R., & Koehnken, G. (1997). *Isolating the effects of the cognitive interview techniques. British Journal of Psychology, 88*(2), 179-197.

Milne, R., & Bull, R. (1999). How do People Remember? *Investigative Interviewing: Psychology and Practice.* West Sussex: Wiley.

Milne, R., & Bull, R. (2002). Back to basics: a componential analysis of the original cognitive interview mnemonics with three age groups. *Applied Cognitive Psychology, 16*(7), 743-753. doi:10.1002/acp.825

Milne, R., & Bull, R. (2003). Does the cognitive interview help children to resist the effects of suggestive questioning? *Legal & Criminological Psychology, 8*(1), 21.

Odinot, G., & Wolters, G. (2006). Repeated recall, retention interval and the accuracy–confidence relation in eyewitness memory. *Applied Cognitive Psychology, 20*(7), 973-985. doi:10.1002/acp.1263

Odinot, G., Wolters, G., & Van Koppen, P. J. (2008). Eyewitness Memory of a Supermarket Robbery: A Case Study of Accuracy and Confidence After 3 Months. *Law and Human Behavior, 33*(6), 506. doi:10.1007/s10979-008-9152-x

*Oxford Dictionary of English* (3rd ed.). (2010). Oxford, UK: Oxford University Press.

Paulo, R. M., Albuquerque, P. B., & Bull, R. (2016a). The enhanced cognitive interview: expressions of uncertainty, motivation and its relation with report accuracy. *Psychology, Crime & Law, 22*(4), 366-381. doi:10.1080/1068316X.2015.1109089

Paulo, R. M., Albuquerque, P. B., & Bull, R. (2016b). Improving the Enhanced Cognitive Interview With a New Interview Strategy: Category Clustering Recall. *Applied Cognitive Psychology, 30*(5), 775-784.

Paulo, R. M., Albuquerque, P. B., Saraiva, M., & Bull, R. (2015). The Enhanced Cognitive Interview: Testing Appropriateness Perception, Memory Capacity and Error Estimate Relation with Report Quality. *Applied Cognitive Psychology, 29*(4), 536-543. doi:10.1002/acp.3132

Rhodes, M. G., & Jacoby, L. L. (2007). Toward Analyzing Cognitive Illusions: Past, Present, and Future. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 379-392). NY: Psychology Press.

Roberts, W. T., & Higham, P. A. (2002). Selecting accurate statements from the cognitive interview using confidence ratings. *Journal of Experimental Psychology: Applied, 8*(1), 33-43. doi:10.1037/1076-898X.8.1.33

Scoboria, A., & Fisico, S. (2013). Encouraging and clarifying "don't know" responses enhances interview quality. *Journal of Experimental Psychology: Applied, 19*(1), 72-82. doi:10.1037/a0032067

Snook, B., Eastwood, J., Stinson, M., Tedeschini, J., & House, J. C. (2010). Reforming Investigative Interviewing in Canada. *Canadian Journal of Criminology and Criminal Justice, 52*(2), 215-229. doi:10.3138/cjccj.52.2.215

Soukara, S., Bull, R., Vrij, A., Turner, M., & Cherryman, J. (2009). What really happens in police interviews of suspects? Tactics and confessions. *Psychology, Crime & Law, 15*(6), 493-506. doi:10.1080/10683160802201827

Talarico, J. M., & Rubin, D. C. (2003). Confidence, Not Consistency, Characterizes Flashbulb Memories. *Psychological Science, 14*(5), 455-461. doi:doi:10.1111/1467-9280.02453

Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. *The science of self-report: Implications for research and practice, 29*, 47.

Van Bergen, S. (2008). Squire Subjective Memory Questionnaire. In T. M. Giesbrecht & C. D. de Ruiter (Eds.), *Forensisch psychodiagnostisch gereedschap* (pp. 137-142). Amsterdam: Harcourt Book Publishers.

Van Bergen, S., Horselenberg, R., Merckelbach, H., Jelicic, M., & Beckers, R. (2010). Memory distrust and acceptance of misinformation. *Applied Cognitive Psychology, 24*(6), 885-896. doi:10.1002/acp.1595

Van Bergen, S., Jelicic, M., & Merckelbach, H. (2009). Are subjective memory problems related to suggestibility, compliance, false memories, and objective memory performance? *The American journal of psychology*, 249-257.

Van Koppen, P. J. (2007). De goede getuige die af en toe faalt [The good witness who fails now and then]. *Tijdschrift voor Criminologie, 49*, 407 - 417.

Vrij, A. (2003). Interviewing Suspects. In A. Memon, A. Vrij, & R. Bull (Eds.), *Psychology and Law: Truthfulness, Accuracy and Credibility* (2 ed.). West Sussex: Wiley.

## Appendix 1

### German Translation of the SSMQ

Fragebogen

Beantworte die folgenden Fragen je mit den folgenden Antwortmöglichkeiten:

-4 = katastrophal
-3 = sehr schlecht
-2 = schlecht
-1 = relativ schlecht
0 = normal
1 = relativ gut
2 = gut
3 = sehr gut
4 = perfekt

1. Meine Fähigkeit, um in meinem Gedächtnis nach Erinnerungen von Namen oder Ereignissen zu suchen, von denen ich weiß, dass ich über sie verfüge, ist...
2. Ich denke, dass meine Familie und Bekannten mein Gedächtnis in diesem Moment einschätzen als...
3. Meine Fähigkeit, um Dinge aus meinem Gedächtnis aufzurufen, wenn ich mich stark darum bemühe, ist...
4. Meine Fähigkeit, mir Dinge die ich gelernt habe zu merken, ist...
5. Wenn ich in einem Monat gefragt werde, dann ist meine Fähigkeit, mich an Dinge von diesem Fragebogen zu erinnern...
6. Meine Fähigkeit, um eine Erinnerung aufzurufen, die mir auf der Zunge liegt, ist...
7. Meine Fähigkeit, um mich an Ereignisse zu erinnern, die vor langer Zeit stattgefunden haben, ist...
8. Meine Fähigkeit, mir Namen und Gesichter von Menschen, die ich treffe zu merken, ist...
9. Meine Fähigkeit, um mich zu erinnern, was ich gemacht habe, nachdem ich ein paar Minuten nicht daran gedacht habe, ist...
10. Meine Fähigkeit, mich an Dinge zu erinnern, die länger als ein Jahr her sind, ist...
11. Meine Fähigkeit, mich zu erinnern, was ich gelesen oder im Fernsehen gesehen habe, ist...
12. Meine Fähigkeit, um mich an Dinge, die in meiner Kindheit passiert sind, zu erinnern, ist...
13. Meine Fähigkeit, mir bewusst zu machen, dass die Dinge, denen ich Aufmerksamkeit widme, auch wirklich in meinem Gedächtnis verankere, ist...

14. Meine Fähigkeit, zu begreifen, was andere Menschen mir erklären, ist...
15. Meine Fähigkeit, um in meinem Gedächtnis zu durchgraben und mich zu erinnern, was ich vor ein paar Minuten gemacht habe, ist...
16. Meine Fähigkeit, meine Aufmerksamkeit auf das zu richten, was um mich herum passiert, ist...
17. Meine allgemeine Aufmerksamkeit für Dinge, die um mich herum passieren, ist..
18. Meine Fähigkeit zu verfolgen, was Menschen sagen, ist...

## Appendix 2

### Instructions for the Online SAI

"It is important that you read the following information carefully and follow the instructions.

In the space below you can write down all the details you are able to recall, such as the people involved, the setting and the actions of the people. Write down the details as you remember them. It's not important to recall and note down details in a different order than they occurred.

Please write up all that you can remember, with **as much detail as possible**. Even details that you deem to be irrelevant or that you're not quite sure of. Each little piece of information could be important for the police investigation.

Feel free to write in full sentences or bullet points. We do not expect that you can recall everything, but please do ensure that your story is as **complete** and **correct** as possible.

(Control condition: no DNG instruction.)

(Simple DNG condition: 'Do not leave out any details, but **do not guess** details you cannot remember.')

(Explicit DNG condition: 'Do not leave out any details, but **do not guess** details you cannot remember. If you cannot truly recall something, **do not fill in the gaps or make up any details** – write down **solely what you yourself have perceived**.')

Stop whenever nothing more comes to mind.

Before you begin with writing down your memories, take a few minutes to fully bring to mind a mental image of the events. **Recall the surroundings, the people and actions, the objects and sounds**. It can help to close your eyes for this part.

Begin when you succeeded in bringing to mind as good a mental image as possible."