Narrative Recall in an Investigative Interview:

Insight into Witness Metacognition

Elizabeth Leonora Fontaine

BSc(Hons), BHlthSc(Hons)

School of Psychology

University of Adelaide

Australia

A thesis submitted in total fulfilment of the requirements for the degree of

Doctor of Philosophy

July 2016

**TABLES OF CONTENTS**

# ABSTRACT

Compared to other interview procedures, the Cognitive Interview produces a larger amount of information without compromising accuracy, and uses techniques that support memory retrieval and socio-communication. Metacognition plays a key role in regulating recall performance but it is unclear how metacognition regulates narrative recall in response to these techniques. Importantly, the grainsize of information elucidates the metacognitive mechanisms regulating recall, yet it is unknown how Cognitive Interview techniques affect narrative grainsize. This thesis examined how these techniques impact narrative performance (quantity, grainsize, and accuracy) and, by applying Koriat and Goldsmith's (1996) framework of metacognition to narrative recall, elucidated the regulatory role of metacognition in the efficacy of the Cognitive Interview.

Experiment 1 tested if the mental-reinstatement-of-context instruction improves monitoring performance, and if the naivety instruction (i.e., the interviewer states their naivety about the witnessed event) encourages the decision to produce more informative testimony. Both instructions produced a greater quantity of information but only the naivety instruction elicited finer-grained accounts. Results suggest that a statement of naivety promotes the decision to give a more informative report, and the mental-reinstatement-of-context instruction reduces the monitoring sensitivity to errors.

Experiment 2 examined the mechanism that may lead a witness to respond to the naivety instruction. Specifically, it was assumed that the witness' decision to report is influenced by their belief in the statement of naivety. When the interviewer made a naivety statement, participants rated their belief in the interviewer's naivety higher and produced more informative reports. Results suggest belief is a necessary state for the efficacy of the naivety instruction. Additionally, Experiment 2 examined if the report-detail instruction also encourages a witness' decision to produce more informative testimony and, importantly, if

this moderates the efficacy of the naivety instruction. Participants produced more informative accounts, and interactions on quantity and grainsize precision, indicate that the report-detail instruction moderates the impact of the naivety statement.

Experiment 3 applied Ackerman and Goldsmith's (2008) dual-criterion model to narrative recall, to examine how the report-detail (informativeness incentive) and do-not-guess (accuracy incentive) instructions impact witness knowledge state. Linguistic qualifiers (e.g., "I think") were also examined for how they communicate recall uncertainty. The study tested if: (a) the report-detail instruction manifests unsatisficing knowledge in more informative, less accurate reports communicated with greater uncertainty (i.e., more linguistic qualifiers); and (b) the do-not-guess instruction manifests conservative satisficing knowledge in less informative, more accurate reports communicated with less uncertainty. The report-detail instruction produced more information (in quantity and finer grainsize) without compromising accuracy or recall uncertainty, suggesting satisficing knowledge is used to give detailed accounts. The do-not-guess instruction produced *more* correct information, suggesting that the instruction enhances monitoring performance.

Across all studies, accuracy was uncompromised when instructions produced more informative reports, suggesting the primary goal in narrative reporting is informativeness and not accuracy. This thesis makes theoretical contributions in applying metacognition theory to narrative recall, and elucidating how component Cognitive Interview techniques impact report informativeness (quantity and grainsize). Findings are useful to practitioners with understanding how different techniques produce informative and accurate testimony.

# DECLARATION

I certify that this work contains no material that has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

_____

Elizabeth Leonora Fontaine

July 2016

# ACKNOWLEDGMENTS

To my supervisors, Dr Carolyn Semmler and Dr Matthew Dry, I am deeply appreciative of your guidance, insightful comments and encouragement throughout my PhD journey. I am particularly grateful for the freedom you extended me to undertake a research project that impassioned me. The journey has been extraordinarily interesting, but it has been the passion for this project that helped me endure some exquisitely challenging times. I am especially thankful for the time and effort you both took to help me improve my writing style.

To Colin Cunningham, Ian Davies, Andriana Tran, and Lachlan Ryan, I am particularly grateful for your assistance with this project. Your involvement in several aspects of the project (interviewing participants, transcribing interviews and coding) has supported the integrity of the research, and provided the much needed help to finish it. To Prof Nicholas Burns, I am sincerely thankful for the access to your laboratory and research facilities so that I could conduct my experiments. To the School of Psychology professional staff, especially Lynda Klopp, Carola Sanders, Angela Allen, and Deidre Simpson, the administrative assistance you provided to this project is gratefully appreciated. To Lynda Klopp, and Karen Roberts from the University of Adelaide Counselling and Disability Service, I am deeply grateful for the care and support you provided for my back injury. Without your assistance, my rehabilitation and the progress of this project would have been significantly hampered. To every person who participated in my experiments, thank you for your involvement because without it, this project would not have come to fruition.

To Amanda Osborn, a very special thanks for your friendship throughout this journey, a journey we shared. To my brother, Sean, your tenacity, courage, and humour in tackling life's challenges, has been (and always will be) a constant source of inspiration. I dedicate this thesis to you. Finally, to my partner, I am tearfully grateful for your encouragement to persevere when times were tough.

**CHAPTER 1**

**Introduction**

Consider the following scenario. Two witnesses are interviewed by an investigator and report on what they observed. The first witness provides the investigator with broad descriptions, comprising 10 details that are all correct. The second witness provides more thorough descriptions, comprising 20 details, but five are wrong. From the investigator's perspective, the goal is to obtain an informative and accurate report because important decisions depend on it. The first witness is less informative, though more accurate, than the second witness. Hypothetically, if you are the investigator and only had one witness to interview, whose report would you prefer? Instead, if you could influence the informativeness and accuracy of witness recall, what techniques would you use to elicit a more informative report from the first witness, or a more accurate report from the second witness?

My thesis was motivated by this last question and with applying Koriat and Goldsmith's (1996) framework of metacognition to narrative recall. Briefly, this framework formulates how metacognitive monitoring and control processes regulate memory reporting. How these processes regulate narrative reporting is less well understood, but this knowledge is needed to develop new interview techniques and understand why current best-practice methods work. My research aims to understand how different interview techniques influence metacognition, and how metacognitive processes mediate narrative recall.

To begin, this chapter overviews the importance of investigative interviewing and the Cognitive Interview protocol for witness interviews (Section 1.1). The necessity to elicit a narrative report with open questions (e.g., "What happened?") is highlighted. The reader is then introduced to Koriat and Goldsmith's (1996) framework of metacognition, and current known strategies that regulate memory reporting (Section 1.2). How these strategies relate to

narrative recall, and why they might be irrelevant to open questioning, is established. I then present the conceptual model that underpins my empirical studies (Section 1.3), an application of Koriat and Goldsmith's framework to narrative recall in the investigative interview setting. The main issue that arises in applying Koriat and Goldsmith's framework to narrative recall is noted, as is the approach that I took to avoid this. Finally, the aims of the research are presented with an outline of the thesis (Section 1.4).

## 1.1    Investigative Interviewing

Investigative interviewing is important across many contexts, for example, policing, national security and workplace safety. Although the outcome of an investigation may differ (e.g., to apprehend a criminal, prevent a terrorist attack or make a workplace safer), the goal for all investigative interviews is the same: to obtain an informative and accurate report of what happened. Critically, how an investigation proceeds, and its successful outcome, rely on decisions that use information contained in the report.

In the policing context, witness testimony plays a central role in justice systems worldwide (Dando, Geiselman, MacLeod, & Griffiths, 2016) and interviews with cooperative witnesses produce the most leads to follow in a criminal investigation (Schollum, 2005). This last point has been demonstrated in the United Kingdom (Kebbell & Milne, 1998) and Australia (Hill & Moston, 2011). Therefore, it is important that witness reports are informative, to provide as many leads as possible. It is equally important that witness reports are accurate so that police resources are not wasted in chasing irrelevant leads. Legal proceedings, subsequent to the criminal investigation, also require informative and accurate reports to provide credible and reliable testimonial evidence.

The success in obtaining an informative and accurate report depends on the interviewer's competency in using appropriate interviewing techniques. Such techniques include: building rapport with the witness, using instructions to assist memory retrieval, using

instructions to encourage informative and accurate responses, and using different communication modes (e.g., verbal, nonverbal; Vrij, Hope, & Fisher, 2014). Of central importance, however, is that the interviewer must always use open questions to elicit a narrative report, and not interrupt the witness during their narration of events (Fisher & Geiselman, 1992; Powell, 2002). All of the techniques noted are used in the Cognitive Interview protocol (Fisher & Geiselman, 1992). This is regarded as one of two[1] interviewing protocols having the "most significant international impact on investigative interviewing practice" (p. 131; Vrij et al., 2014).

### 1.1.1 The Cognitive Interview: evolution of a best-practice interviewing protocol

In 1984, psychological scientists developed the original Cognitive Interview by applying principles of memory to witness recall (Geiselman et al., 1984). The interview protocol comprised four retrieval mnemonics. These are described next, according to the applied memory retrieval principle.

#### *1.1.1.1 Encoding specificity and the mental-reinstatement-of-context mnemonic*

The encoding specificity principle applies when memory traces are activated by matching retrieval conditions to encoding conditions (Tulving & Thomson, 1973). In the context of witness memory, the principle applies, for example, when a witness returns to the scene of a crime and the surrounds trigger recollections of the event. However, it is often impractical to take a witness back to the location where they witnessed the crime. Amongst a range of reasons, this can include it being expensive and time consuming, and may be overly stressful for the witness to physically return to the location.

---

[1] The other is the National Institute of Child Health and Human Development interviewing protocol (NIHCD; Lamb, Orbach, Hershkowitz, Esplin, & Horowitz, 2007). This was developed because children are particularly vulnerable witnesses to interview (Vrij et al., 2014).

Alternatively, in an interview room, recollections may be triggered by instructing the witness to visualise the crime scene; to remember what they could hear, see and smell; and to recall what they were doing, how they felt, and what they were thinking. Combined, these instructions are called the mental-reinstatement-of-context instruction. This instruction operationalises the encoding specificity principle by having the witness mentally reinstate the context of the crime (Fisher & Geiselman, 1992).

The mental-reinstatement-of-context instruction is generally accepted as the most effective of the four mnemonics, receiving good empirical support (e.g., Dando, Wilcock, & Milne, 2009; Dietze, Powell, & Thomson, 2012; Emmett, Clifford, Young, Kandova, & Potton, 2006; Milne & Bull, 1999). However, the efficacy of the mental-reinstatement-of-context instruction may be moderated by individual differences in cognitive style (Emmett, Clifford, & Gwyer, 2003), witness age (Dietze et al., 2012), and how central or relevant the reported information is to the event recalled (Wong & Read, 2011).

### 1.1.1.2 *Multi-componency and the report-everything, change-order, and change-perspective mnemonics*

The multi-component theory of memory posits that memory is an association of multiple traces that can be retrieved in whole, or in part, by one or more probes (Bower, 1967). Three further mnemonics were developed using this theory (Fisher & Geiselman, 1992). The first mnemonic, the report-everything instruction, encourages the witness to report all details that come to mind, no matter how trivial, unimportant or incomplete they may seem. These details may comprise part, or multiple, memory traces that may act as probes to additional recollections of the crime. The second mnemonic, the change-order instruction, guides the witness to report the incident in a different temporal sequence. Commonly, the witness is instructed to start from the end of the incident and recount what happened, backwards in time (Geiselman et al., 1984; Geiselman, Fisher, MacKinnon, & Holland, 1985,

1986; Milne & Bull, 2002). Memory traces activated with this instruction may include those that are overshadowed during forward recall and schema-driven recollections (Geiselman & Callot, 1990; Geiselman, Fisher, Cohen, Holland, & Surtes, 1986). The third mnemonic, the change-perspective instruction, guides the witness to put themselves in the shoes of another bystander, or the suspect, and provide a report from that person's perspective.

The change-order and change-perspective instructions have been found to have little to no effect on recall, and only lengthened the time taken to conduct an interview (Bensi, Nori, Gambetti, & Giusberti, 2011; Boon & Noon, 1994; Dando, Wilcock, Behnkle, & Milne, 2011; Davis, McMahon, & Greenwood, 2005; Mello & Fisher, 1996; Memon, Wark, Bull, & Koehnken, 1997). Pragmatically, the two instructions are also rarely used by police investigators (Clifford & George, 1996). Specifically, in a survey of 96 British police officers, the change-order and change-perspective instructions were perceived as the least useful, and the officers used these instructions least frequently (Kebbell, Milne, & Wagstaff, 1999). Curiously, as far as I have determined from the literature, the efficacy of the report-everything instruction has not been tested in isolation to other Cognitive Interview mnemonics. However, the report-everything instruction is perceived by police officers to be the most effective of the Cognitive Interview methods and they use it most frequently (Kebbell et al., 1999).

### 1.1.1.3 *Enhancing the Cognitive Interview: the importance of open questions*

In the late 1980's, a critical analysis of police interviews found interviewers used methods that were detrimental to eliciting informative and accurate reports (Fisher & Geiselman, 1992; Fisher, Geiselman, Raymond, Jurkevich, & Warhaftig, 1987). The two most common detrimental methods were (a) frequent interruptions, and (b) the excessive use of closed questions to elicit testimony (e.g., "What colour hair did the thief have?"; Fisher, Geiselman, & Raymond, 1987). Of concern, in more recent analyses of police interviews, these detrimental methods are found to prevail (Compo, Gregory, & Fisher, 2012; McLean,

1995; Powell, Cavezza, Hughes-Scholes, & Stoove, 2010; Snook & Keating, 2011; Thoresen, Lønnum, Melinder, & Magnussen, 2009). Interruptions and closed questions are detrimental to witness testimony for the following reasons.

Interrupting a witness discourages an informative report because interruptions interfere with memory retrieval (Fisher & Geiselman, 1992). Specifically, interruptions disrupt attention and disrupted attention may cause activated memories to be forgotten, resulting in less information recalled (e.g., Baddeley, 2001; D. L. Nelson & Goodmon, 2003). Moreover, interruptions condition the witness to give shortened responses in order to provide a response before the next interruption is anticipated (Fisher & Geiselman, 1992). Similarly, closed questions encourage short replies because they elicit specific responses, including "yes"/"no" answers (Fisher, 2010; Ord, Shaw, & Green, 2008). For example, if the interviewer wanted to know what a perpetrator was wearing, a series of closed questions might include: "Was the thief wearing a hat?", "What colour was the hat?", "Was the thief wearing a top?", and "What colour was the top?". Clearly, each of these questions would elicit a short response. Further, closed questioning discourages information being reported that is outside the scope of the questions asked (Fisher, 2010). Therefore, to elicit an informative report with closed questions, the interviewer would need to ask all possible and relevant questions (Fisher & Geiselman, 1992). This however, runs the risk of eliciting an inaccurate report because closed questioning encourage errors (Fisher, 2010), especially when a question leads the witness to respond (e.g., "Was the thief's top blue?"; Loftus, 1975).

Although there is forensic value in asking non-leading closed-questions[2] (N. Brewer, Muller, Nagesh, Hope, & Gabbert, 2010; Hope, Gabbert, Brewer, Tull, & Nagesh, 2010), closed questioning is a cognitively demanding task for the interviewer (Fisher & Geiselman,

---

[2] Both open and closed questions have forensic value, however, the interviewer should take a funnel approach in eliciting a witness report, and only use closed questions to seek additional details and greater clarity about information elicited with open questions (Fisher, 2010).

1992). This is because this style of questioning places the interviewer in control of the flow of information whilst the witness passively responds (Fisher, Geiselman, & Raymond, 1987). Instead, to increase the likelihood of informative and accurate reports, it is critical that the witness controls the flow of information (Fisher, 2010). This is best achieved when the interviewer uses open questions (e.g., "What did the thief look like?"; Fisher, 2010). Importantly, open questions invite narrative responses, more informative than specific responses elicited with closed questions, and open questions encourage a witness to describe only what they know, thereby, motivating accurate responses (i.e., Fisher, 2010). As an aside, open questioning can be considered to implicitly transfer control to the witness. Transfer of control may also be explicitly communicated, for example, when the interviewer makes a naivety statement (Brubacher, Poole, & Dickinson, 2015). A naivety statement tells the witness that the interviewer was not present at the crime and, therefore, does not know what the witness does (Fisher & Geiselman, 1992). I will return to the naivety statement, and why it might elicit informative testimony, in Section 1.3.

### 1.1.1.4   *Efficacy of the Cognitive Interview*

The Cognitive Interview is a very reliable tool for eliciting informative reports (for meta-analytic reviews see Köhnken, Milne, Memon, & Bull, 1999; and Memon, Meissner, & Fraser, 2010). This has been established with the original Cognitive Interview (Geiselman et al., 1984; Geiselman et al., 1985), the enhanced Cognitive Interview (Fisher, Geiselman, Raymond, et al., 1987), and modified versions of the Cognitive Interview – often shortened for frontline police officers who deal with volume crime (e.g., robbery, theft and assault) and time critical situations (Dando, Wilcock, & Milne, 2008; Kebbell et al., 1999). Modifications have included: removal of the change-order and change-perspective mnemonics (Bensi et al., 2011; Boon & Noon, 1994; Dando et al., 2011; Davis et al., 2005; Geiselman & Callot, 1990), replacement of the mental-reinstatement-of-context mnemonic with an instruction to sketch

the crime scene (Dando, Wilcock, & Milne, 2009), development of a self-administered protocol for mass witness situations (e.g., Gabbert, Hope, & Fisher, 2009; Gabbert, Hope, Fisher, & Jamieson, 2012; Gawrylowicz, Memon, & Scoboria, 2014; Hope, Gabbert, & Fisher, 2011), and adding a timeline to assist recall (Hope, Mullis, & Gabbert, 2013).

The Cognitive Interview has been found to be effective across laboratory and field tests, using mock and real witnesses (Clifford & George, 1996; Fisher, Geiselman, & Amador, 1989). It has also been effective across many different cultures, including Australia (e.g., Davis et al., 2005), the United States (e.g., Geiselman, Fisher, MacKinnon, et al., 1986), the United Kingdom (e.g., Dando, Wilcock, Milne, & Henry, 2009), Brazil (Stein & Memon, 2006), France (Py, Ginet, Desperies, & Cathey, 1997), Germany (Aschermann, Mantwill, & Köhnken, 1991), Italy (Bensi et al., 2011), and Spain (Campos & Alonso-quecuty, 1999). Importantly, the Cognitive Interview is effective with vulnerable populations such as children (Verkampt & Ginet, 2010), the intellectually challenged (Clarke, Prescott, & Milne, 2013; Milne, Clare, & Bull, 1999; Milne, Sharman, Powell, & Mead, 2013), and the elderly (Wright & Holliday, 2007). Further, the Cognitive Interview has been shown to increase resistance to the negative effects of suggestibility (Memon, Holley, Wark, Bull, & Köhnken, 1996) and the susceptibility to misinformation effects (Holliday et al., 2012).

Across all studies, the Cognitive Interview has had a strong impact on eliciting a greater quantity of correct details (Cohen's $d = 1.20$; Memon et al., 2010) although a small but significant increase in errors has also been observed (Cohen's $d = 0.24$; Memon et al., 2010). Notably, the increase in errors has not had a detrimental effect on report accuracy (Memon et al., 2010). On the whole, the Cognitive Interview does exceedingly well in helping an investigator achieve their goal to obtain an informative and accurate report.

## 1.2    The Theoretical Framework of Metacognition

Metacognition is defined as the "capacity to monitor and control cognitive activities, such as remembering and perceiving" (Arango-Munoz, 2011, p. 74). Therefore, the act of communicating memories, or reporting behaviour, is a product of memory processes – encoding, storage and retrieval – and metacognition. Memory processes are critical for reporting behaviour, as evidenced by the development of the Cognitive Interview mnemonics, and the importance of open questioning to elicit reports. However, this thesis focusses on metacognition and the mediating role it might play in explaining the efficacy of the Cognitive Interview. Certainly, this does not imply that previous research has not explored this idea already (e.g., Allwood, Ask, & Granhag, 2005; Granhag, Jonsson, & Allwood, 2004; Gwyer & Clifford, 1997; Mello & Fisher, 1996; Roberts & Higham, 2002), and I review the findings from the relevant studies in Chapter 2. However, I took a novel approach to understand the relationship between the efficacy of the Cognitive Interview and metacognition (see Section 1.3.1). Underpinning this approach was Koriat and Goldsmith's (1996) framework of metacognition.

Koriat and Goldsmith's (1996) framework shows how, in response to questioning, metacognitive monitoring and control processes regulate a person's reporting behaviour[3]. Critical to both monitoring and control is the concept of probability. Monitoring involves a probability assessment that an answer retrieved from memory is correct. Control involves a threshold probability that must be exceeded before an answer is reported. Notably, Koriat and Goldsmith's framework formulates how people strategically regulate their reporting behaviour in response to closed questioning. How this formulation relates to open questioning is the foundation of this thesis.

---

[3] Their framework makes no claim about retrieval processes per se but that the best-candidate answer is a joint product of retrieval and monitoring processes (Koriat & Goldsmith, 1996).

### 1.2.1    Strategic regulation of reporting behaviour

When a person is given the option to report[4] or an incentive to reduce the number of errors reported, they can enhance their recall accuracy by regulating their reporting behaviour (Koriat & Goldsmith, 1994, 1996). This behaviour expresses itself in the quantity (Koriat & Goldsmith, 1996) and the level of generality, or grainsize, of information reported (Goldsmith & Koriat, 1999). The mechanism to control these behaviours is the report threshold (Goldsmith, Koriat, & Weinberg-Eliezer, 2002; Koriat & Goldsmith, 1996).

By adjusting the report threshold, information that is less certain to be correct, is withheld or reported at a broader level of generality (i.e., coarse-grained). This has been demonstrated in the domain of general knowledge (e.g., Goldsmith et al., 2002; Koriat & Goldsmith, 1996) and witness memory (e.g., Weber & Brewer, 2008). Across these studies, experimental methodologies have used closed questions to elicit responses that reveal this strategic behaviour. However, as highlighted earlier, open questioning is the preferred method to elicit reports during an investigative interview. How witnesses regulate their narrative reporting behaviour is less well understood, except that open prompts give the witness the freedom (i.e., report option) to select the memories to report (e.g., Buratti, MacLeod, & Allwood, 2014; Evans & Fisher, 2011), and to report only what they know (Fisher, 2010). Therefore, revealing strategic reporting behaviour using report option and accuracy incentives is less relevant to investigative interviewing. Instead, it is pertinent to consider how a witness might regulate their reporting behaviour in response to instructional demands made by the interviewer.

---

[4] Report option is the freedom to withhold an answer to a question. This occurs when a person is told they may say "I don't know" to questions.

### 1.3    A Model of Metacognition for the Investigative Interview Setting

I applied Koriat and Goldsmith's (1996) framework of metacognition to investigative interviewing by constructing a model that (a) explains naturalistic witness behavior, and (b) underpinned my empirical studies. This model is presented in Figure 1.1 and it contains two key elements. The first is the concept of instructional demand – inspired by Koriat and Goldsmith's notion of situational demand (e.g., report option). Instructional demands are made by an interviewer in the form of different instructions. These demands influence narrative recall via long-term memory and metacognition. For example, instructional demands in the form of the Cognitive Interview mnemonics assist memory retrieval (Fisher & Geiselman, 1992). Instructional demands might also affect metacognitive monitoring and control, and it is this aspect of the model that is the focus of this thesis. For example, it has been speculated that a naivety statement (when the interviewer expresses lack of knowledge about the crime) might demand more information by challenging social norms of communication (Milne, 2004), norms that are guided by Grice's (1975) maxims of conversation. This is explored further in the background literature presented in Chapter 2.

The second element to the model is the interviewer's goal to elicit an informative and accurate report. It is at this point that I will define the two dimensions to an 'informative' report. One dimension is the quantity of details that the report contains, and the other dimension is the report grainsize (see Figure 1.1). It is well known that the Cognitive Interview elicits a greater quantity of information (e.g., Memon et al., 2010). However, almost nothing is known about how the Cognitive Interview influences report grainsize[5]. This gap in the literature is addressed by this thesis because report grainsize can provide insight into metacognitive control over reporting behaviour (i.e., Goldsmith et al., 2002).

---

[5] One exception to this is the finding that witnesses, interviewed with the Cognitive Interview after a temporal delay, maintained report accuracy by providing coarser reports (Fisher, 1996). How different Cognitive Interview techniques influence report grainsize is not known.
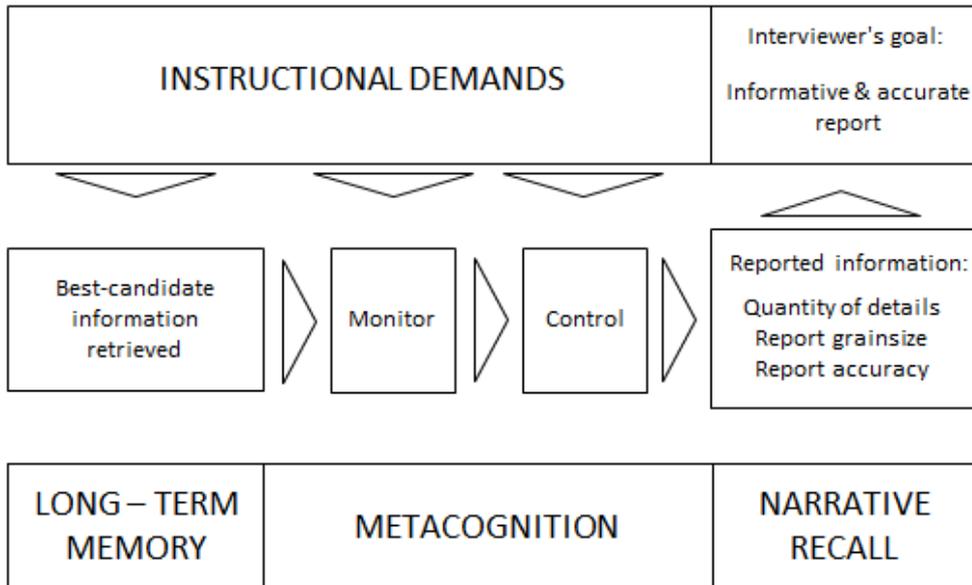
*Figure 1.1.* The metacognition model conceptualised by applying Koriat and Goldsmith's framework of metacognition to narrative recall in the investigative interview setting. Underpinning the model is the use of open questions to elicit a witness narrative report.

### 1.3.1 The methodological approach used in the thesis

The model presented in Figure 1.1 attempts to link metacognition theory, established from closed question methodologies and traditional indices of metacognition (e.g., calibration, discrimination), to naturalistic witness behaviour. It was noted earlier (Section 1.2) that critical to metacognitive monitoring and control is the concept of probability. As such, traditional indices of metacognition require that confidence judgements are made about individual pieces of information (Schraw, 2009). Yet, it is poor practice to interrupt a witness, to elicit confidence judgments about reported details, because interruptions interfere with memory retrieval and diminish report informativeness (Baddeley, 2001; Fisher & Geiselman, 1992; D. L. Nelson & Goodmon, 2003). Therefore, because I used a naturalistic interview paradigm, to situate this thesis in applied psychological science, I took a top-down approach to metacognitive assessment. This was done by mapping narrative performance onto recall outcomes predicted by (a) Koriat and Goldsmith's (1996) framework, and (b) control over

report grainsize (e.g., Goldsmith & Koriat, 2008). Details about these predictions is given in Chapter 2.

## 1.4    Key Aims and Outline of the Thesis

The purpose of this thesis was to observe how instructional demands made with component Cognitive Interview techniques influence narrative recall. An outcome of the thesis is to provide insights into metacognition under narrative recall. To support the research, a key aim was to assess narrative informativeness and accuracy. Informativeness comprises two dimensions that are relevant to narrative recall assessment – the quantity of details reported and report grainsize (see Figure 1.1). Methods, that converge on a generalised formal approach, are common for assessing the quantity of details (and report accuracy) in narrative recall (e.g., Aschermann et al., 1991; Gabbert et al., 2009; Geiselman et al., 1984; Geiselman et al., 1985; Hope et al., 2013; Memon, Holley, Milne, Köhnken, & Bull, 1994). However, there is no similarly generalised approach for assessing the grainsize of narrative recall[6]. Therefore, a key aim was to develop a method to assess narrative grainsize. Moreover, the literature demonstrates how grainsize control regulates reporting behaviour when closed question methodologies are used (e.g., Goldsmith et al., 2002; Weber & Brewer, 2008). Therefore, another key aim was to understand how grainsize control regulates *narrative* reporting behaviour because this behaviour is more relevant to investigative interviewing. Theoretically, this is important to link understanding of metacognitive control over grainsize, determined from traditional methodologies, to naturalistic witness behaviour. Without this link, the full potential for understanding monitoring and control processes cannot be realized.

The motivation for this thesis came from a review of the studies that have investigated the role of metacognition during a Cognitive Interview (Chapter 2). Specifically, this

---

[6] There are two studies (Douglass, Brewer, Semmler, Bustamante, & Hiley, 2013; Evans & Fisher, 2011) that have assessed narrative grainsize but the approach taken in these studies was very different.

motivation grew from the paucity of knowledge related to narrative grainsize, and the limits in applying traditional indices of metacognition to narrative recall when it is elicited with best-practice interview techniques. A series of empirical studies (Experiments 1 to 3) explored the effect that different instructional demands have on narrative informativeness and accuracy. The findings are presented in Chapters 3 to 5 and have theoretical implications for the metacognitive mechanisms involved in mediating narrative recall. The findings extend metacognition theory to naturalistic behaviour and have practical implications for investigative interviewers who elicit witness reports. More broadly, the findings are applicable to investigative interviewing across all contexts (policing, national security, workplace safety, etc.). The thesis concludes with a general discussion and suggestions for future research in Chapter 6.

## 1.5    Summary

The Cognitive Interview is a best-practice protocol that elicits a greater quantity of information from witnesses, across many samples of the population and across many cultures. Further, the Cognitive Interview does this without compromising testimonial accuracy. This is important because the information is likely to provide more investigative leads than information elicited with poor interviewing methods. Although Cognitive Interview techniques support memory retrieval, less is known about how these techniques influence metacognition. One reason for this is the difficulty in applying traditional indices of metacognition during naturalistic behaviour. This thesis aimed to address this by using a top-down approach and, thereby, linking metacognition theory to narrative recall. This was aided by addressing a gap in the literature regarding narrative grainsize. In summary, the quantity of information is only one dimension relevant to the interviewer's goal to obtain an informative report. The other dimension is report grainsize and almost nothing is known about how the Cognitive Interview affects this.

**CHAPTER 2**

**Background Literature on Adult Metacognition in a Cognitive Interview**

In this chapter the reader is given an overview of existing research that has examined how the Cognitive Interview impacts adult metacognition (Section 2.1). Research findings are relevant to understanding monitoring performance, however, I will highlight that the absence of report grainsize analysis poses questions for the validity of the conclusions. Following this, Koriat and Goldsmith's (1996) framework of metacognition is summarised (Section 2.2), to provide the backdrop to explore how different interview techniques might impact metacognition, and why the Cognitive Interview does not produce a quantity-accuracy trade-off (Section 2.3). Specifically, to demonstrate how some techniques might influence monitoring performance and other techniques might influence the decision to report, the reader will be presented with recall outcomes predicted by (a) Koriat and Goldsmith's framework, and (b) grainsize reporting (e.g., Goldsmith et al., 2002). Throughout, the reader will be presented with the gaps in the literature that this thesis will address.

### 2.1    The Cognitive Interview and Witness Metacognition

A limited number of studies have investigated how the Cognitive Interview impacts metacognition (Allwood et al., 2005; Granhag et al., 2004; Gwyer & Clifford, 1997; Mello & Fisher, 1996; Roberts & Higham, 2002) and found little evidence to suggest that metacognition is responsible for the enhanced efficacy of the Cognitive Interview. This conclusion is based on the finding that participants' monitoring performance was similar between cognitive and comparison interviews[7], yet cognitive interviewed participants produced more details (Allwood et al., 2005; Granhag et al., 2004; Gwyer & Clifford, 1997;

---

[7] Comparison interviews were similar to the Cognitive Interview in many ways (e.g., using rapport to help interviewees feel comfortable; not interrupting narrative recall) but differed by omitting the mnemonic instructions (e.g., mental-reinstatement-of-context, report-everything).

Mello & Fisher, 1996). Monitoring was assessed using traditional indices of metacognition (i.e. calibration and discrimination[8]) and whilst a debate exists over the optimal approach to calculate these indices (e.g., Murphy, 1973; T. O. Nelson, 1984, 1986, 1996; Schraw, 1995, 2009; Yaniv, Yates, & Smith, 1991), the commonality is the necessity to elicit confidence judgements in the accuracy of recalled information. However, report grainsize was not examined in these studies. This is a potential shortcoming to the research findings as will be highlighted. Notably, of concern is the generality (or grainsize) of information that an interviewer requests a confidence judgment about. For example, the interviewer might ask a witness to give a confidence judgment about a specific detail (e.g., the colour of a perpetrator's jacket) or about a statement of related details (e.g. the clothing the perpetrator was wearing). In the latter case, the witness might average their confidence in the individual details of the statement, or anchor it to their confidence in a particular detail (e.g., Kahneman & Tversky, 1982). Regardless of how a witness derives their confidence in a statement of details, it is problematic to compare item-specific judgments with judgments for an aggregate of items. This is because item-specific judgments may lead to overconfidence and aggregate judgments may lead to underconfidence, in recall accuracy (see review by Schwarz, 2015). In summary, grainsize has been ignored in the Cognitive Interview literature and the elicitation of confidence for specific items versus aggregate items in memory, has not been accounted for in this literature.

---

[8] Calibration measures how well confidence in recall performance matches recall accuracy, and discrimination measures how well confidence discerns correct from incorrect responses (the reader can find detailed definitions and formulae in Schraw, 2009). Calibration can be expressed as the correspondence (e.g., correlation coefficient) between a set of confidence judgments and performance scores (also known as a relative accuracy index), or as the discrepancy between a confidence judgment and performance score (i.e., judgement precision, also known as an absolute accuracy index). A well calibrated individual will display positive correspondence, and little (to no) discrepancy, between confidence and performance. However, discrimination is considered the more crucial indicator of monitoring performance (although calibration is more commonly reported; Lichtenstein, Fischhoff, & Phillips, 1982; Schraw, 2009) because it is possible that a person can be well calibrated in their performance assessment while having poor discrimination ability (Koriat & Goldsmith, 1996).

One of the key ways that monitoring is assessed is by examining confidence-accuracy calibration. In two studies, the researchers did not find reliable support that participants were well calibrated and this suggested that monitoring is ineffective during a Cognitive Interview (Gwyer & Clifford, 1997; Mello & Fisher, 1996). In the study reported by Mello and Fisher (1996), they cited a small sample size ($N = 10$) as a possible reason for not finding reliable calibration. However, with both studies, the confidence-accuracy relationship might have been unreliable due to grainsize confounds. Gwyer and Clifford (1997) elicited confidence judgments throughout each interview and it is not clear if they controlled for grainsize when probing confidence in the accuracy of information. It is therefore possible that their calibration assessment was unreliable because it was confounded with under- and overconfidence at various grainsize of information. On the other hand, Mellow and Fisher (1996) elicited a global judgment of confidence at the conclusion of each interview. This approach has the potential to produce an unreliable calibration assessment since the probed confidence is an aggregate judgement at the broadest level of grainsize – i.e., *all* recalled details.

In contrast, Roberts and Higham (2002) did find support for effective monitoring during a Cognitive Interview. However, they did not use a comparison interview to determine if this finding was unique to cognitive interviewing or not. Regardless, the researchers used a novel methodology to probe confidence whilst avoiding the need to interrupt recall. During the interview, the interviewer wrote down statements made by the witness. These statements were dissected into smaller information units, and these units were repeated back to the participant (after their interview) to probe their confidence in the accuracy of each unit. As an example, Roberts and Higham provided the following statement and resultant information units (p. 36):

> Statement of information: "The man standing by the tree, smoking a cigarette looking up at the window which had a light on and a person was walking around the room".

> Units of information: (a) man standing by tree, (b) smoking a cigarette, (c) looking up at window, (d) light on in the room, and (e) person walking around the room.

Roberts and Higham (2002) found that their participants were well calibrated, particularly for the statements that they rated with high confidence. Specifically, this confidence-accuracy relationship was restricted to forensically relevant information. This was information deemed relevant (cf. peripheral) to an investigation and/or court proceeding[9]. The findings suggest that a witness' monitoring performance is "fairly good" (p. 40) for forensically relevant information but that it is less effective for peripheral information. However, as noted with the earlier studies (i.e., Gwyer & Clifford, 1997; Mello & Fisher, 1996), the grainsize of information may have confounded the results, especially if forensically relevant information was reported at a different grainsize to peripheral information. To some extent, the researchers may have controlled grainsize because units (of information) were used to probe confidence. It is still possible, however, that the grainsize of these units was mixed. For example, compare the units cited above for (a) and (e). Arguably, "man standing by tree" is finer-grained than "person walking around the room" because the gender of the person is specified in (a) and not (e).

Allwood et al. (2005) also used their participants' own statements, dissected into units of information, to probe confidence in the accuracy of their recall. In contrast to the methodology of Roberts and Higham (2002), the information units were represented to each participant two weeks after their interview. Critically, participants were instructed to report the confidence they presently felt in each unit, and not to think back on the confidence they had experienced two weeks earlier (during the interview). The researchers found that their

---

[9] The researchers used four police officers and a crown counsel to judge if the information would be relevant to an investigation and/or court proceeding.

Cognitive Interview group, although equally calibrated, showed poorer discrimination than their comparison interview group. This suggested that the comparison group were better at discriminating correct from incorrect information and, therefore, displayed better monitoring performance (Allwood et al., 2005). The researchers proposed that poorer discrimination was an "indication that the witnesses in the [Cognitive Interview] condition reported items that they might have chosen not to report if given only the [comparison interview] instructions" (p. 194). However, I argue for an alternative account of Allwood et al.'s (2005) results, based on the grainsize of descriptions reported <u>during</u> the interview, and the confidence in these descriptions <u>after</u> a temporal delay. Given that the Cognitive Interview group reported a greater quantity of details, it is possible that their descriptions were also finer-grained. For example, in describing a perpetrator's clothing, a statement such as "black long-sleeved jumper and dark blue jeans" is finer-grained, and contains a greater quantity of details, than "jumper and jeans". Noting this possibility for finer-grained descriptions, I will now turn to two studies that may help explain why Allwood et al.'s (2005) Cognitive Interview group appeared to show poorer monitoring performance.

In the first study, Fisher (1996) observed that the accuracy of reports elicited with a Cognitive Interview at two points in time was unaffected by retention interval. He explored this with a post hoc grainsize analysis and found that the reports elicited after a temporal delay were coarser-grained. The second study (Evans & Fisher, 2011), followed-up on this finding with a systematic investigation that examined how report accuracy is maintained when memory deteriorates over time. The researchers found that participants interviewed[10] after a one week retention interval provided coarser-grained reports than participants interviewed after a 10min retention interval. However, the report accuracy of both groups was

---

[10] A full Cognitive Interview was not used to elicit a narrative report, although some methods were employed. Participants were told the interviewer had no knowledge of the crime (depicted on the film the participant had viewed), and they were asked to describe everything in as much detail as possible but not to guess.

similar and this suggested that the participants in the delay group maintained accuracy by

coarsening the grainsize of their reports (Evans & Fisher, 2011). The conclusion it may be

possible to draw from these two studies is that people will be less confident in their *finer-*

*grained* statements, when these statements are presented after a temporal delay. The

implication for Allwood et al.'s (2005) study, assuming that their Cognitive Interview group

produced finer-grained statements, is that the participants may have become less confident in

the detail of their descriptions with time. Hence, it might be an invalid approach to use the

confidence probed after a temporal delay to assess monitoring performance *during* an

interview.

Whilst report grainsize has implications for the validity of the aforementioned

research findings (i.e., Allwood et al., 2005; Granhag et al., 2004; Gwyer & Clifford, 1997;

Mello & Fisher, 1996; Roberts & Higham, 2002), the focus that each study had on monitoring

performance is also problematic. As highlighted in the Introduction (see Figure 1.1, p. 12),

narrative recall is mediated by both monitoring and control processes. Therefore, any

relationship between confidence and recall accuracy – when this relationship is used to assess

monitoring performance – will be impacted by the decision to report (i.e., Koriat &

Goldsmith, 1996). Consequently, it is important to distinguish how the report threshold

contributes to an assessment of metacognitive processes. Narrative recall, however, poses a

difficult conundrum for this assessment because nothing is known about the information that

a person decides to withhold. To demonstrate why this withheld information is so useful to

metacognitive insight, I will comment on the two-phase forced-free methodology developed

by Koriat and Goldsmith (1996; see also Goldsmith & Koriat, 2008), that underpins their

framework of metacognition. This methodology has been adopted, and adapted, by many

researchers and it has been instrumental in understanding how metacognition impacts

reporting behaviour in the context of closed questioning (e.g., Ackerman & Goldsmith, 2008;

Goldsmith & Koriat, 2008; Goldsmith, Koriat, & Pansky, 2005; Goldsmith et al., 2002; Luna,

Higham, & Martín-Luengo, 2011; Luna & Martin-Luengo, 2012; Luna, Martin-Luengo, & Brewer, 2015; McCallum, Brewer, & Weber, 2016; Weber & Brewer, 2008).

Briefly, Koriat and Goldsmith's (1996) methodology uses a set of closed questions (e.g., general knowledge) in two phases of testing. The first phase is a forced-response test that requires participants answer every question and, following each answer, provide a confidence judgment in the response. The second phase is a free-response test, comprising the same set of questions used for the forced-response test but the participant is now free to withhold answers they are less sure are correct. The strength of this two-phase methodology lies in providing the researcher a means to measure monitoring performance (using the results from the forced-response phase) separately to the decision to report (using the results from the free-response phase). The reader is referred to Koriat and Goldsmith's paper (pp. 500-502) for the specific calculation details.

Additionally, from a signal detection theory perspective (SDT; i.e., Green & Swets, 1966), the two-phase methodology allows researchers to calculate the response bias and discriminability indices of a Type-2 SDT model (for a review of Type-1 and Type-2 SDT tasks see Galvin, Podd, Drga, & Whitmore, 2003). This is because the decision to report (i.e., response bias) defines the Type-2 task (Higham, 2007; for an excellent example of this in the witness metacognition literature, see McCallum et al., 2016 ). Importantly, the order in which the testing phases are presented does not appear to impact the results and the metacognitive inferences that can be made (Koriat & Goldsmith, 1996; however, phase order may impact witness identification performance, see Perfect & Weber, 2012). That is, the forced-response test may be administered after the free-response test. This last point is relevant to the study reported by Granhag et al. (2004), who used a two-phase free-forced methodology to understand how the Cognitive Interview impacts monitoring performance. However, a

limitation with their methodology was that the memory cues were mixed across testing phases, and why this might have implications for their findings will be examined next.

In Granhag et al.'s (2004) study, the first-phase of testing – the free-response test – comprised a Cognitive Interview (or comparison interview[11]) that produced a narrative report about the film participants had viewed. Following the interview phase, a forced-response test was administered and this comprised 45 two-alternative closed questions about the film. For each question, participants provided an answer and gave a confidence judgement in the accuracy of their response. The researchers used the results from the forced-response questions to assess monitoring performance for the interview phase. They found no difference in participants monitoring ability (discrimination or calibration) between the Cognitive and comparison interviews, suggesting monitoring performance was similar across interview protocols.

The limitation with Granhag et al.'s (2004) methodology, however, is it is fundamentally different to Koriat and Goldsmith's (1996) approach. Koriat and Goldsmith used the same set of questions across both testing phases so that the memory cues were equivalent, in order to make reliable metacognitive inferences. Granhag et al. (2004) mixed the memory cues in their two-phase methodology, with interview instructions used in the first phase and closed questions used in the second phase. The validity of this mixed methodology, to infer monitoring performance *during* the interview phase, is potentially problematic. It is possible that Granhag et al.'s (2004) finding only suggests that the monitoring performance of one group of people, forced to respond to 45 closed questions, was the same as another group of people forced to respond to the same 45 closed questions. Clearly, this conclusion is

---

[11] The comparison interview used rapport building, elicited a narrative report, and asked questions to clarify parts of the narrative that were unclear. The researchers also used a control condition but the participants in this group were not interviewed (they were only asked the 45 closed questions).

unremarkable but it does leave open questions regarding how monitoring performance is influenced by the Cognitive Interview.

In summary, existing research suggests monitoring during a Cognitive Interview is effective in recalling forensically relevant information (i.e., Roberts & Higham, 2002), but it is similarly effective during a comparison interview (i.e., Allwood et al., 2005; Granhag et al., 2004; Gwyer & Clifford, 1997; Mello & Fisher, 1996). This would suggest that the greater quantity of details – or greater report informativeness in this dimension – produced with a Cognitive Interview is not attributable to metacognitive processes. However, the grainsize of information, not assessed by the researchers (cf. Evans & Fisher, 2011; Fisher, 1996), has implications for the validity of their findings. Grainsize has critical relevance to narrative recall because report informativeness comprises both quantity and grainsize dimensions. Therefore, exploring how the Cognitive Interview influences report grainsize is a significant gap in the literature that this thesis addresses.

Further, traditional indices of metacognition are problematic for the investigative interview setting because they require confidence judgements. Pragmatically, when confidence is probed during the witness' narration of events, the interruptions will restrict report informativeness (e.g., Baddeley, 2001; D. L. Nelson & Goodmon, 2003). In contrast, when confidence is probed subsequent to the interview, it may be an insensitive measure of metacognition. Consequently, the approach taken in this thesis was to avoid confidence judgments entirely and, instead, map recall performance onto predictions made using (a) Koriat and Goldsmith's (1996) framework of metacognition, and (b) grainsize reporting (e.g., Goldsmith et al., 2002). This is examined in the next section.

## 2.2   Overview of Koriat and Goldsmith's (1996) Framework

Metacognition can be thought of as two gates – monitoring and control – that must be passed before information is reported. At the monitoring gate, information retrieved from

memory is assessed and assigned some confidence value that it is correct (Koriat & Goldsmith, 1996). Monitoring performance is defined by how well this assessment discriminates correct from incorrect information. After the monitoring evaluation, information is then subjected to a metacognitive control decision to communicate or withhold it. The decision to communicate is made when the confidence value assigned at the monitoring gate, exceeds the minimum-confidence criterion that defines the report threshold (Koriat & Goldsmith, 1996). Reporting behaviour is therefore a function of monitoring performance and the report threshold. Importantly, this overview of Koriat and Goldsmith's (1996) framework accounts for a phenomenon in reporting behaviour known as the quantity-accuracy trade-off (Koriat & Goldsmith, 1994). The implication that the quantity-accuracy trade-off has for narrative recall, and the efficacy of the Cognitive Interview, is explored next.

### 2.2.1    The quantity-accuracy trade-off and Cognitive Interview efficacy

People who are given the option to say "I don't know" to closed questions (e.g., "How old was he?"), or monetary incentives to reduce the number of incorrect answers, can improve their response accuracy by withholding answers they are less certain to be correct (Koriat & Goldsmith, 1994, 1996). The mechanism underlying this is the report threshold and it operates by increasing the minimum-confidence criterion used in the decision to report (Koriat & Goldsmith, 1996). In this way, a quantity-accuracy trade-off manifests in recall performance (e.g., Koriat & Goldsmith, 1996). However, the size of the trade-off depends on monitoring performance – as the ability to discriminate correct from incorrect answers improves, greater increases in recall accuracy can occur at lower costs to quantity performance (Koriat & Goldsmith, 1996).

In the context of an investigate interview, closed questions should only be used to seek additional details and greater clarity about information produced with open prompts (e.g., "Tell me what happened") and open questions (e.g., "What was the thief wearing?"; Fisher,

2010). Therefore, the interviewer should take a funnel approach in questioning a witness and use closed questions only after the majority of testimony has been produced with open prompts/questions (Fisher, 2010). Importantly, open prompts/questions encourage a witness to describe only what they know (e.g., Fisher & Geiselman, 1992), inherently giving the witness freedom (i.e., report option) to withhold information they are less certain about. Consequently, report option is less relevant to witness narratives, and monetary incentives are unethical as they could be considered bribing the witness. Notwithstanding, metacognitive processes might still play a role in explaining why the Cognitive Interview produces a greater quantity of information without compromising accuracy (i.e., Memon et al., 2010).

Moreover, why a quantity-accuracy trade-off is not observed might relate to the Cognitive Interview as an omnibus method. For example, some techniques might demand greater quantity and other techniques might demand greater accuracy, and the interaction of these performance outcomes might prevent an observable quantity-accuracy trade-off. Further, since narrative informativeness comprises both quantity <u>and</u> grainsize dimensions, grainsize might also play a role in explaining the quantity-accuracy relationship. Curious as to why a trade-off is not observed, and possible contributing role of grainsize, I was motivated to explore how individual Cognitive Interview techniques impact metacognition and narrative recall.

### 2.3    Metacognitive and Narrative Impact of Component Techniques

In taking a top-down approach to metacognition assessment, it was important to examine component techniques to disentangle the effect that some techniques might have on monitoring performance and other techniques might have on the decision to report. The investigation commenced with considering (a) how the mental-reinstatement-of-context instruction might influence monitoring (Section 2.3.1), and (b) how the naivety instruction might influence the report threshold (Section 2.3.2).

### 2.3.1   Mental-reinstatement-of-context instruction and monitoring performance

The mental-reinstatement-of-context instruction was designed to enhance memory retrieval by providing mnemonic assistance to the witness (Fisher & Geiselman, 1992). The instruction is considered the most effective component of the Cognitive Interview (Memon & Higham, 1999) and has received good empirical support that it enhances adult recall (Dando, Wilcock, & Milne, 2009; Dietze et al., 2012; Emmett et al., 2006; Milne & Bull, 1999). An exception is the finding by Milne and Bull (2002) that the instruction did not significantly enhance recall (although it produced extra details), but they noted a small sample of adults (*N* = 34, split between five interview conditions) as a possible reason for this non-significant effect. This exception aside, the mental-reinstatement-of-context instruction enhances recall by eliciting a greater amount of correct information (Dando, Wilcock, & Milne, 2009; Dietze et al., 2012; Emmett et al., 2006; Milne & Bull, 1999). Further, it does not appear to produce more recall errors (Dando, Wilcock, & Milne, 2009; Dietze et al., 2012) or affect report accuracy (Dando, Wilcock, & Milne, 2009), although less research has examined the effect on errors (i.e., Emmett et al., 2006) and accuracy (i.e., Dietze et al., 2012). Overall, the findings are consistent with enhanced memory retrieval but this does not rule out the idea that the instruction might also improve monitoring performance. This proposition is supported by a review of work on encoding specificity and memory cueing that suggests a role for monitoring (see pp. 36-37 in Goldsmith & Koriat, 2008)

Assuming that the mental-reinstatement-of-context instruction improves monitoring performance, the impact on the quantity of information recalled can be conceptualised with a model grounded in signal detection theory (SDT; i.e., Green & Swets, 1966). Figure 2.1 presents this model and a Type-2 rather than a Type-1 SDT model[12] is used because (a) the decision to report (response bias or ln β in SDT parlance but shown as 'Prc' in Figure 2.1)

---

[12] For a review of Type-1 and Type 2 SDT see Galvin, Podd, Drga & Whitmore (2003).

defines the Type-2 task, and (b) the separation in the distribution peaks defines monitoring

discrimination (d-prime in SDT parlance but shown as 'd' in Figure 2.1).
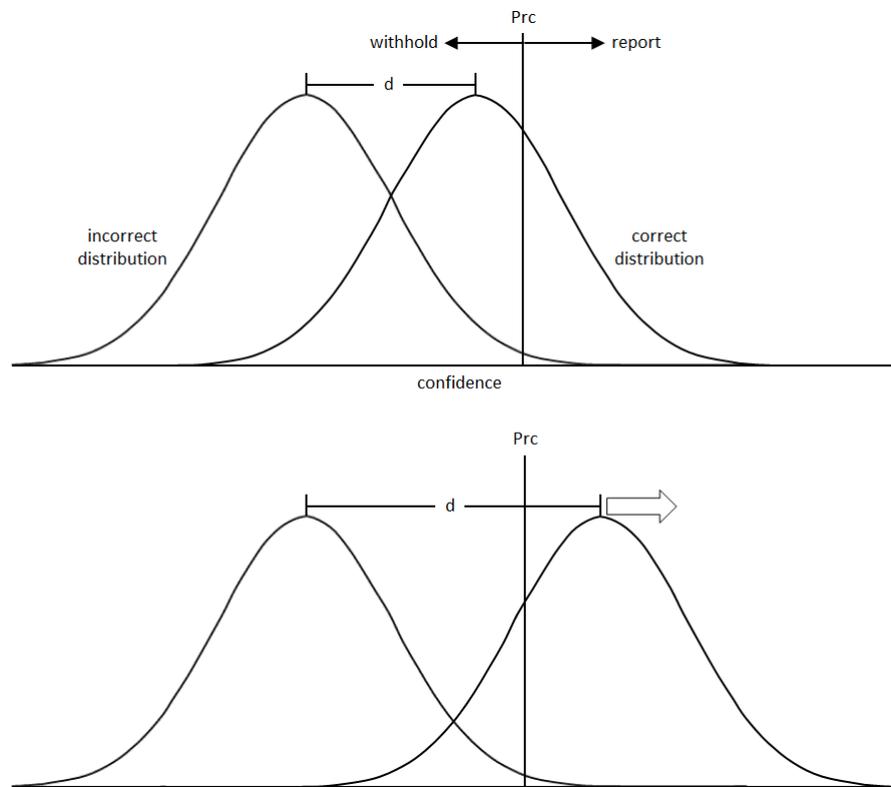


*Figure 2.1.* A Type-2 signal detection model used to conceptualise how the mental-reinstatement-of-context

instruction impacts monitoring performance. Correct and incorrect information is distributed along confidence,

the variable used by a witness to discriminate correct from incorrect information (d), and decide whether to

report or withhold information (Prc). The top diagram represents baseline metacognition when the interviewer

does not use the mental-reinstatement-of-context instruction. The bottom diagram shows the impact that the

mental-reinstatement-of-context instruction might have on improving monitoring performance (in the direction

of the open arrow), thereby, producing more correct information and not errors. This scenario assumes the report

threshold (Prc) is unaffected by the mental-reinstatement-of-context instruction.

The model is presented for conceptual purposes only because it is impossible to

calculate the metacognition indices (i.e., ln β or d-prime) without knowing the attributes of

the information withheld from a narrative (i.e., is it correct or incorrect?; Higham, 2007). As

Figure 2.1 shows, if the mental-reinstatement-of-context instruction improves a witness'

discrimination ability (assuming their report threshold is unaffected), then more correct

details, and not errors, are reported. This is consistent with the literature that suggests the mental-reinstatement-of-context instruction does not produce more recall errors (e.g., Dando, Wilcock, & Milne, 2009; Dietze et al., 2012).

In terms of Koriat and Goldsmith's (1996) framework, Figure 2.2 shows how an improvement in discrimination ability impacts recall performance. The figure presents a selection of simulated quantity and accuracy profiles constructed by Koriat and Goldsmith (1996; the reader is referred to pp. 495-498 in their paper for the simulation calculations) for various monitoring discriminability (d) and report thresholds (Prc). In detail, the quantity profiles represent input-bound performance, and the accuracy profiles represent output-bound performance. To explain the difference between input-bound and output-bound performance, consider the following scenario.

A person is asked 20 questions and they are given the option to report (i.e., freedom not to answer). The person responds to 15 questions (they decide not to answer the remaining five questions) and answers 10 correctly. This person's input-bound performance is the number of correct answers as a proportion of the total number of questions (i.e., $10/20 = .5$). Their output-bound performance is the number of correctly answered questions that they responded to (i.e., $10/15 = .75$). Critically, input-bound and output-bound performance[13] is measurable when closed questions are used because the number of to-be-remembered details is constrained by the number of questions asked. In contrast, when a narrative report is elicited with open prompts or questions, input-bound performance cannot be determined because the quantity of details that may be recalled is unconstrained. However, the performance profiles in Figure 2.2 are still relevant to narrative recall because they show the

---

[13] These measures of recall performance reflect a fundamental difference between traditional and naturalistic research approaches to memory assessment (Koriat & Goldsmith, 1994). Naturalistic research (e.g., witness memory) is concerned with how much of what is said is accurate, and traditional research is concerned with how much of a stimulus (e.g., list of words learnt) is accurately recalled.

relationship between quantity and accuracy, as a function of monitoring and the report threshold. The solid black arrows in Figure 2.2 demonstrate how an improvement in monitoring performance (at an arbitrarily set report threshold) will increase the number of correct details reported and increase report accuracy[14].
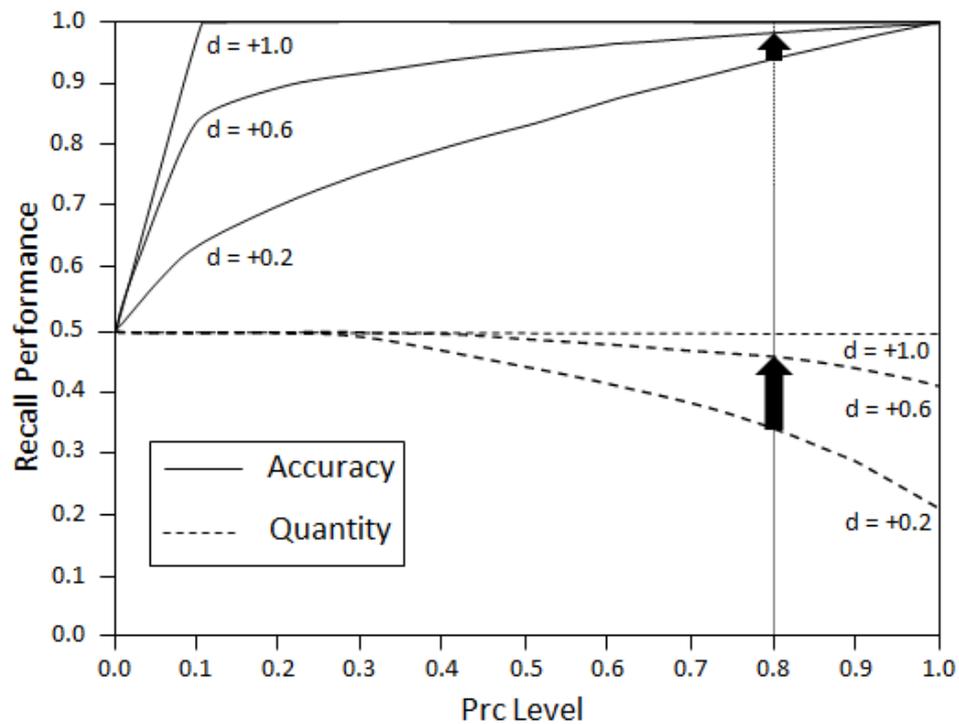


*Figure 2.2.* Simulated recall profiles (input-bound quantity correct and output-bound accuracy) for varying monitoring discriminability (d) and reporting thresholds (Prc). Greater values of 'd' reflect better monitoring performance and greater values of 'Prc' reflect a stronger tendency to withhold information. The black arrows show how improved monitoring discrimination (at an arbitrarily chosen report threshold) increases the quantity of correct details reported and increases report accuracy. The figure is replicated from Koriat and Goldsmith (Figure 4, 1996). See in text explanation for why quantity and accuracy performance is shown as a proportion on the ordinate.

In summary, if the mental-reinstatement-of-context instruction improves monitoring performance, then a greater quantity of correct details <u>and</u> greater report accuracy is expected. However, the literature suggests that the instruction elicits more correct details without

---

[14] This assumes calibration is resistant to the mnemonic influence of the mental-reinstatement-of-context instruction, an assumption that is reasonable given debiasing efforts generally do not reduce the phenomenon of overconfidence (Fischhoff, 1982; Kahneman & Tversky, 1982).

compromising accuracy (e.g., Dando, Wilcock, & Milne, 2009). In the next section, I consider

the role that grainsize might play in reconciling this difference between predicted and

observed quantity and accuracy performance.

### 2.3.1.1  *Mental-reinstatement-of-context instruction and narrative grainsize*

Grainsize will be discussed in more detail in relation to the decision to report (Section

2.3.2.1, p. 36) because the granularity of information is considered to be a principle of

communication (i.e., grainsize guides the decision to report; Fisher, 1996). Nevertheless, there

are two points I want to make about grainsize in relation to monitoring performance, because

they might elucidate how the mental-reinstatement-of-context instruction impacts grainsize.

The first point is related to the finding that, by coarsening their responses, people can

maintain their report accuracy over a temporal delay (Evans & Fisher, 2011; Fisher, 1996;

Goldsmith et al., 2005). It may be possible to conclude from this finding that the mental-

reinstatement-of-context instruction might not impact report accuracy because it produces

coarser-grained reports. Contradicting this idea, however, is the observation that the

instruction produces a greater quantity of information (e.g., Dando, Wilcock, & Milne, 2009;

Dietze et al., 2012; Emmett et al., 2006; Milne & Bull, 1999). This observation would suggest

that the mental-reinstatement-of-context instruction might produce finer-grained reports

because Evans and Fisher (2011) found finer-grained reports contained more details. In

reconciling this quantity-grainsize-accuracy paradox, I wondered if the mental-reinstatement-

of-context instruction produces more information, without compromising accuracy, because

the witness makes a greater number of coarse-grained statements (thereby maintaining

accuracy but producing more details). To explain, consider a report that contains 10

statements about different things (e.g., there might be a statement about the clothes worn by a

perpetrator), with each statement containing five details of information (e.g., the clothing

statement might include details about colour, style, material, etc.). In total, this report contains

50 details (i.e., 10 x 5 details). Now consider a second report that is equivalently accurate to the first report but contains 20 coarser-grained statements (e.g., each statement contains only four details of information). In total, this report contains 80 details (i.e., 20 x 4 details) – 30 details more than the first report and it is also coarser-grained but similarly accurate.

The second point is related to the finding that, provided there is no temporal delay in testing (i.e., Goldsmith et al., 2005), monitoring performance is similar for coarse-grained and fine-grained responses (e.g., as evidence by the discrimination index used in Goldsmith et al., 2002; and Weber & Brewer, 2008). This finding would suggest that if the mental-reinstatement-of-context instruction were to impact the grainsize of descriptions, it is unlikely to occur due to a greater ability to discriminate coarser-grained information. Inconsistent with this supposition is a finding by a recent study that suggests discrimination is poorer for coarse-grained responses (Sauer & Hope, in press). The researchers speculated this might occur if monitoring, for example, is not supported by the retrieval of additional contextual information. Given that the mental-reinstatement-of-context instruction was designed to support the retrieval of contextual information (i.e., Geiselman et al., 1984), it is reasoned that if the instruction produces coarser-grained reports, this is unlikely to occur as a result of poorer monitoring performance. In summary, it is unclear how the mental-reinstatement-of-context instruction might impact report grainsize but I speculate that it might produce coarser-grained reports.

In conclusion, the mental-reinstatement-of-context instruction was developed to assist witnesses retrieve information from memory (Geiselman et al., 1984) and it is considered to be the most effective of the Cognitive Interview techniques (Memon & Higham, 1999). Importantly, the mental-reinstatement-of-context instruction produces more correct information but not errors (e.g., Dando, Wilcock, & Milne, 2009; Dietze et al., 2012), an association that suggests the instruction might also improve monitoring performance. This

possibility was explored in Experiment 1 (Chapter 3). Moreover, the mental-reinstatement-of-context instruction does not impact report accuracy, suggesting a potential role for grainsize in maintaining accuracy. What this role is, was examined in Experiment 1 (Chapter 3).

### 2.3.2  Naivety instruction and the decision to report

The naivety instruction is so named because the interviewer tells the witness that, since they were not at the scene of the crime, they do not know what happened (Brubacher et al., 2015). The purpose of the instruction is to explicitly transfer the control of information flow to the witness, to encourage a more informative report (Fisher & Geiselman, 1992). It has been suggested that one of the ways the naivety instruction achieves this, is by challenging everyday communication norms (Milne, 2004). These norms are guided by Grice's (1975) maxims of cooperative conversation. The following is a list of these maxims and their normative influence on interpersonal communication:

- The Maxim of Quantity guides a speaker to provide sufficient information to make their point to their listener.

- The Maxim of Quality guides a speaker to say only what they believe to be true.

- The Maxim of Relation guides a speaker to focus on what is relevant to the topic of conversation.

- The Maxim of Manner guides the speaker to speak with clarity and orderliness.

Starting from an early age, people become well versed in communication norms and it has been suggested that they are probably the most important factor in selecting information to communicate (Blank, 2009). This is relevant to an investigative interview because although it is a formal communication setting (cf. conversing with a friend in a café), the interpersonal interaction between the interviewer and witness is likely to be influenced by communication norms. Therefore, it is important to consider how these norms affect the witness' decision to

select information to communicate, and how this decision influences their narrative report. For example, detailed descriptions are generally avoided in everyday conversation (Grice, 1975) but they are critical to a police investigation, to provide as many leads to follow.

One reason people avoid detailed descriptions is because a speaker will assume their listener uses similar frames of reference (e.g., scripts, schemas, abstract representations of a scene) and will, therefore, eliminate information that is highly predictable (Schuurmans & Vandierendonck, 1985). For example, in describing the layout of an office, a desk was more likely to be mentioned to an (imagined) alien than to a fellow student (Schuurmans & Vandierendonck, 1985). The researchers concluded from this finding that people expect other humans to know an office will contain a desk but that an alien is not expected to know this. Similarly, in the context of an interview, a witness might withhold information that they think the interviewer should already know. Consequently, if the interviewer explicitly states they do not know what the witness does, a statement of naivety might encourage the witness to produce this information.

Curiously, although the naivety instruction is a recommended interviewing technique (e.g., Memon, Wark, Bull, et al., 1997; Memon, Wark, Holley, Bull, & Köhnken, 1997), to my knowledge there has been no published study investigating how it influences adult recall. There is some evidence from the child witness literature that the naivety instruction increases the quantity of information reported by children (see review by Brubacher et al., 2015). This finding supports the idea that a statement of naivety encourages additional information that would otherwise be withheld. Notwithstanding, this outcome with child witnesses may or may not be generalizable to adult witnesses.
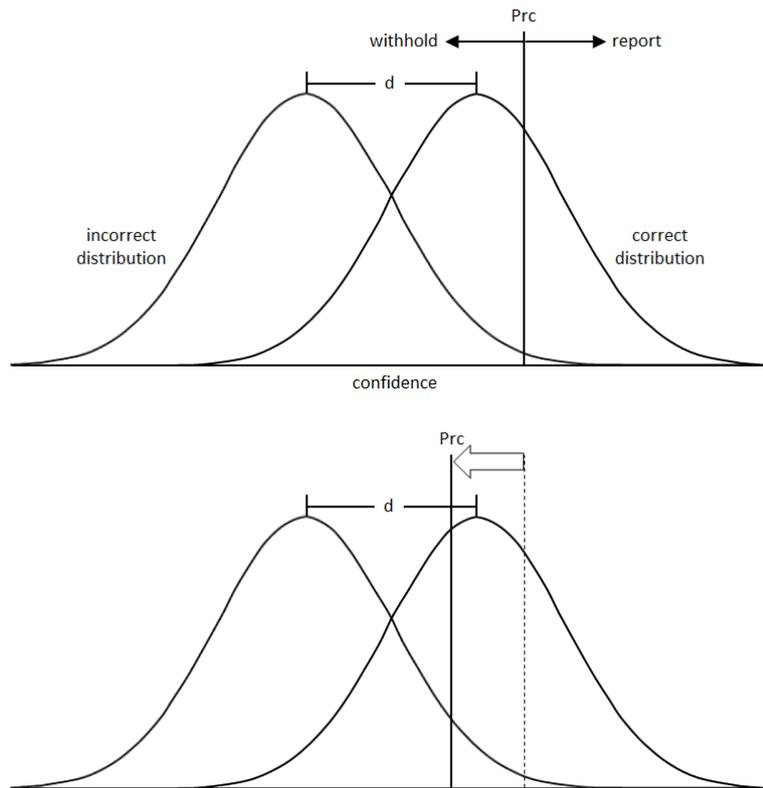
*Figure 2.3.* A Type-2 signal detection model used to conceptualise how the naivety instruction impacts the report threshold. Correct and incorrect information are distributed along confidence, the variable used to discriminate correct from incorrect information (d), and decide whether to report or withhold information (Prc). The top diagram represents baseline metacognition when the interviewer does not use the naivety instruction. The bottom diagram shows the impact that the naivety instruction might have on reducing the report threshold (open arrow), thereby, producing more correct information and errors. This scenario assumes monitoring performance (d) is unaffected by the naivety instruction.

If the naivety instruction encourages an adult witness to violate communication norms and volunteer more information, then the assumed mechanism to achieve this is the report threshold. This is because the report threshold governs the decision to communicate or withhold information (Koriat & Goldsmith, 1996). Conceptually, Figure 2.3 – using the same Type-2 SDT model that was presented earlier – shows the impact that the naivety instruction might have on the quantity of information reported. The figure shows how a reduction in the report threshold (i.e., by deciding to volunteer more information) produces a greater quantity of correct details and errors. Further, Figure 2.4 demonstrates with Koriat and Goldsmith's

(1996) simulated profiles, the potential impact on narrative quantity and accuracy performance. The solid black arrows in Figure 2.4 show how a report threshold reduction (assuming monitoring performance is unaffected) will increase the number of correct details reported but at a cost to report accuracy.
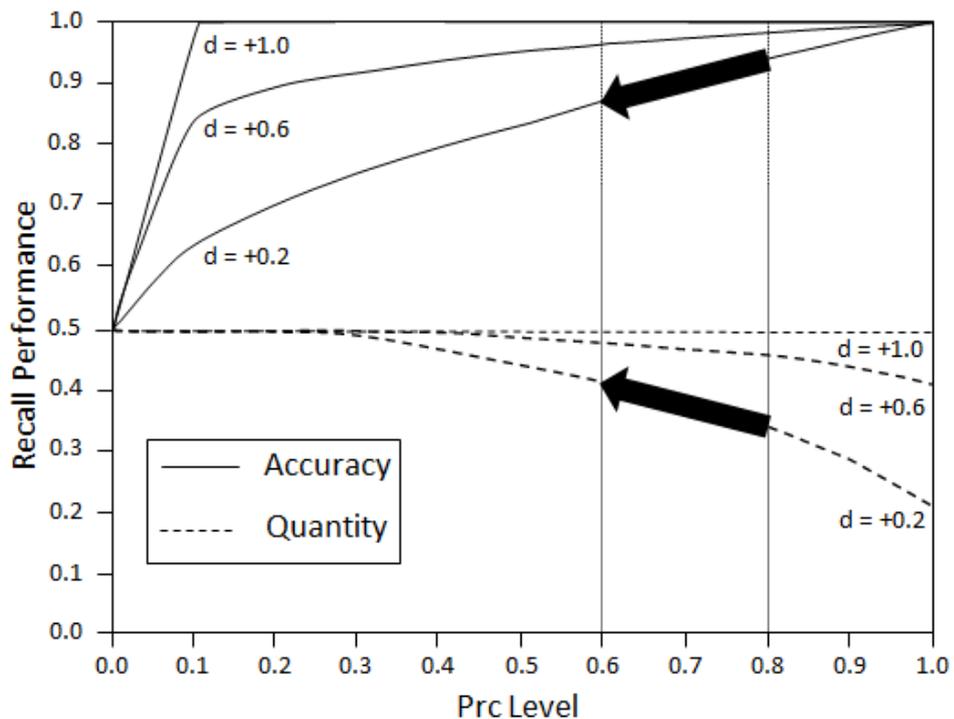


*Figure 2.4.* Simulated recall outcomes (quantity correct and accuracy) for varying monitoring discriminability (d) and reporting thresholds (Prc). Greater values of 'd' reflect better monitoring performance and greater values of 'Prc' reflect a stronger tendency to withhold information. The solid black arrows show how an increase in the quantity of correct details reported, and a decline in report accuracy, occurs when the report threshold is relaxed (discrimination ability is arbitrarily chosen). The figure is replicated from Koriat and Goldsmith (Figure 4, 1996). Why quantity and accuracy performance is shown as a proportion on the ordinate is explained in text (see Section 2.3.1, p. 28). The importance of the figure is conceptualising how the quantity and accuracy of narrative recall may be affected by the naivety instruction, if the instruction relaxes the report threshold.

In summary, if the naivety instruction encourages a witness to reduce their report threshold, then a quantity-accuracy trade-off is expected in their narrative report. How the naivety instruction influences report grainsize – the other dimension to narrative informativeness – is explored in the next section.

### 2.3.2.1  *The naivety instruction and narrative grainsize*

If the naivety instruction encourages a witness to reduce their report threshold and produce a less accurate account, as was argued in the previous section, then their narrative report should also be finer-grained. This hypothesis draws together several findings established in the context of closed questioning. First, coarse-grained answers are more likely to be accurate than fine-grained answers and this manifests a grainsize-accuracy trade-off in recall (Yaniv & Foster, 1995, 1997). Yaniv and Foster (1997) suggested that the grainsize-accuracy trade-off occurs because response grainsize reflects a compromise in the communication norms (i.e., Grice, 1975) for informativeness (Maxim of Quantity) and accuracy (Maxim of Quality). However, there is a tendency to avoid overly coarse-grained responses even when they are most likely to be accurate (Yaniv & Foster, 1997) because people perceive overly coarse-grained responses as uninformative (McCallum, Brewer, & Weber, 2015; Reid, Brewer, & Weber, 2013a, 2013b).

Second, the report threshold is the mechanism that regulates response grainsize and the grainsize-accuracy trade-off (Goldsmith et al., 2002). Specifically, a fine-grained answer is volunteered when the confidence assigned to it during monitoring exceeds the report threshold, otherwise the coarse-grained alternative is communicated (Goldsmith et al., 2002). This finding was established with general knowledge reporting but, critically, it has been replicated in witness memory research (Weber & Brewer, 2008). A significant outcome of Weber and Brewer's research is they found a grainsize-accuracy trade-off for both numerical (e.g., age range versus specific age) and verbal (e.g., tone versus specific colour) responses. This suggests that grainsize regulation is not limited to numerical responses (Weber & Brewer, 2008), a salient point for the focus that this thesis has on narrative recall because many descriptors used by witnesses are verbal.

Third, response grainsize is strategically regulated to accommodate competing goals for informativeness and accuracy (Goldsmith et al., 2002). Specifically, it has been found that the report threshold can be adjusted to accommodate either informativeness or accuracy incentives (Goldsmith et al., 2005; Goldsmith et al., 2002). For example, when inaccurate responses were heavily penalised, by raising their report threshold, participants sacrificed grainsize for accuracy (Goldsmith et al., 2005). In contrast, when correct fine-grained responses were richly rewarded, by lowering their report threshold, participants sacrificed accuracy for grainsize (Goldsmith et al., 2002). More recently, research has found fine-grained responses tend to be volunteered when the benefits of informativeness are likely to be large, and the costs to accuracy are likely to be small (N. Brewer et al., 2010; Hope, Gabbert, & Brewer, 2011; Hope et al., 2010). In summary, a grainsize-accuracy trade-off manifests in recall when the report threshold is strategically adjusted to accommodate competing goals for informativeness and accuracy. This has been demonstrated by incentivising informativeness or accuracy performance.

In the context of an investigative interview, it could be argued that incentives are unethical to apply in the form of monetary rewards or penalties. However, if the naivety instruction encourages a witness to violate the communication norm guided by the Maxim of Quantity (i.e., Grice, 1975), the statement of naivety itself is an informativeness incentive. That is, the witness might reduce their report threshold to communicate finer-grained descriptions (i.e., Goldsmith et al., 2002) and, in doing so, sacrifice accuracy (i.e., produce a grainsize-accuracy trade-off; Yaniv & Foster, 1995).

In conclusion, the naivety instruction produces more information in child witness reports (Brubacher et al., 2015) but it is unknown if this finding is generalizable to adult witnesses. This thesis addresses this gap in the literature (see Chapters 3 and 4). Theoretically, the naivety instruction is speculated to lead a witness to violate communication norms (Milne,

2004) and the proposed mechanism to achieve this is the report threshold. In this manner, the witness' narrative report will not only contain a greater quantity of details (see Section 2.3.2) but will also be finer-grained. However, greater narrative informativeness is expected to compromise narrative accuracy.

## 2.4    Summary

The Cognitive Interview produces a greater quantity of information without compromising accuracy. Existing research (Allwood et al., 2005; Granhag et al., 2004; Gwyer & Clifford, 1997; Mello & Fisher, 1996; Roberts & Higham, 2002) suggests metacognition is not responsible for the enhanced efficacy of the Cognitive Interview. However, the validity of this conclusion is drawn into question because there is a significant research gap in the role that grainsize plays in the metacognition assessment. Moreover, as an omnibus method comprising multiple techniques, the quantity-accuracy trade-off might not occur with a Cognitive Interview if some techniques demand greater informativeness, and other techniques demand greater accuracy.

By applying Koriat and Goldsmith's (1996) framework of metacognition to the investigative interview setting, I considered how instructional demands – in the form of different interview instructions – impacted metacognition and narrative performance. Specifically, my empirical work began with examining how the mental-reinstatement-of-context instruction impacts monitoring performance (Chapter 3) and how the naivety instruction impacts the decision to report (Chapters 3 and 4).

# CHAPTER 3

## Experiment 1

This chapter presents the first of three empirical studies, exploring how different Cognitive Interview instructional demands influence narrative recall, and what insights into metacognition this provides. The purpose of the study was to apply Koriat and Goldsmith's (1996) framework of metacognition to naturalistic witness reports, and address two gaps in the Cognitive Interview literature.

Koriat and Goldsmith's (1996) framework is theoretically important because it shows how, in response to questioning, metacognitive monitoring and control processes regulate a person's reporting behaviour. Notably, their framework was developed in the context of closed questioning. I was motivated to apply their framework in the context of open questioning, to understand how instructional demands made with different instructions influence narrative reports. To achieve this, I first developed a metacognition model for the investigative interview setting (see Figure 1.1, p. 12).

The model conceptualises the relationship between the interviewer, whose goal is to obtain an informative and accurate report (with the aid of interview instructions), and the witness, whose narrative report (in response to the interview instructions) is mediated by metacognitive monitoring and control processes. Importantly, an informative report comprises the quantity of details used by a witness in their account, and the grainsize of their descriptions. Thus, quantity and grainsize are two dimensions to report informativeness. This leads me to the two, though related, gaps in the Cognitive Interview literature that this chapter addresses.

The first gap is with understanding how the Cognitive Interview affects report grainsize (cf. Fisher, 1996). This is important for two reasons. First, the grainsize of

descriptions places a question mark over the finding that cognitive interviewed witnesses show similar monitoring performance to comparison interviewed witnesses (i.e., Granhag et al., 2004; Gwyer & Clifford, 1997; Mello & Fisher, 1996). Second, report grainsize might explain why a quantity-accuracy trade-off is not observed during a Cognitive Interview (i.e., Memon et al., 2010). Detailed argument for these two reasons can be found in Chapter 2.

Additionally, the Cognitive Interview is an omnibus method. Therefore, some instructions might demand recall that is mediated by monitoring performance, and other instructions might demand recall that is mediated by the decision to report. Therefore, to address the first gap, with understanding how the Cognitive Interview influences report grainsize, it was important to tease apart the individual contributions that different instructions have on all three variables (quantity, grainsize and accuracy). This was important because I took a top-down approach to metacognition assessment, by mapping narrative performance onto recall predictions. These predictions were made using (a) Koriat and Goldsmith's (1996) framework of metacognition, and (b) grainsize regulation (Goldsmith et al., 2002). Further, because grainsize regulation is assumed to be a metacognitive control mechanism (i.e., Goldsmith, Pansky, & Koriat, 2013), it was critical to use report grainsize to make inferences about metacognitive processes. I chose to test the mental-reinstatement-of-context and naivety instructions.

The mental-reinstatement-of-context instruction was chosen for two reasons. The first reason was to test the proposal that the instruction might improve monitoring performance, by improving a witness' ability to discriminate correct details from errors (see Section 2.3.1, p. 26). The second reason was to support the validity of my experimental approach. The instruction has received good empirical support that it enhances recall (e.g., Dando, Wilcock, & Milne, 2009; Dando, Wilcock, Milne, et al., 2009; Dietze et al., 2012; Emmett et al., 2006), therefore, replicating this finding would validate the experimental method.

The naivety instruction was also chosen for two reasons. The first was to test the proposal that the instruction encourages a witness to relax their report threshold and produce a more informative report (see Section 2.3.2, p. 32). The other reason was to address the second gap in the literature; understanding how the instruction influences adult recall. For these reasons, the instruction was given to participants after viewing two types of stimuli: film and picture.

Of note, the mental-reinstatement-of-context instruction was not used with the picture stimulus because the efficacy of the instruction has been replicated across different stimuli (i.e., film, staged, slide, narrative; see Memon et al., 2010). Further, the narrative responses to the naivety instruction could be compared across stimuli without making the experiment excessively long and risking participant fatigue (by also using the mental-reinstatement-of-context instruction with the picture).

### 3.1    Method

#### 3.1.1    Participants

The sample comprised 93 participants recruited from two pools of people, the general community and first year Psychology students. Community participants were given the chance to win AUD$200 and students participated for course credit. Informed written consent, including permission to be video-recorded, was obtained from all participants. Twelve participants were excluded from the study because equipment malfunctioned (4 people), participants' English was not fluent (4 people), instructions were not followed (3 people), and the interviewer coughed excessively throughout one person's recall[15]. The mean age, of the remaining 81 participants, was 30.72 years and ranged from 18 – 77 years. The

---

[15] Eighty participants were required to detect a medium effect size with statistical power of .80. Additional participants were recruited when exclusions occurred. Risk factors for exclusion were better managed in the subsequent studies because participants were more effectively screened and given more detailed instructions.

sample comprised 31 males ($M$ = 32.58 years, $SD$ = 17.29) and 50 females ($M$ = 29.56 years, $SD$ = 12.36). All remaining participants had normal or corrected-to-normal vision and were fluent in English.

### 3.1.1.1 Interviewer

One interviewer (40-year-old male) was used and volunteered his time for a small honorary payment per interview. He was naive to the experimental aims of the research. Before the study commenced, the interviewer familiarised himself with the interview instructions, printed on separate cards, for each experimental condition. He then conducted several practice interviews over the course of a single day, using myself and other volunteers as mock witnesses. This training helped to maintain experimental control.

During the study, to help build rapport, the interviewer was friendly with participants, smiled, and made casual conversation by asking how their day was going. For every interview, the interviewer gave instructions slowly and coherently to the participant, and he actively listened to their recall by using head nods and neutral acknowledgements (e.g., "uh-huh").

### 3.1.2 Materials

The stimulus materials used were a film and a picture, described separately below. All other materials used in the study are detailed in Section 3.1.3.

### 3.1.2.1 Film stimulus

A two-minute film showed a non-violent, credit card theft. A young male perpetrator entered a restaurant whilst another young male left his credit card with the waiter at the front desk. The perpetrator then asked the waiter about booking availability. He stole the credit card when the waiter's back was turned to check a diary. The waiter then answered a telephone. Whilst he was speaking on the telephone the perpetrator exited the restaurant and ran away.

### *3.1.2.2 Picture stimulus*

An A4-sized colour picture showed a cobbled street scene (see Appendix A). Three blocks of colourful buildings surrounded a cobbled road intersection. Two cars were parked on the right hand side. A woman swept the footpath on the left hand side.

## 3.1.3 Procedure

Participants were tested individually and to view the film, they sat in a partitioned cubicle, approximately 55 cm from a high resolution, 17 in, colour video monitor. Participants used headphones to listen to the film. When the film finished, participants were asked if they had seen the film prior to the study, and if they recognised any people in the film. All participants answered "no" to both questions.

After a 35 min distraction period, participants were led to another room, where the interviewer was waiting. To help build rapport, the interviewer and participant sat positioned at right angles at a desk rather than opposite each other. Participants were interviewed with one of four interview conditions, randomly assigned to each participant.

The mental-reinstatement-of-context (MRC) and naivety (NVT) instructions were manipulated in a two-way factorial, between-subjects design to elicit a narrative report. The four interview conditions comprised: (a) MRC-absent, NVT-absent; (b) MRC-present, NVT-absent; (c) MRC-absent, NVT-present; and (d) MRC-present, NVT-present. These conditions are explained next.

Participants in the control condition (MRC-absent, NVT-absent) were asked to describe everything in as much detail as possible, without guessing. The request made by the interviewer essentially comprises two instructions: (a) to describe everything in as much detail, and (b) not to guess. Since the interviewer's goal is to obtain an informative and accurate report, the control interview was a basic approach to achieve this by using, for the

purpose of definition, the report-detail (i.e., describe everything in detail) and do-not-guess instructions. Specifically, with this basic approach, the report-detail instruction was used to motivate the participants to give an informative account, similarly suggested by Bensi, Nori, Gambetti and Giusberti (2011) for instructing their participants to "recall as many details as possible" (p. 314). The do-not-guess instruction was used because, although narrative recall gives a witness the freedom to report only what they know (i.e., Fisher & Geiselman, 1992), it has been suggested that the instruction not to guess provides additional report option (i.e., freedom to withhold information; Koriat & Goldsmith, 1996) during narrative recall (Evans & Fisher, 2011). The basic approach to obtain an informative and accurate account, provided the baseline to compare the influence that the mental-reinstatement-of-context and naivety instructions had on report informativeness and accuracy. Notably, the report-detail and do-not-guess instructions were used across all interview conditions.

When the mental-reinstatement-of-context instruction was present, the interviewer stated: "Remember being seated to watch the film. Sometimes it helps us to remember things if our eyes are closed or we stare at a blank spot in front of us. So if you would like to, you can close your eyes or stare at the wall while you remember what happened on the film (pause). What I want you to do is create a picture in your mind of what you witnessed. What could you see? What could you hear? Think about how you were feeling while you were watching the film (pause). What thoughts you were having (pause)."

When the naivety instruction was present, the interviewer stated: "I have not seen it [the film] myself so I do not know what you witnessed. You have all the information and I'd like you to share it with me".

Participants were not interrupted during their reports. When the participant appeared to have finished volunteering information, the interviewer waited for a short period and then

asked if there was anything else they wanted to add. The film interview concluded after additional information was volunteered or the participant declined to add anything else.

Participants were then informed that a picture would be shown briefly. The picture was held by the interviewer so only the participant could see it. It was held in place for 5 s, approximately 55 cm from the participant, and then replaced face down on the table. Participants were interviewed immediately to elicit a narrative report, using one of two interview conditions randomly assigned to each participant. Participants in the control condition (the naivety instruction was absent) were simply asked to describe what they saw on the card. When the naivety instruction was present, the interviewer stated: "I do not know what is on the card so I only have your description to rely on".

Picture interviews followed the same procedure as for the film interviews (i.e., the interviewer did not interrupt the participant and he listened actively to their report). The follow-up question, asking participants if they wanted to add anything else, was omitted. After interviewing was finished, each participant was debriefed.

The actual interview instructions used in the study, for the film and picture interviews, can be found in Appendix B. The instructions were read out verbatim to participants. Every interview was video-recorded for later transcription and scoring.

### 3.1.4   Dependent variables to assess report informativeness and accuracy

An informative report comprises the quantity of details, used to describe things, and the grainsize of descriptions. Thus, report informativeness was assessed with quantity and grainsize variables. The quantity variables comprised the number of correct details and errors recalled. How this information was coded in the interview transcripts, is described in Section 3.1.5.

The grainsize variable was grainsize precision. This variable represents the proportion of information recalled that was fine-grained. To calculate grainsize precision, I first developed the concept of 'grainsize chunk'. 'Grainsize chunk' is a non-technical label applied to a statement of related information (e.g., a statement about the thief's upper-body clothing). Grainsize chunks could be either coarse-grained or fine-grained statements, and how these were coded is described in Section 3.1.5. Grainsize precision was calculated by dividing the number of fine-grained chunks by the total number of grainsize chunks recalled (fine-grained + coarse-grained).

Report accuracy was assessed with an accuracy variable. This variable represents the proportion of details recalled that were correct. It was calculated using the quantity variables, by dividing the number of correct details by the total number of details recalled (correct + errors).

For all variables, except accuracy and grainsize precision, individual scores could only attain a value of '1'.

### 3.1.5   Scoring narrative recall

Video recordings of every interview were transcribed verbatim. However, the interview instructions were omitted from the transcripts to ensure the coders, during quantitative scoring, were blind to the experimental condition. Coding keys were produced for this thesis so that the quantity (correct and errors) and grainsize (coarse-grained and fine-grained) of information could be scored. These keys comprised a list of (a) rules to apply during scoring, (b) details that could be scored, and (c) scoring examples. The coding keys and scoring procedures are overviewed in Section 3.1.5.1 for the quantity and accuracy variables, and Section 3.1.5.2 for the grainsize variables (the reader is referred to Appendices C to F for further detail on the scoring rules, coding keys and examples).

### *3.1.5.1   Quantity coding key, and scoring quantity and accuracy*

I adopted the coding procedure of Milne and Bull (2002) to produce the quantity coding key for this thesis. Briefly, the procedure entailed a two-step process. The first step involved describing the film and picture in detailed written format such that the details comprised information that could be coded in the transcripts. The second step in producing the final coding key was more time-consuming because it involved an iterative process of pilot coding of five randomly selected transcripts. Two volunteers were involved in this process and they used the (evolving) coding key to first score a transcript from the practice interviews that the interviewer conducted, and then provide feedback to me about the key's usability. I then updated the coding key to incorporate their feedback by adding details to the key that could be scored for the next iteration of pilot coding. Importantly, for each iteration of coding I calculated the inter-rater reliability (percentage agreement) and found the key's reliability improved with each coding iteration. The inter-rater reliability started at less than 50% on the first coding attempt and it rose to greater than 80% on the fifth coding attempt. After the five iterations of coding, the coding key was deemed a final version. However, additional details were subsequently added to the key over the course of coding more than 400 transcripts for the three studies reported in this thesis.

For the quantity variables, details were scored correct if the information in the transcript matched the coding key (e.g., stating the thief wore a jacket and he was wearing one) or, if the detail was not present in the key, could be confirmed from the film. Details were scored as incorrect if the information did not match the coding key or film (e.g., stating the thief wore a jumper). Details were scored as confabulated if the information could be verified from the film as being made-up by the participant (e.g., stating the thief wore a hat when it could be seen in the film that he was not wearing one). Because the number of confabulated details was very low, confabulated and incorrect scores were combined into a

single variable, errors, to more broadly represent error in recall (similarly done by Memon et al., 1994). Fine-grained descriptors were scored in preference to coarse-grained descriptors (e.g., if both "card" and "American Express" were stated for the credit card, then "American Express" was scored).

Suppositions that related to feeling (e.g., "He looked upset") or thinking state (e.g., "He planned to steal the card") were not scored. Information was also not scored if it could not be verified from the film (e.g., "There was someone in the 4WD" but this could not be verified from the film). Qualified statements were treated as unqualified (e.g., "I think he was wearing a jacket" was treated as "He was wearing a jacket").

### 3.1.5.2  *Grainsize coding key and scoring procedure*

Procedures for coding narrative grainsize have been published (i.e., Douglass et al., 2013; Evans & Fisher, 2011), however, I developed a procedure that incorporated all information reported, including the relation of details to actions. Existing grainsize coding procedures do not appear to use action-related information (i.e., Douglass et al., 2013; Evans & Fisher, 2011). I considered it important to include action-related information in assessing narrative grainsize because actions are critical to episodic memory reporting.

The grainsize coding key was developed with the assistance of a South Australian police officer (Chief Inspector rank), using a two-step process. For the first step, grainsize chunks were defined by grouping related information together (e.g., information related to the thief's upper body clothing, information related to the customer leaving his credit card, etc.). For the second step, and drawing from transcripts randomly selected, examples of coarse-grained (e.g., "jacket") and fine-grained (e.g., "black and white sports jacket") statements were defined for each grainsize chunk. Transcripts were then scored for the number of fine-grained and coarse-grained chunks reported. If information was related to the same grainsize chunk, fine-grained chunks were scored in preference to coarse-grained chunks (e.g., if both

"jacket" and "black and white sports jacket" were reported for the thief's upper body clothing grainsize chunk, then "black and white sports jacket" was scored). A tally of coarse-grained and fine-grained chunks reported by each participant derived the number of grainsize chunks recalled. The key variable of interest, grainsize precision, was calculated by dividing the number of fine-grained chunks by the total number of grainsize chunks recalled (fine-grained + coarse-grained).

### 3.1.5.3  *Inter-rater reliability*

All transcripts were scored by myself, and seventeen interviews (20%), randomly selected, were scored independently by a second coder. As noted earlier, the interview instructions were omitted from the transcripts to ensure scoring was done blind to the interview condition.

The inter-rater reliability was assessed by percentage agreement and this approach was chosen over the kappa statistic (i.e., Cohen, 1960) for the following reasons. Whilst the kappa statistic accounts for the possibility that coders might agree in their coding only because they might guess codes when uncertain (Cohen, 1960), the coders always had the option to not score (i.e., they were not forced to make a judgement) and therefore less likely to make coding guesses. Further, the prevalence of occurrences (i.e., codable details – correct, incorrect or confabulated) and non-occurrences (i.e., non-codable features – suppositions, digressions, repeated information) did not appear especially biased either way. Therefore, if the coders did make coding guesses, these guesses would have less impact on the inter-rater calculation (i.e., Kazdin, 1982). Finally, when codes are not equiprobable, the kappa statistic will be lower (Sim & Wright, 2005). This is a non-trivial issue for narrative recall because codable details are dominantly correct (e.g., as evidenced by the strong impact that the Cogntive Interview has on correct recall; Memon et al., 2010).

For the total quantity of information (correct + errors), an inter-rater reliability of 80.3% was established for the film stimulus, and 79.4% for the picture stimulus. Disagreements were reviewed and the majority were related to human error in simply missing details. The raw inter-rater score is presented, however, to demonstrate the robustness of the quantity coding key. My scores, except when they were changed during discussion with the second coder, were retained for the main analysis.

For the total number of grainsize chunks (fine-grained + coarse-grained), after disagreements were reviewed, an inter-rater reliability of 97.5% was established for the film stimulus. The grainsize coding key was updated with additional examples, to help coders in future studies determine the category of grainsize (i.e., coarse- or fine-grained) for each grainsize chunk reported. Again, my scores, except when they were changed during discussion with the second coder, were retained for the main analysis. However, due to the small number of grainsize chunks reported for the picture, coders jointly scored the grainsize of each transcript.

### 3.2 Results

Descriptive statistics for the different interview conditions are presented in Table 3.1 for the film stimulus, and Table 3.2 for the picture stimulus. The tables show that most dependent variables were influenced by the interview condition, except accuracy for the film stimulus, and errors and accuracy for the picture stimulus.

#### 3.2.1 Data screening

Parametric assumptions were checked for each film and picture interview condition. This was done with the Shapiro-Wilk normality test and the Levene's test for homogeneity of variance (using the median). Outliers were examined with standardized data and identified when $z$-scores were less than -2.50 or greater than 2.50.

### 3.2.1.1 Film stimulus: parametric assumptions and outliers

All dependent variables met parametric assumptions. No outliers were found in correct recall or grainsize precision but they were found in errors (3 outliers) and accuracy (1).

### 3.2.1.2 Picture stimulus: parametric assumptions and outliers

Correct recall met parametric assumptions. All other dependent variables were skewed with equal variance. No outliers were found in grainsize precision but they were found in correct recall (2 outliers), errors (2), and accuracy (2).

Table 3.1

*Mean Recall [and 95% confidence intervals] by Interview Condition, for the Film Stimulus*

|  | Interview condition | | | |
|---|---|---|---|---|
|  | MRC-a NVT-a | MRC-a NVT-p | MRC-p NVT-a | MRC-p NVT-p |
|  | ($n = 20$) | ($n = 20$) | ($n = 21$) | ($n = 20$) |
| Variable | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* |
| Correct | 95.00 (33.97) | 99.00 (33.26) | 115.30 (35.01) | 125.40 (33.11) |
|  | [79.10, 110.90] | [83.86, 114.14] | [98.91, 131.69] | [109.90, 140.90] |
| Errors | 5.10 (4.54) | 5.67 (3.45) | 7.95 (4.70) | 7.90 (4.39) |
|  | [2.98, 7.23] | [4.09, 7.24] | [5.75, 10.15] | [5.85, 9.95] |
| Accuracy | .95 (.03) | .94 (.04) | .94 (.02) | .94 (.03) |
|  | [.94, .96] | [.92, .96] | [.92, .95] | [.93, .95] |
| GS Precision | .43 (.17) | .45 (.13) | .49 (.11) | .50 (.16) |
|  | [.35, .52] | [.39, .51] | [.44, .54] | [.43, .57] |

*Note.* Correct = Number of correct details; Errors = number of errors; Accuracy = proportion of correct details; GS Precision = proportion of fine-grained recall.

MRC-a = mental-reinstatement-of-context instruction absent; MRC-p = mental-reinstatement-of-context instruction present; NVT-a = naivety instruction absent; NVT-p = naivety instruction present.

Table 3.2

*Mean Recall [and 95% confidence intervals] by Interview Condition, for the Picture Stimulus*

| | Interview condition | |
|---|---|---|
| | NVT-a | NVT-p |
| | (*n* = 29) | (*n* = 32) |
| Variable | *M (SD)* | *M (SD)* |
| Correct | 17.03 (7.74) | 21.50 (6.40) |
| | [14.09, 19.98] | [19.19, 23.81] |
| Errors | 2.17 (1.51) | 2.88 (2.90) |
| | [1.60, 2.75] | [1.83, 3.92] |
| Accuracy | .89 (.07) | .89 (.10) |
| | [.86, .91] | [.86, .93] |
| GS Precision | .31 (.20) | .48 (.25) |
| | [.23, .39] | [.39, .57] |

*Note.* Correct = Number of correct details; Errors = number of errors; Accuracy = proportion of correct details; GS Precision = proportion of fine-grained recall.

NVT-a = naivety instruction absent; NVT-p = naivety instruction present.

### 3.2.1.3  Permutation ANOVA

Permutation ANOVA is a robust statistical method that is useful to apply when data violate parametric assumptions and/or contain outliers (Anderson, 2001). The method is particularly useful for factorial ANOVA designs where there is no equivalent non-parametric test (see Anderson, 2001, for a thorough review of permutation tests, why they are useful, and how to apply them to complex experimental designs). Notably, permutation ANOVA does not replace parametric ANOVA – the tests are run in conjunction – but permutation methods eliminate the need for traditional approaches (e.g., transformations, removing outliers, trimming the mean, etc.) to make data amenable to parametric tests. I will explain permutation ANOVA in the context of my experimental design and Manly's method of data

permutation (Manly's and several other methods are reviewed by Anderson, 2001). I chose to use Manly's method because it is a straightforward process and it uses unrestricted sampling of the raw observation data.

With respect to the present experimental design for the film stimulus, imagine there are four glass jars and each jar represents one of the four interview conditions. Now imagine a marble represents the response data of each participant, and the numbers of marbles in each jar is equivalent to the number of participants in each interview condition. Let this sample of marbles be called the original sample. Let us analyse this original sample with parametric ANOVA, to derive $F$-values for the main effects and interaction term.

Now imagine we pour all of the marbles from each jar into a raffle barrel, turn the barrel a few times and then pick a marble at random, returning it to one of the four glass jars (also randomly chosen). We continue this process until there are no marbles remaining in the barrel. Let the sample of marbles, now distributed across the four jars, be called the new sample. This process conceptualises Manly's method because the participants' raw data is used and there are no restrictions placed on the permutation process. Let us analyse the new sample with parametric ANOVA, to derive $F$-values for the main effects and interaction term. If we redo the entire process (i.e., pour the marbles into the barrel, pick one at random, etc.), say, 2,000 times we will have created 2,000 new samples and calculated 2,000 parametric ANOVAs. This entire process is known as permutation ANOVA.

The 2,000 $F$-values (for each main effect and the interaction term) derived from the permutation process will form a distribution that the original sample $F$-values can be compared against. This comparison is the most critical aspect of permutation ANOVA. This is because the output from the permutation calculation is a permutation $p$-value, representing the likelihood that the original sample $F$-values occurred by chance. For example, if the original sample returned a significant result for one of the main effects, and we applied an

alpha level of .05 to permutation ANOVA, then obtaining a permutation $p$-value less than .05 suggests that the original sample $F$-value (for the main effect) is unlikely to be a Type I error. Alternatively, obtaining a permutation $p$-value greater than .05 suggests that the original sample $F$-value is likely to be a Type I error. Further, if the original sample returned a non-significant $F$-value but the permutation $p$-value is less than .05, this suggests that the original $F$-value is likely to be a Type II error. In short, permutation ANOVA is essentially a tool that tests the reliability of applying parametric ANOVA to data that does not meet parametric assumptions. The number of data permutations can be any quantity, however, the more that are computed will derive a smoother distribution of $F$-values and greater reliability in the permutation $p$-value.

For my sample, I used Manly's method of data permutation and 2000 permutations in the permutation ANOVA analysis. Permutation ANOVA, and the permutation $p$-value, was calculated in R, an open-source language and environment for statistical computing (version 3.2.2; R Core Team, 2014), using code that I adapted from Howell (2009). The reader can find the code that I used in Appendix G. For each dependent variable (for both the film and picture stimulus) that showed a significant parametric ANOVA result, the permutation $p$-value that was obtained with permutation ANOVA was less than .05. This suggested that the parametric ANOVA result was reliable and unlikely to be a Type I error, regardless that there were outliers and skewness observed in the data.

### 3.2.2 Film stimulus narrative performance

In the following analyses, the focus was on testing (a) if the mental-reinstatement-of-context instruction improves monitoring performance, and (b) if the naivety instruction challenges communication norms and encourages a witness to reduce their report threshold.

If the mental-reinstatement-of-context instruction impacts monitoring, this was expected to manifest a greater quantity of correct details, and not errors, in reports. However,

enhanced monitoring also predicts greater report accuracy (i.e., Koriat & Goldsmith, 1996), a finding inconsistent with the literature that suggests the mental-reinstatement-of-context instruction does not improve testimonial accuracy (e.g., Dando, Wilcock, & Milne, 2009). Therefore, report grainsize was examined for the role it plays in maintaining accurate accounts.

If the naivety instruction impacts the report threshold, this was expected to manifest finer-grained reports containing a greater quantity of correct details. However, as predicted by the grainsize-accuracy (i.e., Goldsmith & Koriat, 2008) and quantity-accuracy (i.e, Koriat & Goldsmith, 1996) trade-offs, greater report informativeness was expected to compromise accuracy and produce more recall errors.

Inferential analyses were calculated using univariate factorial ANOVA (alpha level set at .05) in IBM SPSS Statistics 21. A Type II sum of squares calculation was used due to the unbalanced design albeit the interview conditions only differed in size by one person. Effect size, eta squared ($\eta^2$), was calculated using the *lsr* package (version 0.5; Navarro, 2015) in R (version 3.2.2; R Core Team, 2014). For significant effects, point estimates of the mean are given with 95% confidence intervals. These were calculated using the bias-corrected and accelerated bootstrap method in SPSS, using 2000 permutations. Qualitative descriptions for eta squared adopted Cohen's (1988) recommendations (i.e., .01 = 'small', .06 = 'medium', .14 = 'large').

In presenting the results, I focus on the main effects that the mental-reinstatement-of-context and naivety instructions had on report informativeness and accuracy (see Figures 3.1 to 3.5). There was no interaction observed between the instructions on any dependent variable: correct recall, $F(1, 77) = 0.16$, $p = .69$, $\eta^2 < .01$; errors, $F(1, 77) = 0.11$, $p = .75$, $\eta^2 < .01$; grainsize precision, $F(1, 77) < 0.01$, $p = 1.00$, $\eta^2 < .01$; and accuracy $F(1, 77) = 1.57$, $p = .22$, $\eta^2 = .02$.

### 3.2.2.1 *How the mental-reinstatement-of-context and naivety instructions impacted report informativeness*

In this section, the results are presented separately for the quantity and grainsize dimensions of report informativeness.

#### 3.2.2.1.1 *Quantity dimension of report informativeness*

As expected, the mental-reinstatement-of-context instruction increased correct recall, $F(1, 77) = 9.67$, $p < .01$, $\eta^2 = .11$ (see Figure 3.1). Interviews with the mental-reinstatement-of-context instruction, had a medium to large sized effect on producing more correct details ($M = 120.35$, $SE = 5.38$, $CI_{95}[109.73, 129.56]$) compared to the interviews without the instruction ($M = 97.05$, $SE = 5.19$, $CI_{95}[86.70, 107.08]$).

Unexpectedly, the mental-reinstatement-of-context instruction also increased recall errors, $F(1, 77) = 7.10$, $p = .01$, $\eta^2 = .08$ (see Figure 3.2). Interviews with the mental-reinstatement-of-context instruction, had a medium sized effect on producing more errors ($M = 7.93$, $SE = 0.71$, $CI_{95}[6.63, 9.22]$), than interviews without the instruction ($M = 5.39$, $SE = 0.62$, $CI_{95}[4.26, 6.65]$).

Turning to the naivety instruction, it was surprising to find that it did not influence correct recall, $F(1, 77) = 0.87$, $p = .35$, $\eta^2 = .01$ (see Figure 3.1), or errors, $F(1, 77) = 0.08$, $p = .78$, $\eta^2 < .01$ (see Figure 3.2). These findings were unexpected because, with the picture stimulus (see Section 3.2.3.2, p. 61), the naivety instruction did influence the quantity dimension of report informativeness.

*Figure 3.1.* Mean correct recall (+SE) for each interview condition. MRC = mental-reinstatement-of-context instruction. NVT = naivety instruction.



*Figure 3.2.* Mean errors (+SE) for each interview condition. MRC = mental-reinstatement-of-context instruction. NVT = naivety instruction.

### 3.2.2.1.2 *Grainsize dimension of report informativeness*

Curiously, the mental-reinstatement-of-context instruction did not influence grainsize precision, $F(1, 77) = 2.96$, $p = .09$, $\eta^2 = .04$ (see Figure 3.3), as I speculated it might (see Section 2.3.1.1, p. 30).

Further, the naivety instruction did not influence grainsize precision, $F(1, 77) = 0.11$, $p = .74$, $\eta^2 < .01$ (see Figure 3.3). This was also an unexpected finding because the naivety instruction influenced grainsize precision in regards to the picture stimulus (see Section 3.2.3.2, p. 61).
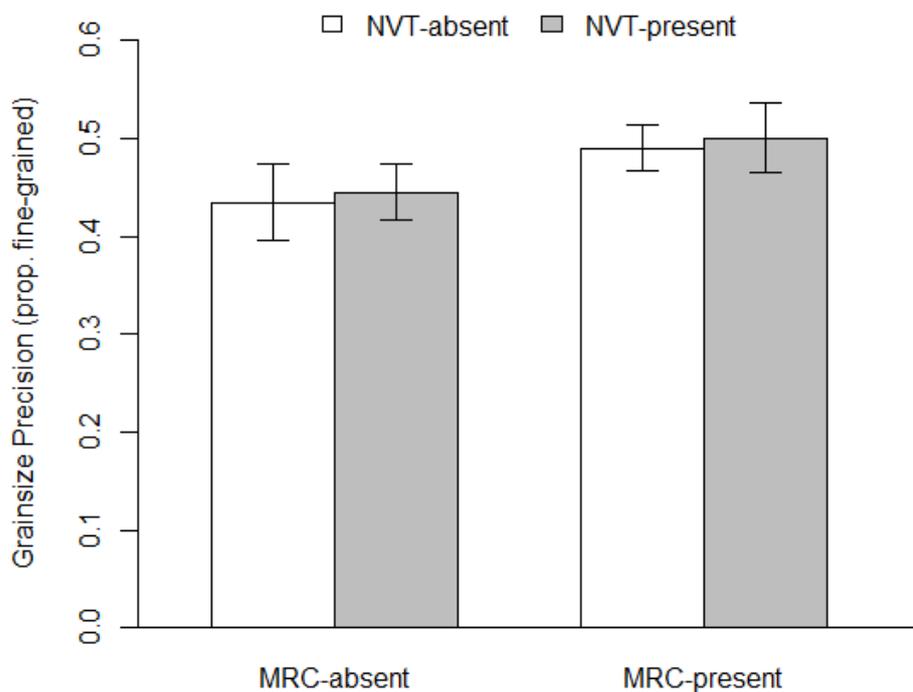


*Figure 3.3.* Mean grainsize precision (+SE) for each interview condition. MRC = mental-reinstatement-of-context instruction. NVT = naivety instruction.

### 3.2.2.2 *How the mental-reinstatement-of-context and naivety instructions impacted report accuracy*

Contrary to expectations, neither the mental-reinstatement-of-context, $F(1, 77) = 1.23$, $p = .27$, $\eta^2 = .02$, nor the naivety, $F(1, 77) = 0.18$, $p = .67$, $\eta^2 < .01$, instructions influenced

recall accuracy (see Figure 3.4). The ceiling effect that can be seen in Figure 3.4 is a likely

reason that there was no effect observed. This finding has broader theoretical and practical

implications that will be addressed in the General Discussion (see Chapter 6). However, a

ceiling effect in narrative accuracy is not a negative outcome for the investigative

interviewer's goal to obtain an accurate report. Moreover, it was interesting to find that the

mental-reinstatement-of-context instruction did not influence report accuracy because this
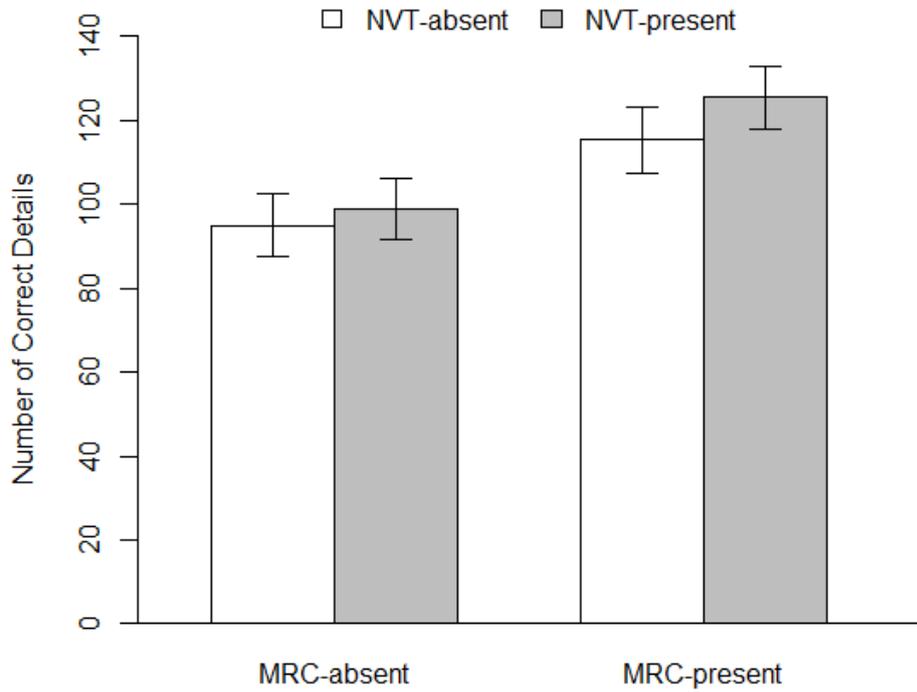
replicates Dando et al.'s (2009) finding.
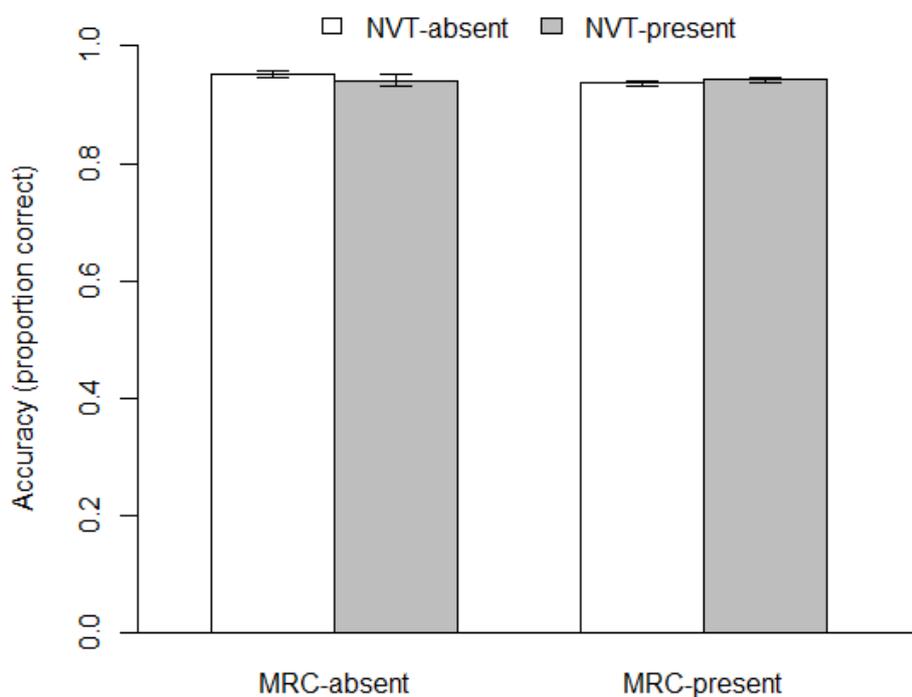


*Figure 3.4.* Mean accuracy (+SE) for each interview condition. MRC = mental-reinstatement-of-context
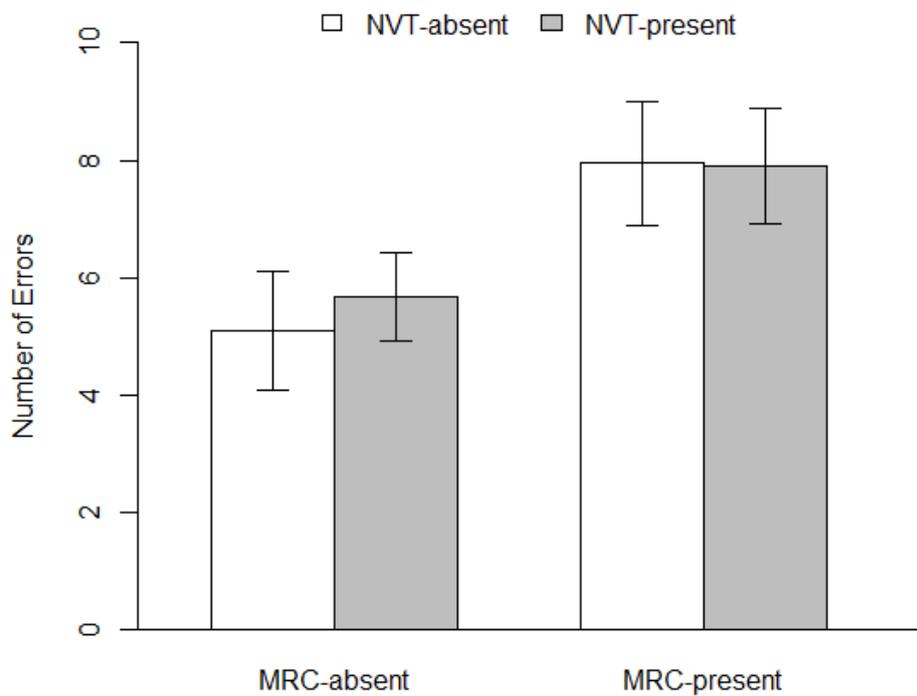
instruction. NVT = naivety instruction.

### 3.2.3   Picture stimulus narrative performance

In these analyses, I wanted to observe how the naivety instruction influenced narrative

performance when a different to-be-remembered stimulus is used. Specifically, the focus was

on observing recall effects similar to those that were expected for the film stimulus. That is, if

the naivety instruction encourages a witness to reduce their report threshold, this was

expected to manifest a finer-grained report containing a greater quantity of correct details and errors, and greater report informativeness was expected to compromise accuracy

Inferential analyses were calculated using one-way ANOVA (alpha level set at .05) in IBM SPSS Statistics 21. Effect size was calculated with eta squared, $\eta^2$, using the *lsr* package (version 0.5; Navarro, 2015) in R (version 3.2.2; R Core Team, 2014) For significant effects, point estimates of the mean are given with 95% confidence intervals. These were calculated using the bias-corrected and accelerated bootstrap method in SPSS, using 2000 permutations. Qualitative descriptions for eta squared adopted Cohen's (1988) recommendations (i.e., .01 = 'small', .06 = 'medium', .14 = 'large').

### 3.2.3.1 *Analysis of presentation time*

The time taken for the interviewer to hold the picture in view for participants, increased systematically with progression of the study. Viewing time ranged from 3.88 s to 15.10 s ($M = 9.40$, $SE = .29$, $CI_{95}[8.83, 9.97]$).

Shorter viewing time may have unduly influenced recall so I explored this with correlations. I chose not to use ANCOVA, with presentation time as the covariate, because the assumption cannot be made that the covariate (i.e., presentation time) is independent of the experimental effect (i.e., naivety instruction). Specifically, the impact of any instructional demand on recall performance ultimately depends on memory encoding, and encoding depends on presentation time (e.g., Coyne, 1985; Melcher, 2006). Returning to the correlational analysis, more correct details were recalled with longer viewing time, $r(81) = .32$, $p < .01$. There was no relationship between viewing time and errors, $r(81) = .07$, $p = .51$, accuracy, $r(81) = .14$, $p = .20$, or grainsize precision, $r(81) = .05$, $p = .63$.

The confounding influence of viewing time was removed by arbitrarily excluding the lower 25th percentile of data. This derived a sub-set of data ($n = 61$) with a lower cut-off of

7.67 s viewing time ($M = 10.47$, $SE = .25$, CI$_{95}$[9.98, 10.97]). Thereafter, no significant correlations existed between viewing time and any of the dependent variables: correct recall, $r(61) = .10$, $p = .43$, errors, $r(61) = -.02$, $p = .86$, accuracy, $r(61) = .09$, $p = .47$, and grainsize precision, $r(61) = -.02$, $p = .89$. The sub-set of data was used in subsequent analyses, including derivation of the descriptive statistics that are presented in Table 3.2.

### 3.2.3.2  *How the naivety instruction impacted report informativeness and accuracy*

Focusing on the quantity dimension, as expected, the naivety instruction increased correct recall, $F(1, 59) = 6.07$, $p = .02$, $\eta^2 = .09$. Interviews that used the instruction, had a medium to large sized effect on producing more correct details ($M = 21.50$, $SE = 1.13$, CI$_{95}$[19.39, 23.71]) compared to the interviews devoid of the naivety instruction ($M = 17.03$, $SE = 1.44$, CI$_{95}$[14.32, 19.66]). Surprisingly, and contrary to what was expected, the naivety instruction did not influence errors, $F(1, 59) = 1.36$, $p = .25$, $\eta^2 = .02$.

Turning to the grainsize dimension, also as expected, the naivety instruction influenced grainsize precision, $F(1, 59) = 8.37$, $p = .01$, $\eta^2 = .12$. Interviews with the naivety instruction, had a large sized effect on producing finer-grained reports ($M = 0.48$, $SE = 0.04$, CI$_{95}$[0.39, 0.56]), than the interviews without the instruction ($M = 0.31$, $SE = 0.04$, CI$_{95}$[0.24, 0.39]).

In terms of report accuracy, and contrary to what was expected, the naivety instruction did not influence recall accuracy, $F(1, 59) = 0.07$, $p = .80$, $\eta^2 < .01$. As an aside, when report accuracy is collapsed across interview condition, it was interesting to find that the reports about the film ($N = 81$, $M = .94$, $SE < .01$, CI$_{95}$[.94, .95]) were more accurate than the reports about the picture ($N = 61$, $M = .89$, $SE = .01$, CI$_{95}$[.87, .91]). This is examined in more detail in

Overall, in producing more informative reports about the picture, the naivety instruction elicited finer-grained information that contained more correct details but not errors. These results support the idea that the naivety instruction led participants to reduce their report threshold. However, the expected trade-off in report accuracy did not occur. As for the film stimulus, this finding has broader theoretical and practical implications that will be addressed in the General Discussion (see Chapter 6). The next section examines why the naivety instruction did not produce similar informativeness effects with the film stimulus.

### 3.2.4  Naivety instruction and informativeness variance across stimuli

When the interviewer made a statement of naivety, participants produced more informative reports about the picture but not about the film. On the face of it, this could suggest that the naivety instruction only affects report informativeness for pictures. However, I will discuss an alternative account in order to rule it out, before exploring more interesting reasons to explain the variance observed in report informativeness across stimuli.

The alternative account relates to the categories of information that could be reported on for each stimulus. Notably, there were three categories of information that could be reported on for the film (Person, Surrounding, and Action), and only two categories that could be reported on for the picture (Person and Surrounding). The following briefly describes each category (see Appendices C to F for more detail).

Person information comprised descriptions about what a person, in the film or picture, looked like. Surrounding information comprised descriptions about what the surroundings (e.g., streetscape), and objects (e.g., vehicles) within the surroundings, looked like. Action information comprised descriptions about what a person did or said. There was no Action category for the picture stimulus because it was a still shot of a scene.

By having the additional Action category of information to report on for the film, this might have moderated the report informativeness response to the statement of naivety. To explore this, Action scores were removed, and inferential analyses were re-calculated for the film stimulus, using one-way ANOVA in IBM SPSS Statistics 21 (collapsed across the mental-reinstatement-of-context instruction). The impact of the naivety instruction on report informativeness for the film, remained unchanged in both the quantity and grainsize dimensions (all $F$s < 1). This suggests that the additional category of information (i.e., Action), available for participants to report on for the film, does not account for the variance in report informativeness observed for the naivety statement across stimuli. This leads me to comment on two possible accounts that might explain the results instead.

Briefly, one account relates to a potential methodological confound that may have moderated participants' belief in the interviewer's statement of naivety about the film but not the picture. This is discussed in more detail in Section 3.3.2.1 (see p. 70). The second account is theoretically motivated because there is a possibility that the naivety instruction makes a similar demand on report informativeness, as the report-detail instruction. In the present study, the report-detail instruction – to describe everything in as much detail as possible – was used in all film interviews but it was omitted in the picture interviews. The implication that the report-detail instruction might have on moderating the naivety instruction, is discussed in more detail in Section 3.3.2.2 (see p. 72).

### 3.2.5 Accuracy variance across stimuli

The finding that report accuracy (collapsed across interview condition) was lower for the picture ($N = 61$, $M = .89$, $SE = .01$, $CI_{95}$[.87, .91]) than it was for the film ($N = 81$, $M = .94$, $SE < .01$, $CI_{95}$[.94, .95]), was unexpected. Again, having the additional Action category of information to report on for the film, might account for this variance in report accuracy. To explore this, Action scores were removed and report accuracy was re-calculated for the film

stimulus (collapsed across interview condition). Report accuracy remained higher for the film ($M = .94$, $SE = .01$, $CI_{95}[.92, .95]$) than the picture. This suggests that the additional category of information, available for participants to report on for the film, does not account for the variance in report accuracy observed across stimuli. This is an interesting finding because it is consistent with an alternative explanation that is related to the instruction not to guess.

The do-not-guess instruction was present in all film interviews yet it was absent in the picture interviews. This is notable because, in response to accuracy incentives, people can improve their recall accuracy by increasing their report threshold (i.e., Koriat & Goldsmith, 1996). Similarly, the do-not-guess instruction might have incentivised greater report accuracy in the film interviews, by encouraging participant's to increase their report threshold. This idea was investigated in the third study (see Chapter 5).

### 3.3    Discussion

A key finding was that the mental-reinstatement-of-context instruction does not impact grainsize precision but that the naivety instruction does impact this aspect of narrative recall. This is an important outcome of the study because grainsize regulation is assumed to be a metacognitive control mechanism (i.e., Goldsmith et al., 2013). Therefore, it is possible to make inferences about how metacognition mediates narrative recall, when instructional demands are made with the mental-reinstatement-of-context and naivety instructions.

Briefly, the results suggest that the mechanism to produce an informative report with the mental-reinstatement-of-context instruction, does not involve the decision to report. This is consistent with the idea that the mental-reinstatement-of-context instruction might improve monitoring performance, as theorised in Section 2.3.1 (see p. 26). In contrast, the results suggest that the mechanism that produces an informative report with the naivety instruction, does involve the decision to report. This is consistent with the idea, theorised in Section 2.3.2 (see p. 32), that the naivety instruction encourages a witness to violate communication norms

and reduce their report threshold. The remainder of this chapter will discuss the empirical findings in greater detail, and will focus on each instruction in turn.

### 3.3.1 Impact of the mental-reinstatement-of-context instruction on monitoring

If the mental-reinstatement-of-context instruction improves the ability to discriminate correct information from errors, then participants were expected to report more correct details and not errors (see Section 2.3.1, p. 26). Participants not only produced a greater quantity of correct details, they also reported more errors (albeit a small increase). This increase in errors might be unique to the present research because other studies (Dando, Wilcock, & Milne, 2009; Dietze et al., 2012) have not found that the mental-reinstatement-of-context instruction produces more errors. However, the errors observed raise questions about the sensitivity of the monitoring mechanism when a witness mentally reinstates the context of the crime. Importantly, the mechanism must be sensitive enough to assign appropriately low confidence to errors so that they may be withheld from reporting.

Mnemonic assistance (by the mental-reinstatement-of-context instruction) is assumed to increase the amount of information retrieved from memory, however, more errors might also be retrieved if the retrieval mechanism is imperfect. One way that these errors might occur is if the memory cues, self-generated by mentally reinstating the context of the crime, encourage schema-consistent intrusions (e.g., Greenberg, Westcott, & Bailey, 1998). If the monitoring mechanism is insensitive to these additional errors, and does not assign them appropriately low confidence, then they might be reported (this is discussed further in Section 3.3.1.1, p. 66). Alternatively, mnemonic assistance might help create a monitoring illusion, such that all information (correct and errors) retrieved from memory is monitored more favourably as being correct. One way this illusion might occur is if information comes to mind more easily and inflate the confidence assigned to it during monitoring (i.e., Kelley & Lindsay, 1993). In this manner, the monitoring mechanism might be sensitive to retrieval

fluency (e.g., Alter & Oppenheimer, 2009) but not errors (this is discussed further in Section 3.3.1.2, see p. 67). In summary, it is an important issue to consider how the mental-reinstatement-of-context instruction impacts monitoring sensitivity so that errors may be withheld from testimony.

### 3.3.1.1   Self-generated cues and schema-consistent intrusions

The mental-reinstatement-of-context instruction assists recall by guiding the witness to probe their memory of the crime with self-generated cues. Cues are self-generated because the interviewer instructs the witness to recollect their thoughts and feelings experienced at the time of the event. Although speculative, these self-generated cues might produce schema-consistent intrusions (e.g., Greenberg et al., 1998), that are recall errors associated with the witness' schemas. For example, a robbery schema might cause a witness to erroneously recall a gun if their schema rules that robbers use guns (Tuckey & Brewer, 2003). How memory retrieval cues are generated might have important implications for monitoring sensitivity and recall performance.

The mental-reinstatement-of-context instruction was designed to operationalise the encoding specificity principle (Geiselman et al., 1984). The effectiveness of this memory principle depends on the amount of feature overlap between the encoded event and the memory retrieval cues (Flexser & Tulving, 1978). In other words, recall will only improve to the extent that retrieval cues are diagnostic of the targeted memory traces (i.e., cue distinctiveness; Pansky, Koriat, & Goldsmith, 2005). Typically, the encoding specificity effect is demonstrated with cues that have been externally generated by the researcher (e.g., cue words; Zeelenberg, 2005). When cues are generated in this way, there is a greater possibility that the cues are diagnostic of targeted traces because the cues replicate, in whole (e.g., a word) or in part (e.g., a word stem), the stimuli learnt (e.g., word list).

In contrast, the mental-reinstatement-of-context instruction helps a witness probe their memory traces with self-generated cues. These self-generated cues appear diagnostic of the memory trace, as evidenced by the efficacy of the mental-reinstatement-of-context instruction (e.g., Dando, Wilcock, & Milne, 2009; Dietze et al., 2012; Emmett et al., 2006). However, some cues might be non-diagnostic of the memory trace, and elicit more errors due to schema-consistent intrusions. Perhaps, the film stimulus used in the present research was unique in producing additional errors as schema-consistent intrusions[16]. An area for future research is to examine if self-generated cues produce schema-consistent intrusions, and under what conditions these are most problematic for the efficacy of the mental-reinstatement-of-context instruction.

### 3.3.1.2 *The retrieval fluency heuristic and inflated confidence*

The mental-reinstatement-of-context instruction might bring information to mind more easily and activate the retrieval fluency heuristic. When this heuristic is engaged, recall is mediated by monitoring processes (Pansky et al., 2005) because information that comes to mind easily, inflates metacognitive judgements such as confidence, familiarity and liking (Alter & Oppenheimer, 2009, for review), or rightness (Thompson, Turner, et al., 2013). The stronger these judgements are, the less people engage in assessing the quality of their memory output (e.g.,Thompson, Ackerman, et al., 2013). Although fluently retrieved answers are more likely to be correct, inflated confidence can create a monitoring illusion that fluently recalled errors are also correct (Pansky et al., 2005). Additional analyses were done to see if the

---

[16] Anecdotally, schema-consistent intrusions occurred in the study. Some participants erroneously recalled the thief wore a hooded-jumper (television news footage frequently shows thieves wear hooded-jumpers) and the waiter wore an apron (waiters typically wear aprons). Moreover, schema-consistent intrusions have been found to occur during a Cognitive Interview (Geiselman & Callot, 1990). These intrusions may be moderated with the change-order instruction, one of the original Cognitive Interview mnemonics that was designed to reduce schema-consistent intrusions, by making an witness recall events in reverse order (Geiselman, Fisher, MacKinnon, et al., 1986). However, as noted in the Introduction (Chapter 1) the change-order instruction has been found to have little to no effect on recall, and only lengthened the time taken to conduct an interview (e.g., Bensi et al., 2011).

mental-reinstatement-of-context instruction produced fluent recall, and infer the activation of the retrieval fluency heuristic.

### 3.3.1.3 How the mental-reinstatement-of-context instruction impacted retrieval fluency

If the mental-reinstatement-of-context instruction produces fluent recall, then shorter pauses in recall might exist when information is retrieved more fluently. Although speech pauses serve to catch breath, they might also serve to retrieve information from memory. If pauses are used to retrieve information from memory, and information comes to mind more easily, then it is expected that more information will be recalled per unit of reporting time (i.e., recall rate). However, part of the mental-reinstatement-of-context instruction it to tell the witness to get a good picture in mind (of the witnessed event) before giving their account, and this might confound their recall rate. Empirically, if participants took the time to get a good mental picture, then their response latency should be longer than participants who were not interviewed with the mental-reinstatement-of-context instruction.

Response latency was measured with the time elapsed between the interviewer saying his last word (of the interview instruction) and the participant responding with their first word. Interestingly, participants were slower to initiate their report in the interviews that used the mental-reinstatement-of-context instruction ($Mdn = 3.08$ s, $SE = 0.29$, $CI_{95}[2.54, 3.76]$), than the interviews without the instruction ($Mdn = 1.91$ s, $SE = 0.21$, $CI_{95}[1.63, 2.38]$), $U = 505.50$, $n_{MRC\text{-}absent} = 41$, $n_{MRC\text{-}present} = 40$, $z = -2.97$, $p < .01$, $CI_{95}[-1.78, -0.40]$, $r = -.33$. This suggests that participants did take the time to get a good mental picture, before giving their report. Therefore, response latency was subtracted from reporting time, to remove this confound from the analysis of participants' recall rate.

Returning to the primary analysis of interest, if the mental-reinstatement-of-context instruction helps retrieve information more fluently, then participants' recall rate (number of details per second of reporting time) was expected to be faster. Contrary to this prediction,

participants' recall rate was slower in the interviews that used the mental-reinstatement-of-context instruction ($M = 0.65$, $SE = 0.03$, CI$_{95}$[0.59, 0.70]), than the interviews without the instruction ($M = 0.76$, $SE = 0.04$, CI$_{95}$[0.68, 0.83]), $t(79) = 2.28$, $p = .03$, CI$_{95}$[0.01, 0.20], $d = 0.51$. However, this finding might be unreliable for two reasons (and perhaps more).

First, participants gave longer reports when responding to the mental-reinstatement-of-context instruction. The information contained in these reports included information that was repeated, but not scored. Therefore, given that the participants' reporting time was measured as the total time taken to report all of the information (including repeated information), then their reporting time would have been inflated. Second, the interviewer gave neutral feedback (i.e., "mm mm", "ok") more abundantly in longer reports because participants said more. The accumulated time to give this feedback might have further inflated participants reporting time. Given the complexity with interpreting what the additional analyses might mean, an area for future research is to determine if the mental-reinstatement-of-context instruction activates the retrieval fluency heuristic, and what impact this has on recall performance.

### 3.3.2   Impact of the naivety instruction on the decision to report

The present study is the first known empirical investigation to demonstrate the efficacy of the naivety instruction with adults. Although limited to the picture stimulus, the results suggest that a statement of naivety produces finer-grained reports that contain a greater quantity of correct details. In terms of the quantity dimension, this result is important because it replicates the finding from the child witness literature (i.e., see review by Brubacher et al., 2015). However, there remains a gap in understanding how the naivety instruction impacts the children's grainsize precision and so this is an area for future research.

The present findings are theoretically important because they support the idea that a naivety statement will encourage a witness to violate communication norms and produce more informative reports (i.e., Milne, 2004). Moreover, participants regulated their response

grainsize (i.e., Goldsmith et al., 2002), supporting the idea that the naivety instruction impacts the decision to report (see Section 2.3.2.1, p. 36). Curiously, however, the naivety instruction did not produce more errors. Why this occurred will be discussed in more detail in Chapter 4 (see Section 4.3.3, p. 98), in conjunction with the results of Experiment 2.

Unfortunately, a limitation with the study was that the naivety instruction did not produce more informative reports about the film stimulus. The following discussion will explore two possible explanations for why this might have occurred. One account relates to a potential methodological confound that might have diminished the participants' belief in the interviewer's statement of naivety. The other account is theoretically motivated, and examines how the naivety and report-detail instructions[17] might have similarly encouraged participants to violate communications norms.

### 3.3.2.1   Belief in the interviewer's statement of naivety

When communication norms are challenged with a naivety statement (i.e., Milne & Bull, 1999), the report threshold is the assumed mechanism to communicate finer-grained descriptions (i.e., Goldsmith et al., 2002). However, an effective statement of naivety requires that the witness actually believes the interviewer does not know what happened, otherwise the witness might not violate communication norms. Therefore, belief in the interviewer's stated lack of knowledge, is assumed to be the key mechanism to violate communication norms in the first instance. Certainly, there are reasons a witness might assume the interviewer has prior knowledge of the event and disbelieve their statement of naivety. For example, schemas and frames (abstract representations of a scene or episode) form the basis of shared knowledge and might impact the assumed level of knowledge (e.g., W. F. Brewer & Treyens, 1981; Schuurmans & Vandierendonck, 1985; Tuckey & Brewer, 2003; Vandierendonck &

---

[17] The report-detail instruction was present in all film interviews but it was absent in the picture interviews.

Damme, 1988). However, another reason that has implications for the psychological realism of the naivety instruction in the present study, is that the witness might assume the interviewer knows what happened from interviewing multiple witnesses.

The psychological realism of the naivety instruction might not have occurred with the film stimulus because there were methodological differences between film and picture interviews. Specifically, only one film was referred to in the Information Sheet, by myself (in outlining the upcoming task), and by the interviewer ("I understand you have watched a film of a crime"). Participants might have thought it impossible that the interviewer had no knowledge of the crime, when he said "I don't know what you witnessed". They might have assumed he had built his knowledge of the crime through interviewing other participants and, consequently, did not believe his stated lack of knowledge.

In contrast, the psychological realism of the naivety instruction appears to have occurred with the picture stimulus because participants produced more informative reports. The key methodological difference was that the interviewer had a pile of 10 pictures, face down on the table, from which he lifted one to show participants. Although the same picture was shown to every participant, the card pile might have convinced the participants to believe the interviewer when he said "I don't know what is on the card".

In summary, participants might have believed the interviewer's stated lack of knowledge to be more real for the picture than for the film. This might be why, when the interviewer made his statement of naivety, more informative reports were produced for the picture but not the film. Experiment 2 (see Chapter 4) explored this issue and examined if a witness must believe the interviewer's stated naivety to produce a more informative report.

### *3.3.2.2 The report-detail instruction and the decision to report*

The report-detail instruction – to describe everything in as much detail as possible – was used in all film interviews but it was omitted in the picture interviews. Theoretically, the demand that this instruction makes on a witness, might be similar to the demand made by the naivety instruction. That is, the report-detail instruction might also encourage a witness to violate communication norms, to produce a more informative report. Supporting this idea, is the theoretical rationale made by Bekerian and Dennett (1993) for the report-everything instruction. To recap, as highlighted in the Introduction (see Section 1.1.1.2, p. 4), the report-everything instruction was developed to assist memory retrieval by applying the multi-component theory of memory to witness interviews (Geiselman et al., 1984). However, Bekerian and Dennett (1993) suggested that the strategy to report everything "is compatible with the work on signal detection theory, which suggests that a person's willingness to report an event will depend upon the particular response (or confidence level) being adopted" (p. 277). In other words, a witness' willingness to report depends on their report threshold. This theoretical rationale is relevant to the report-detail instruction because, similar to the report-everything instruction, the witness is instructed to report everything about the crime[18]. Moreover, a witness' willingness to report might be consistent with communication norms because they might be less willing to report additional information if they think the interviewer already knows it (i.e., Schuurmans & Vandierendonck, 1985).

If the report-detail and naivety instructions made similar demands on participants in the present study, then this might be the reason why the naivety instruction did not produce more informative reports about the film. The participants, in describing the film in as much detail as possible, might have violated communication norms and reduced their report

---

[18] The report-everything instruction tells an eyewitness to report everything that comes to mind no matter how trivial or unimportant it might seem (Fisher & Geiselman, 1992).

thresholds as much as possible. Consequently, the naivety instruction may have had no influence on the decision to report. In other words, the report-detail instruction might have moderated the naivety instruction and produced a ceiling effect on report informativeness. In contrast, the naivety instruction might have impacted report informativeness during the picture interviews because the moderating effect of the report-detail instruction was absent. Experiment 2 (see Chapter 4) explored the theoretical relevance of this argument.

### 3.4    Summary of Key Findings

- Instructional demands with the mental-reinstatement-of-context and naivety instructions produced more informative reports without compromising accuracy.

- The mental-reinstatement-of-context instruction produced more informative reports in the quantity dimension and, importantly, grainsize precision did not change. This is consistent with the idea that the mental-reinstatement-of-context instruction impacts monitoring performance and not the report threshold. However, participants produced more errors, suggesting that the instruction influenced monitoring sensitivity to errors. That is, the monitoring mechanism may not have assigned appropriately low confidence to errors so that they could be withheld from reporting.

- Specifically, it is possible that mnemonic assistance retrieved additional errors because the retrieval mechanism was imperfect and the monitoring mechanism was insensitive to these errors (e.g., schema-consistent intrusion). Alternatively, mnemonic assistance might have brought information to mind more easily and inflated the confidence assigned by the monitoring mechanism to all information, including errors (e.g., retrieval fluency heuristic).

- The naivety instruction produced more informative reports (about the picture but not the film) in the quantity and grainsize (i.e., finer-grained) dimensions, suggesting that the

report threshold was affected. This is consistent with the idea that the naivety instruction encourages a witness to violate communication norms and reduce their report threshold.

- The naivety instruction might not have influenced film reports if there was a methodological confound that moderated participants' belief in the statement of naivety. Alternatively, the report-detail instruction might have moderated the naivety instruction and produced a ceiling effect on report informativeness. These alternative accounts are examined in the next study (Chapter 4).

**CHAPTER 4**

**Experiment 2**

This chapter presents the second empirical study examining how differing instructional demands influence narrative recall, and the insights into metacognition this provides. The overarching purpose of the study was to apply Koriat and Goldsmith's (1996) framework of metacognition to naturalistic witness reports. The model of metacognition that I developed for the investigative interview setting (see Figure 1.1, p. 12), provided the context for the study. To recap, the model conceptualises the relationship between the interviewer, whose goal is to obtain an informative and accurate report (with the aid of interview instructions), and the witness, whose narrative report (in response to the interview instructions) is mediated by metacognitive processes.

Specifically, the purpose of the present study was to understand why the naivety instruction did not influence the film reports in Experiment 1 (see Chapter 3). The reader is reminded that a statement of naivety might encourage a witness to violate communication norms and produce finer-grained descriptions (Milne, 2004), and the assumed mechanism to achieve this is the report threshold (i.e., Goldsmith et al., 2002). Consistent with this, the first study found that the naivety instruction produced finer-grained reports about the picture stimulus, and these reports contained a greater quantity of details. In contrast, the naivety instruction did not influence either the quantity or grainsize dimensions of report informativeness about the film. To understand why this occurred, the present experiment had two aims.

One aim was to determine if the efficacy of the naivety instruction is influenced by the witness' belief in the interviewer's stated naivety. A methodological confound potentially existed in the first study to moderate the participants' belief in the interviewer's naivety. It is

important that participants believed the interviewer when he said "I don't know what you

witnessed" because this belief is the proposed mechanism that may lead a witness to reduce

their report threshold. Participants in Experiment 1 might have been aware that the same film

was shown to every participant and, therefore, assumed the interviewer knew the contents of

the film from preceding interviews. If so, the participants might not have believed the

interviewer's stated lack of knowledge about the film. In other words, the participants might

have believed it unnecessary to reduce their report threshold and, thereby, produce a more

informative report. In contrast, the participants might have presumed other participants saw

different pictures because the interviewer drew the stimulus picture from a pile of pictures.

The picture pile might have helped ensure participants did not assume the interviewer had

prior knowledge about the scene in the picture. If so, participants might have believed the

interviewer's stated naivety and, thereby, reduced their report threshold to produce a more

informative report.

In the present study, I took the approach to use weak and strong versions of the

naivety instruction, by varying the naivety statement. The aim was to observe if participants

believed more strongly in a stronger statement of naivety, and if a stronger belief manifested

greater report informativeness. Participants' belief was measured with a post-interview

question. Participants were asked to rate how strongly they believed the interviewer had no

knowledge of the crime prior to their interview (see Section 4.1.2, p. 78, for more detail on

the materials used).

The second aim of the present study was to determine if the report-detail instruction

might also lead a witness to reduce their report threshold. This is important to know because it

would provide evidence that the report-detail and naivety instructions make similar demands

on a witness. If so, this would provide a theoretical explanation for why the naivety

instruction did not influence the film reports in Experiment 1 (Chapter 3). In other words, the

report-detail instruction might have led participants to reduce their report threshold as far as they could and this (a) produced a ceiling effect on report informativeness, and (b) moderated the impact of the naivety instruction. In the present study, I used a factorial design to observe how the report-detail and naivety instructions influenced narrative performance. The aim was to observe if an interaction on report informativeness could be observed and clarify the idea that both instructions make similar demands on a witness.

## 4.1 Method

### 4.1.1 Participants

The sample comprised 134 participants recruited from the general university community or from first year Psychology students. General university participants received AUD$10 and first year Psychology students received course credit. Informed written consent, including the permission to be video-recorded, was obtained from all participants. Three participants were excluded from the study because equipment malfunctioned (1 person), and participants did not follow instructions (2 people). The mean age of the remaining 131 participants was 27.21 years and ranged from 18 – 68 years. The sample[19] comprised 52 males ($M = 28.25$ years, $SD = 12.72$), 77 females ($M = 26.38$ years, $SD = 10.08$), and two transgender ($M = 32.00$ years, $SD = 9.90$). All remaining participants had normal or corrected-to-normal vision and were fluent in English.

---

[19] Results from Experiment 1 found that the naivety instruction had a medium to large sized effect on recall for the picture stimulus ($\eta^2 = .09$ for correct recall, $\eta^2 = .12$ for grainsize precision). If recall is similarly impacted by the report-detail instruction, then it was expected that the same effect size would occur for this instruction. A power analysis suggested that the present study required approximately 25 participants per interview condition (using power of .80 and alpha of .05; Cohen, 1992). Therefore, a total sample of 150 participants was needed. Unfortunately, very slow participation and restricted access to the interviewer, reduced the final sample size.

#### *4.1.1.1 Interviewer*

The interviewer (54-year-old male) was an expert in investigative interviewing, being an experienced detective (Chief Inspector level) with the South Australia Police. He volunteered his time and was not familiar with the experimental aims of the research. As with the first study, the interviewer familiarised himself with the interview instructions, printed on a separate card for each experimental condition. He practiced reading the cards out verbatim prior to interviewing participants. This practice helped to maintain experimental control.

To help build rapport, the interviewer was friendly with participants (i.e., made eye contact and smiled) and made casual conversation by asking how their day and/or studies were going. For every interview, the interviewer gave instructions slowly and coherently to the participant. During participant recall, the interviewer used verbal acknowledgments (e.g., "uh-huh", "yep") and head nods minimally since his professional training in investigative interviewing discourages overusing these active listening techniques. Instead, he focused on maintaining eye-contact throughout participant recall.

### 4.1.2   Materials

All of the materials used were similar to the first study, including the film stimulus. However, the picture stimulus was omitted. In addition, a post-interview questionnaire was used and participants were asked to fill it in after their interview was completed. The questionnaire comprised six questions and participants used a Likert scale to respond to each question. One question was of primary interest, and it aimed to measure how strongly participants believed the interviewer's naivety about the content of the film. It was expected that participants' belief would be stronger when the interviewer made a naivety statement than when he did not state his naivety.

### 4.1.3 Procedure

The same procedure, outlined in Chapter 3 (see Section 3.1.3, p. 43), was used with the following exceptions. After a 30 min distraction period, the participants were taken to the room where the interviewer was waiting. The interview room was a located down a hallway, 20 m away from the laboratory. In the first study, the interview room was annexed to the same laboratory where the film was viewed. The close proximity of the interview room might have contributed to a potential methodological confound that moderated participant's belief in the interviewer's statement of naivety. This is because, through a glass window in the door to the interview room, it was possible for participants to see another person being interviewed. This might have led participants to assume that the interviewer already had knowledge about the film and disbelieve his naivety claim.

Participants were interviewed with one of six interview conditions. The report-detail (RDT) and naivety (NVT) instructions were manipulated in a two-way factorial, between-subjects design to elicit a narrative report. The six interview conditions comprised: (a) RDT-absent, NVT-absent; (b) RDT-absent, NVT-weak; (c) RDT-absent, NVT-strong; (d) RDT-present, NVT-absent; (e) RDT-present, NVT-weak; and (f) RDT-present, NVT-strong. These conditions are explained next.

Participants in the control condition (RDT-absent, NVT-absent) were asked to describe what they saw and heard on the film. When the report-detail instruction was present, the interviewer stated: "Tell me everything that you saw and heard on the video and describe this to me in as much detail as you can".

The naivety instruction had three levels (i.e., absent, weak, strong) to strengthen participants' belief in the interviewer's statement of naivety. A naivety statement was absent from the base level (NVT-absent). In the weak version of the naivety instruction (NVT-weak), the interviewer stated: "I have not seen it (the film) myself so I do not know what you

have witnessed. You have all the information so I only have your description to rely on". Notably, this was the same statement of naivety that was used in Experiment 1 (Chapter 3).

In the strong version of the naivety instruction, the interviewer stated: "I understand you have watched one of several films showing different crimes. I have not seen any of the films myself and I don't know which one you have watched. I'm one of several interviewers and this is my first interview of the day so I have not heard from other people what they have watched. I do not know what you witnessed. You have all the information so I only have your description to rely on". Of note, the same film was shown to every participant, however, the strong statement of naivety was deceptive, to communicate more strongly that the interviewer did not know the content of the film.

To aid the deception, I also used different instructions and props. In the base (NVT-absent) and weak naivety (NVT-weak) conditions, I explained to the participant that they would watch a short film. In the strong naivety condition, I explained to the participant that they would watch one of several short films. Further, for the strong naivety condition, eight CD cases (labelled "Crime 1", "Crime 2', etc.) were stacked without particular order next to the computer screen that the participant used to watch the film. No CD cases were placed next to the computer screen in the base and weak naivety conditions.

Participants were randomly assigned to one of the six interview conditions, however, some restriction was made with the strong naivety condition. The strong version of the naivety statement was generally allocated to participants arriving before midday (although morning participants could be interviewed with any of the other five interview conditions). This restriction was made to help participants believe that their interview, as stated by the interviewer, was the interviewer's first for the day.

After a response was elicited with one of the interview conditions, when the participant appeared to have finished volunteering information, the interviewer waited for a short period and then asked if there was anything else they could add. The participant's free-narrative response was concluded after additional information was volunteered or they declined to add anything else. The interviewer then asked three open questions that related to the object stolen, the thief's clothing, and the vehicle parked outside the restaurant. These questions were used to elicit additional narrative responses about the film. Whilst this thesis does not examine the impact of question format on recall – free narratives produce greater accuracy than cued responses (i.e., Evans & Fisher, 2011) – the purpose of the questions was to avoid a ceiling effect on report accuracy (as was observed in Experiment 1, see Section 3.2.2.2, p. 58). Specifically, the purpose was to help observe a trade-off in accuracy that is expected to occur when a more informative report is produced, in either the quantity (i.e., Koriat & Goldsmith, 1994) or grainsize (i.e., Yaniv & Foster, 1995) dimensions of report informativeness. Importantly, open questions were used instead of closed questions because open questions invite narrative responses (i.e., Fisher, 2010), in keeping with the focus on narrative recall. The order of the open questions was randomised across the interview conditions. The final narrative report included the free-narrative response and the additional narrative responses to the three open questions. Participants were not interrupted during their reports. After interviewing was finished, each participant was debriefed.

The actual interview instructions and open questions used in the present study, can be found in Appendix B. The instructions and open questions were read out verbatim to participants. Every interview was video-recorded for later transcription and scoring.

### 4.1.4   Dependent variable to assess belief in the statement of naivety

After their interview, each participant made a rating of how strongly they believed the interviewer had no knowledge of the film before commencing their interview. Participants

could select a rating from '1' (the interviewer had no knowledge) to '10' (the interviewer had complete knowledge). This rating was used to assess participants' belief in the statement of naivety made by the interviewer. Ratings were reverse-scored for descriptive statistics and inferential analysis.

### 4.1.5   Dependent variables to assess report informativeness and accuracy

The same dependent variables outlined in Chapter 3 (see Section 3.1.4, p. 45) were used in the present study. Report informativeness was assessed with quantity and grainsize variables. Quantity variables included the number of correct details and errors recalled. The grainsize variable was grainsize precision and it was calculated by dividing the number of fine-grained chunks by the total number of grainsize chunks recalled (fine-grained + coarse-grained). To recap, 'grainsize chunk' is a non-technical label applied to a statement of related information (e.g., a statement about the thief's upper-body clothing), and grainsize chunks could be either coarse-grained or fine-grained statements. Report accuracy was assessed as the proportion of details recalled that were correct (correct / correct + errors). How these variables were coded is described next.

### 4.1.6   Scoring narrative recall

Quantity and grainsize scoring of the free-narrative response followed the same procedure outlined in Chapter 3 (see Section 3.1.5, p. 46). However, the narrative responses to the three open questions (related to the object stolen, thief's clothing, and the vehicle parked outside the restaurant) were scored only if the information had not been stated during the free-narrative response. For example, if "jacket" was stated during the free-narrative, and "black jacket" was given in response to the question about the thief's clothing, then "black" was new information that was scored. New information was scored as correct or erroneous, following the same rules outlined in Chapter 3 (Section 3.1.5). The grainsize preference rule was also applied (i.e., finer-grained responses were scored in preference to coarser-grained responses).

For example, if "4WD" was stated during the free-narrative, and "Toyota Landcruiser" was stated in response to the question about the vehicle, then "Toyota Landcruiser" was scored in preference to "4WD".

The scores obtained from the free-narrative and open question responses were combined to derive narrative performance for the number of correct details and errors recalled, and the number of grainsize chunks recalled (this comprised a tally of reported coarse-grained and fine-grained chunks). From these, narrative performance for accuracy (correct / correct + errors) and grainsize precision (proportion of grainsize chunks that were fine-grained) was calculated. The reader is referred to Appendix C and E for further detail on the scoring rules, coding keys and examples.

### 4.1.6.1   Inter-rater reliability

All interview transcripts were scored by myself, and twenty-seven transcripts (20%), randomly selected, were scored independently by a second coder. As outlined in Chapter 3 (Section 3.1.5, p. 46), the interview instructions were omitted from the transcripts during transcription to ensure scoring was done blind to the interview condition.

The inter-rater reliability was assessed by percentage agreement (see Section 3.1.5.3, p. 49, for why the kappa statistic was not used). For the total quantity of information (correct + errors), an inter-rater reliability of 80.4% was established. For the total number of grainsize chunks (fine-grained + coarse-grained), an inter-rater reliability of 96.0% was established. As outlined in Chapter 3 (Section 3.1.5.3, p. 49), coders did not discuss disagreements for quantity scores but we did discuss disagreements for grainsize scores. My scores, except when they were changed during discussion with the second coder, were retained for the main analysis.

## 4.2 Results

Descriptive statistics for the different interview conditions are presented in Table 4.1.

The table shows that most variables were influenced by interview condition, except accuracy.

Table 4.1

*Mean Recall and Belief [and 95% confidence intervals] by Interview Condition*

| | Interview Condition | | | | | |
|---|---|---|---|---|---|---|
| | RDT-a NVT-a | RDT-a NVT-w | RDT-a NVT-s | RDT-p NVT-a | RDT-p NVT-w | RDT-p NVT-s |
| | (*n* = 22) | (*n* = 21) | (*n* = 22) | (*n* = 22) | (*n* = 22) | (*n* = 22) |
| Variable | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* |
| Correct | 86.09 (36.70) | 114.86 (44.62) | 100.82 (39.36) | 113.27 (33.85) | 110.68 (42.88) | 118.14 (23.72) |
| | [69.82, 102.36] | [94.55, 135.17] | [83.37, 118.27] | [98.26, 128.28] | [91.67, 129.69] | [107.62, 128.65] |
| Errors | 7.18 (3.36) | 9.24 (5.14) | 7.27 (3.59) | 10.27 (6.52) | 10.27 (6.21) | 8.36 (3.59) |
| | [5.69, 8.67] | [6.90, 11.58] | [5.68, 8.86] | [7.38, 13.16] | [7.52, 13.03] | [6.77, 9.96] |
| Accuracy | .92 (.04) | .92 (.03) | .93 (.04) | .91 (.04) | .91 (.04) | .93 (.03) |
| | [.90, .93] | [.91, .94] | [.91, .94] | [.89, .93] | [.90, .93] | [.92, .95] |
| GS Precision | .44 (.15) | .57 (.14) | .51 (.14) | .51 (.13) | .53 (.12) | .54 (.11) |
| | [.38, .51] | [.51, .63] | [.45, .58] | [.45, .57] | [.47, .58] | [.49, .58] |
| Belief | 4.32 (2.71) | 6.24 (3.37) | 5.91 (2.67) | 4.95 (3.12) | 5.05 (2.95) | 5.68 (1.78) |
| | [3.11, 5.52] | [4.70, 7.77] | [4.72, 7.09] | [3.57, 6.34] | [3.74, 6.35] | [4.89, 6.47] |

*Note.* Correct = number of correct details; Errors = number of errors; Accuracy = proportion of correct details;

GS Precision = proportion of fine-grained recall; Belief = belief rating in the interviewer's stated naivety (higher

number = stronger belief).

RDT-a = report-detail instruction absent; RDT-p = report-detail instruction present; NVT-a = naivety instruction

absent; NVT-w = weak naivety instruction; NVT-s = strong naivety instruction.

### 4.2.1 Data screening

Parametric assumptions for each interview condition were checked using the Shapiro-Wilk normality test and the Levene's test for homogeneity of variance (using the median). Correct recall and grainsize precision met parametric assumptions. All other dependent variables were skewed (in at least one interview condition) with equal variance.

Outliers were examined with standardized data for each interview condition and identified when $z$-scores were less than -2.50 or greater than 2.50. No outliers were found in correct recall or belief but they were found in errors (3 outliers), accuracy (2), and grainsize precision (1).

#### 4.2.1.1 Permutation ANOVA

Permutation ANOVA was used to check if parametric violations and outliers, when present, did not bias the statistical inference that was determined from parametric ANOVA. The same approach was taken as outlined in Chapter 3 (see Section 3.2.1.3, p. 52). For each dependent variable that showed a significant parametric ANOVA result, the permutation $p$-value that was obtained with permutation ANOVA, was less than .05. This suggested that the parametric ANOVA result was reliable and unlikely to be a Type I error.

#### 4.2.1.2 Time of interview

Correlations were explored to assess if morning interviews influenced recall performance (collapsed across interview condition). This was done to determine if recall was affected by restricting the interviews that used the strong version of the naivety instruction, to morning participants. There was no relationship between the time of the day when the interview was conducted and any dependent variable: correct recall, $r(131) = -.05$, $p = .57$; errors, $r(131) = .02$, $p = .84$; accuracy, $r(131) = -.11$, $p = .23$; grainsize precision, $r(131) = -.04$, $p = .71$; and belief, $r(131) = .07$, $p = .41$.

### 4.2.2 Belief in the interviewer's statement of naivety

With these analyses, I wanted to know if the strength of the interviewer's stated naivety influenced participants' belief. Pairwise comparisons were made between the three naivety conditions (i.e., absent, weak, strong) and a Bonferroni adjustment (alpha = .017) was applied to control for familywise error. Data was collapsed across the report-detail factor because the interviewer's knowledge state was not communicated with the report-detail instruction.

A generalised Wilcoxon test was used, following Neuhäuser's (2010) approach, because data was skewed with unequal variance. Neuhäuser's approach uses the Brunner-Munzel test statistic, $W_{BF}$; effect size, $P$ (= P(X<Y) + 0.5*P(X=Y)); and a confidence interval for $P$. If the confidence interval contains the value '.5' then the result is not significant. The test was calculated using the *lawstat* package (version 2.4.1; Gastwirth et al., 2014) in R (version 3.2.2; R Core Team, 2014). For significant effects, point estimates of the median are given with 98.3% confidence intervals. These were calculated using the bias-corrected and accelerated bootstrap method in IBM SPSS Statistics 21, using 2000 permutations. For verbal description of the effect size $P$, $P$ was first converted to Somers' $D$ (2*$P$-1; Newson, 2002) to make it equivalent to effect size $r$ (Ferguson, 2009), and Cohen's (1992) recommendations were adopted for $r$ (i.e., .10 = 'small', .30 = 'medium', .50 = 'large').

Figure 4.1 shows that, as expected, participants had a stronger belief in the interviewer's naivety, when he used either the weak or strong naivety statements, than when he did not state his naivety. Pairwise comparisons revealed a marginally significant, small to medium sized effect, $W_{BF}(76.77) = -2.09$, $p = .04$, $CI_{98.3}[0.23, 0.52]$, $P = .38$, $n_1 = n_2 = 44$, suggesting a stronger belief in the interviewer when he used the strong naivety statement ($Mdn = 6.00$, $SE = 0.40$, $CI_{98.3}[5.00, 6.00]$), than when he did not state his naivety ($Mdn = 5.00$, $SE = 0.84$, $CI_{98.3}[4.00, 6.00]$). Although a nonsignificant trend, there was a small to

medium sized effect, $W_{BF}(83.89) = -1.43$, $p = .16$, $CI_{98.3}[0.26, 0.56]$, $P = .41$, $n_1 = 44$, $n_2 = 43$, suggesting a stronger belief in the interviewer when he used the weak naivety statement (*Mdn* = 6.00, *SE* = 0.70, $CI_{98.3}[4.00, 6.00]$), than when he did not state his naivety (*Mdn* = 5.00, *SE* = 0.84, $CI_{98.3}[4.00, 6.00]$). No difference in belief was found between the weak and strong naivety statements, $W_{BF}(65.34) = -0.25$, $p = .80$, $CI_{98.3}[0.33, 0.64]$, $P = .48$, $n_1 = 43$, $n_2 = 44$. These results suggest that participants believed the interviewer's naivety and their belief was not moderated by the strength of naivety statement.
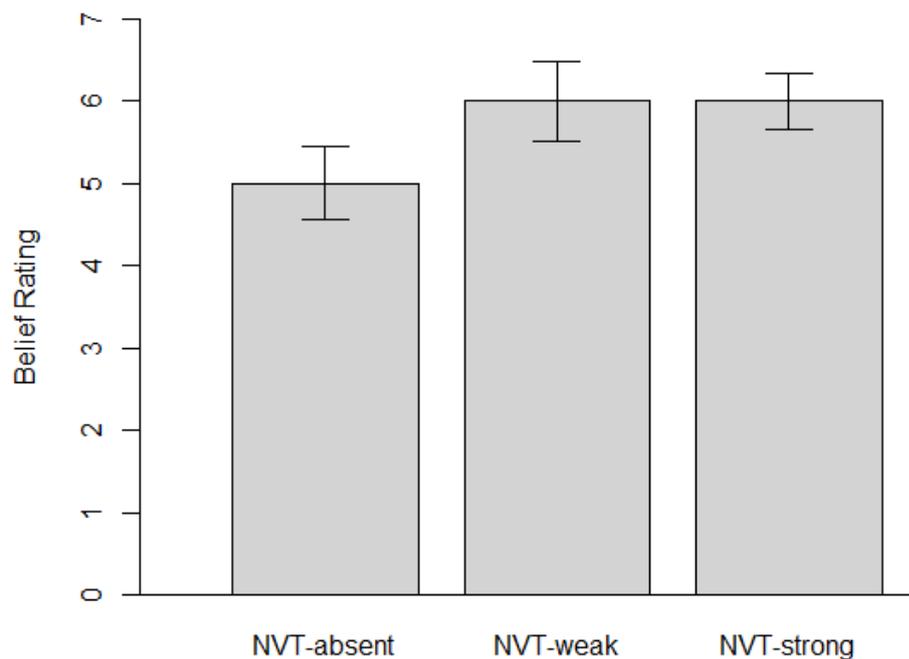


*Figure 4.1*. Median belief (+SE) for each naivety statement (higher rating = stronger belief).

In summary, it was found that the naivety statement impacted participants' belief. This finding is important because it is consistent with the idea that belief (in the naivety statement) may lead a witness to reduce their report threshold. This will be examined in more detail, together with the narrative performance results, in the Discussion (see Section 4.3, p. 96). It was also found that participants' belief was not moderated by the strength of the naivety instruction. This is an important observation because it suggests a methodological confound did not exist in the first study (to moderate participants' belief in the interviewer's naivety

about the film stimulus). I will now turn to the alternative theoretical account that may explain why the naivety instruction did not impact film reports in Experiment 1 (Chapter 3).

### 4.2.3    Narrative performance

In this section, analyses are focused on testing the idea that the report-detail and naivety instructions make similar demands on a witness. Specifically, both instructions might lead a witness to violate communication norms and reduce their report threshold. If so, it is possible that the report-detail instruction produced a ceiling effect on report informativeness, in Experiment 1. The critical test in the present study was to observe if the report-detail instruction moderated the naivety instruction on report informativeness.

In terms of report accuracy, it was expected that a trade-off in accuracy would occur with greater report informativeness, on either the quantity dimension (i.e., Koriat & Goldsmith, 1994) or grainsize dimension (i.e., Yaniv & Foster, 1995). Notably, to help avoid a ceiling effect on report accuracy (that was observed in Experiment 1, see Section 3.2.2.2, p. 58) so that a trade-off could be observed, a series of open questions were added to the interviews in the present study.

Inferential analyses were calculated using univariate factorial ANOVA (alpha level set at .05) in IBM SPSS Statistics 21. Type II sum of squares was used due to the unbalanced design although the interview conditions only differed in size by one person. Effect size, eta squared ($\eta^2$), was calculated using the *lsr* package (version 0.5; Navarro, 2015) in R (version 3.2.2; R Core Team, 2014). For significant effects, point estimates of the mean are given with 95% confidence intervals. These were calculated using the bias-corrected and accelerated bootstrap method in SPSS, using 2000 permutations. Verbal description for eta squared adopted Cohen's (1988) recommendations (i.e., .01 = 'small', .06 = 'medium', .14 = 'large').

### *4.2.3.1 How the report-detail and naivety instructions impacted report informativeness*

In this set of analyses, the main effects were important but I was particularly interested in an interaction between the report-detail and naivety instructions, on report informativeness. Specifically, in the presence of an interaction, I explored the simple main effect for the naivety instruction (a) when the report-detail instruction was absent, and (b) when the report-detail instruction was present. If the report-detail instruction moderates the naivety instruction, then a simple main effect was expected when the report-detail instruction was absent but not when the report-detail instruction was present.

The simple main effect was examined at each level of the report-detail factor (i.e., absent, present), using pairwise comparisons between the three naivety conditions (i.e., absent, weak, strong). A Bonferroni adjustment (alpha = .01) was applied, to control for familywise error across the six pairwise comparisons. The Student's *t* test was used because data met parametric assumptions (effect size was calculated with Cohen's *d*). The test was calculated using the *psych* package (version 1.4.8; Revelle, 2014) in R (version 3.2.2; R Core Team, 2014). For significant effects, point estimates of the mean are given with 99% confidence intervals. These were calculated using the bias-corrected and accelerated bootstrap method in IBM SPSS Statistics 21, using 2000 permutations.

The results are presented separately for the quantity and grainsize dimensions of report informativeness.

### *4.2.3.1.1 Quantity dimension of report informativeness*

As expected, the report-detail instruction increased correct recall, $F(1, 125) = 4.31$, $p = .04$, $\eta^2 = .03$ (see Figure 4.2), and increased recall errors, $F(1, 125) = 4.14$, $p = .04$, $\eta^2 = .03$ (see Figure 4.3). Interviews with the report-detail instruction, had a small to medium sized effect on producing more correct details ($M = 114.03$, $SE = 4.19$, $CI_{95}[105.80, 122.04]$), and a

small to medium sized effect on producing more errors ($M = 9.64$, $SE = 0.69$, CI$_{95}$[8.39, 11.03]), than interviews without the instruction (correct details, $M = 100.37$, $SE = 5.13$, CI$_{95}$[90.31, 109.96]; errors, $M = 7.88$, $SE = 0.51$, CI$_{95}$[6.94, 8.82]).

It was surprising to find that the naivety instruction did not influence correct recall, $F(2, 125) = 1.41$, $p = .25$, $\eta^2 = .02$ (see Figure 4.2) because the statement of naivety increased correct recall in Experiment 1 (i.e., with the picture stimulus, see Section 3.2.3.2, p. 61). It is likely the present result reflects a lack of power with the study because the naivety instruction had an observed power of .30. However, it was curious to find that the naivety instruction did not influence recall errors, $F(2, 125) = 1.68$, $p = .19$, $\eta^2 = .03$ (see Figure 4.3), because this replicates the finding from Experiment 1 (i.e., with the picture stimulus, see Section 3.2.3.2, p. 61) albeit the observed power in the present study was .35.
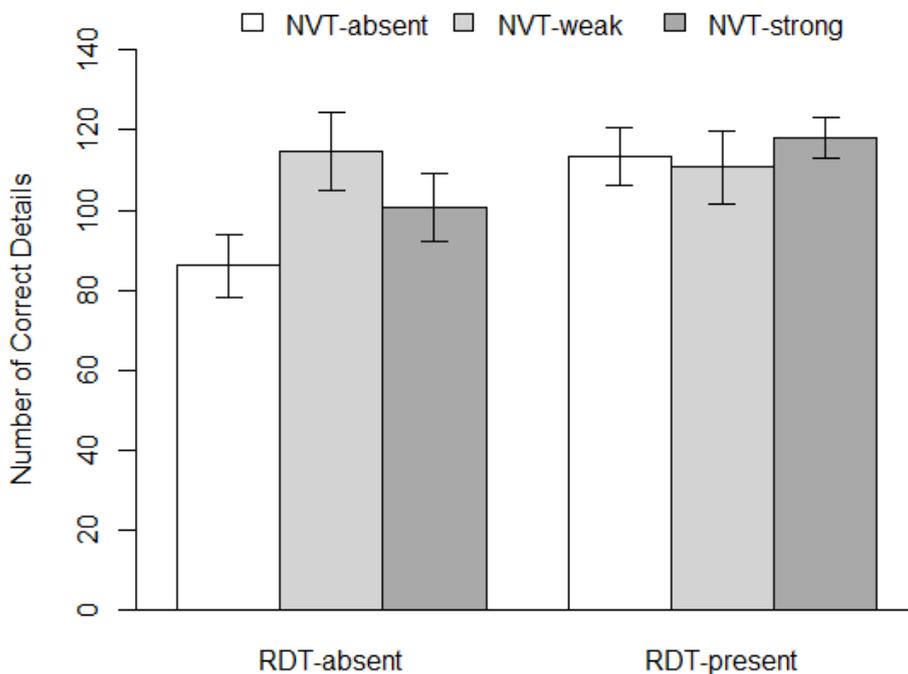


*Figure 4.2.* Mean correct recall (+SE) for each interview condition. RDT = report-detail instruction. NVT = naivety instruction.

There was no interaction observed between the report-detail and naivety instructions on correct recall, $F(2, 125) = 1.99$, $p = .14$, $\eta^2 = .03$, or recall errors, $F(2, 125) = 0.62$, $p = .54$,

$\eta^2 = .01$. However, a closer inspection of Figure 4.2 suggests there might be an interaction on correct recall[20]. This was explored with a simple effect analysis for the naivety instruction at each level of the report-detail factor (i.e., absent, present).
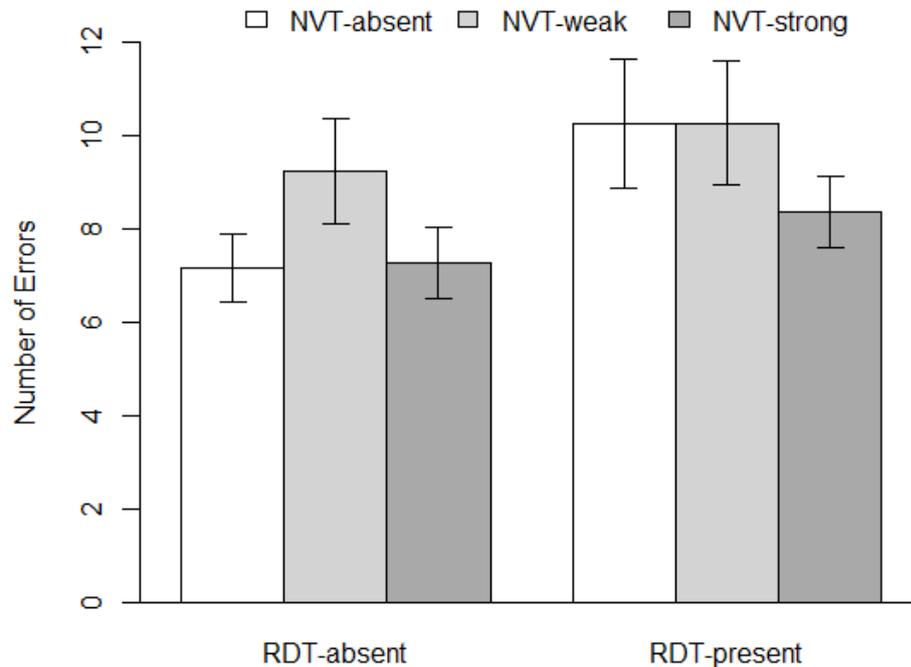


*Figure 4.3*. Mean errors (+SE) for each interview condition. RDT = report-detail instruction. NVT = naivety instruction.

Interestingly, for interviews excluding the report-detail instruction, a marginal but large sized difference in correct recall, $t(41) = -2.31$, $p = .03$, $CI_{99}[-62.35, 4.82]$, $d = 0.71$, suggests more correct details were produced in the interviews with the weak naivety statement ($M = 114.86$, $SE = 9.74$, $CI_{99}[88.47, 139.13]$), than the interviews without a naivety statement ($M = 86.09$, $SE = 7.82$, $CI_{99}[66.32, 105.19]$). However, there was no difference in correct recall for the remaining naivety comparisons: between the strong and weak naivety statements, $t(41) = 1.10$, $p = .28$, $CI_{99}[-20.58, 48.66]$, $d = 0.33$; and between the strong and absent naivety statements, $t(42) = -1.28$, $p = .21$, $CI_{99}[-45.68, 16.23]$, $d = 0.39$.

---

[20] It appears that an interaction exists on errors, however, there is wider variance in the data (see Figure 4.3) and all pairwise comparisons in a simple effect analysis (for the naivety instruction at each level of the report-detail factor), found no significant results.

In contrast, for interviews including the report-detail instruction, the simple effect analysis found no difference in correct recall between the three naivety conditions: strong and weak naivety statements, $t(42) = -0.71$, $p = .48$, $CI_{99}[-35.64, 20.73]$, $d = 0.22$; strong and absent naivety statements, $t(42) = -0.55$, $p = .58$, $CI_{99}[-28.64, 18.91]$, $d = 0.17$; and weak and absent naivety statements, $t(42) = 0.22$, $p = .83$, $CI_{99}[-28.83, 34.02]$, $d = 0.07$.

The interaction results suggest: (a) the report-detail instruction moderated the naivety instruction on correct recall; and (b) only the weak naivety statement increased correct recall, when the report-detail instruction was absent from interviews. A comment on these findings will be made after the grainsize results.

### 4.2.3.1.2 Grainsize dimension of report informativeness

Surprisingly, the report-detail instruction did not influence grainsize precision, $F(1, 125) = 0.52$, $p = .47$, $\eta^2 < .01$ (see Figure 4.4). I am jumping ahead of the present study, but this result is surprising because the report-detail instruction produced finer-grained reports in Experiment 3 (see Chapter 5, Section 5.5.2.1.2, p. 130). Again, it is likely the present result reflects a lack of power with the study because the report-detail instruction had an observed power of .11.

However, as expected, the naivety instruction influenced grainsize precision, $F(2, 125) = 3.21$, $p = .04$, $\eta^2 = .05$ (see Figure 4.4). A Tuckey post hoc test revealed a difference in grainsize precision ($p = .04$, $d = .52$), suggesting that interviews with the weak naivety statement had a medium sized effect on producing finer-grained reports ($M = 0.55$, $SE = 0.02$, $CI_{95}[0.51, 0.59]$), than interviews without a naivety statement ($M = 0.48$, $SE = 0.02$, $CI_{95}[0.43, 0.52]$). There was no difference in grainsize precision between the remaining naivety comparisons: strong and weak naivety statements ($p = .70$, $d = .18$), and strong and absent naivety statements ($p = .21$, $d = .36$). The results suggest that the main effect observed on

grainsize precision is dominated by the weak naivety statement. This was examined further with a simple effects analysis, described next.
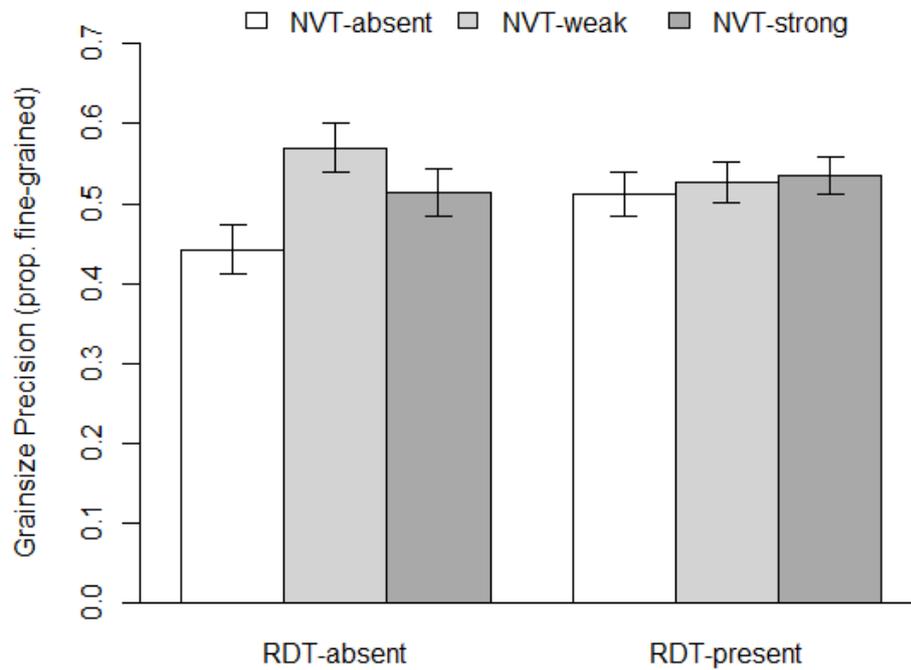


*Figure 4.4.* Mean grainsize precision (+SE) for each interview condition. RDT = report-detail instruction. NVT = naivety instruction.

Although, there was no interaction observed between the report-detail and naivety instructions on grainsize precision, $F(2, 125) = 1.99$, $p = .14$, $\eta^2 = .03$, a closer inspection of Figure 4.4 suggests an interaction. This was explored with a simple effect analysis for the naivety instruction at each level of the report-detail factor (i.e., absent, present).

Interestingly, for interviews excluding the report-detail instruction, a large sized difference in grainsize precision, $t(41) = -2.92$, $p = .01$, $CI_{99}[-0.24, -0.01]$, $d = .89$, suggested finer-grained reports were produced in the interviews with the weak naivety statement ($M = 0.57$, $SE = 0.03$, $CI_{99}[0.49, 0.65]$) compared with interviews devoid of a naivety statement ($M = 0.44$, $SE = 0.03$, $CI_{99}[0.36, 0.52]$). However, there was no difference in grainsize precision between the remaining naivety comparisons: strong and weak naivety statements, $t(41) =$

1.30., $p = .20$, $CI_{99}[-0.06, 1.71]$, $d = .40$; and strong and absent naivety statements, $t(42) = -1.65$, $p = .11$, $CI_{99}[-0.19, 0.05]$, $d = .50$.

In contrast, for interviews including the report-detail instruction, the simple effect analysis found no difference in grainsize precision between the three naivety conditions: strong and weak naivety statements, $t(42) = -0.26$, $p = .79$, $CI_{99}[-0.10, 0.09]$, $d = .08$; strong and absent naivety statements, $t(42) = -0.65$, $p = .52$, $CI_{99}[-0.12, 0.07]$, $d = .20$; and weak and absent naivety statements, $t(42) = -0.38$, $p = .71$, $CI_{99}[-0.12, 0.09]$, $d = .11$.

The interaction results suggest: (a) that the report-detail instruction moderated the naivety instruction; and (b) only the weak naivety statement produced finer-grained reports, when the report-detail instruction was absent from interviews. These findings replicate those established for the interaction on correct recall. It is significant finding that the report-detail instruction moderated the naivety instruction because the weak naivety statement used in the present study, replicates the naivety statement that was used in Experiment 1. In other words, the present results indicate that the reason the naivety instruction did not influence film reports in Experiment 1, is because the report-detail instruction moderated the statement of naivety. It is unclear why the strong naivety statement did not similarly influence grainsize precision (or correct recall) like the weak naivety statement did (perhaps due to insufficient power in the study because the naivety instruction had an observed power on grainsize precision of .60). However, the observation has implications for the mechanism that may lead a witness to reduce their report threshold and this will be addressed in the Discussion (see Section 4.3.1, p. 97). In summary, the results indicate that the report-detail instruction moderated the naivety instruction. Further, the naivety instruction only influenced reported information (correct recall and grainsize precision) when the report-detail instruction was absent from interviews.

***4.2.3.2  How the report-detail and naivety instructions impacted report accuracy***

These analyses focused on finding a trade-off in report accuracy that is expected to occur with greater report informativeness, on either the quantity dimension (i.e., Koriat & Goldsmith, 1994) or grainsize dimension (i.e., Yaniv & Foster, 1995). Curiously, as Figure 4.5 shows, neither the report-detail instruction, $F(1, 125) = 0.02$, $p = .90$, $\eta^2 < .01$, nor the naivety instruction, $F(2, 125) = 1.57$, $p = .21$, $\eta^2 = .02$, influenced recall accuracy, and there was no interaction, $F(2, 125) = 0.38$, $p = .68$, $\eta^2 = .01$.
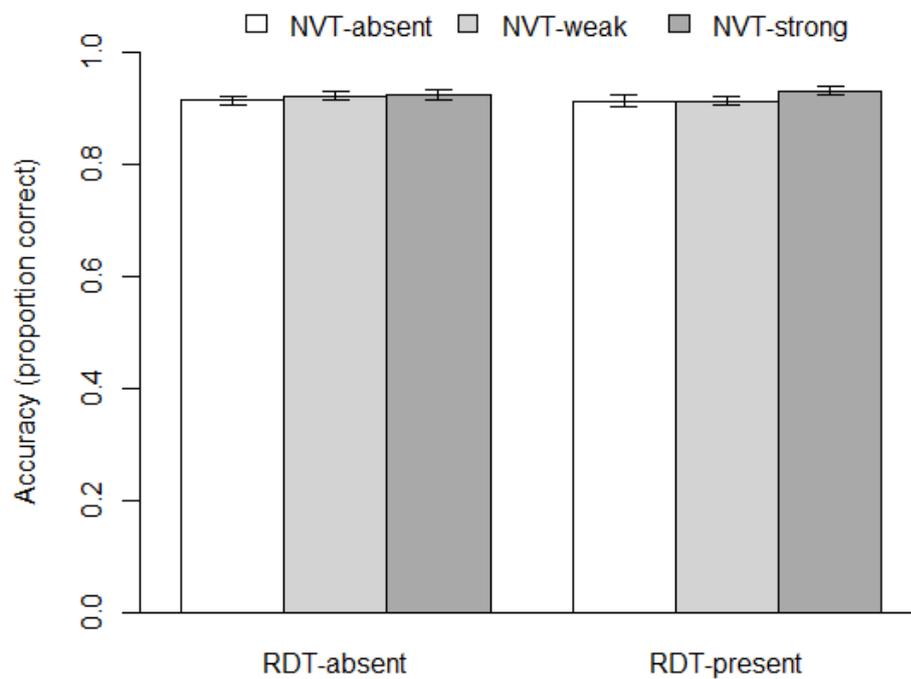


*Figure 4.5.* Mean recall accuracy (+SE) for each interview condition. RDT = report-detail instruction. NVT = naivety instruction.

Notably, the ceiling effect that can be seen in Figure 4.5, is similar to that found in Experiment 1 (see Section 3.2.2.2, p. 58). This effect is interesting because a series of open questions was used in the present study to avoid a ceiling effect on recall accuracy (see Section 4.1.3, p. 79). However, when data is collapsed across interview condition, the open questions reduced recall accuracy, $t(195.10) = 4.54$, $p < .01$, CI$_{95}$[0.01, 0.03], $d = .63$, suggesting narrative reports were less accurate in the present study ($N = 131$, $M = .92$, $SE < 0.01$, CI$_{95}$[0.91, 0.93]) than they were in Experiment 1 (i.e., film stimulus, $N = 81$, $M = .94$,

$SE < 0.01$, CI$_{95}$[0.94, 0.95]). This is examined further in Experiment 3 (see Chapter 5), where an experimental design was used to systematically explore the impact of open questions on narrative performance.

In summary, the report-detail and naivety instructions influenced report informativeness but not accuracy. These findings are significant because they suggest an informativeness-accuracy trade-off might not occur in narrative reports. The broader theoretical and practical implications of this finding will be addressed in the General Discussion (see Chapter 6).

### 4.3    Discussion

Instructions that challenge communication norms are important if they produce more information, as this will help the interviewer achieve their goal to obtain informative testimony. The naivety instruction produced more informative reports, replicating how it impacted narrative performance with the picture stimulus in Experiment 1 (see Section 3.2.3.2, p. 61). This is a significant outcome because it suggests that the efficacy of the naivety instruction is reliable across different stimuli (i.e., film and picture). Importantly, these findings indicate that a statement of naivety will lead a witness to violate communication norms, and produce a more informative report (as proposed by Milne, 2004). Moreover, the interactions found on report informativeness (correct recall and grainsize precision), suggest that the report-detail instruction will also lead a witness to violate communication norms. Theoretically, it appears that both instructions (naivety and report-detail) impact the report threshold, the mechanism assumed to produce greater report informativeness (i.e., Goldsmith et al., 2002; Koriat & Goldsmith, 1996). However, it also appears that the typical informativeness-accuracy trade-off, exemplified in the context of closed questioning (e.g., Koriat & Goldsmith, 1994; Weber & Brewer, 2008; Yaniv & Foster, 1995), does not occur with narrative recall. This will be addressed in the General Discussion

(see Chapter 6). The remainder of the present discussion will examine the mechanisms that may lead a witness to reduce their report threshold. The implications for narrative informativeness and the type of information reported, will be discussed.

### 4.3.1 Producing an informative report with the naivety instruction

Detailed descriptions are typically avoided in everyday conversation (Grice, 1975) because people may withhold information that they expect another person should know (e.g., Schuurmans & Vandierendonck, 1985). The naivety instruction produced more informative reports, suggesting that the communication norms that guide everyday conversation (i.e., Grice, 1975), are also applicable to the investigative interview setting. Moreover, the present study found that the naivety statement encouraged participants to believe the interviewer's stated naivety. Importantly, their belief was equally strong for the different naivety statements used by the interviewer (i.e., weak and strong). The association between belief and report informativeness, indicates that the naivety statement elicits the information that would otherwise be withheld by a witness (i.e., the information they expect the interviewer should know). This supports the idea that belief in the naivety statement is the mechanism that leads a witness to reduce their report threshold.

However, reports were most informative when the interviewer made a weak naivety statement. This finding might simply reflect pre-existing differences in the weak-naivety and strong-naivety groups (e.g., that might arise from random clustering of individual differences traits; Emmett et al., 2003). Alternatively, the results possibly indicate that whilst belief in the naivety statement might be a necessary condition to produce informative reports, it might not always be a sufficient mechanism to do so. That is, it might be critical *how* the interviewer claims naivety. In the present study, when the interviewer stated his naivety weakly, he simply explained that he had not seen the film himself. It is assumed this conveyed the message "I have no knowledge of the crime", and the belief ratings support this. In contrast,

when the interviewer stated his naivety strongly, he explained (a) he was one of several interviewers, (b) the interview was his first for the day, and (c) he understood the participant had watched one of several different films. Perhaps, this not only conveyed the message "I have no knowledge of the crime" (and the belief ratings support this) but also "I won't be able to verify your account". Although highly speculative, perhaps the participants withheld information that they thought the interviewer could not verify. An area for future research is to examine what forms of the naivety statement best encourage a witness to produce a more informative report.

### 4.3.2 Producing an informative report with the report-detail instruction

The report-detail instruction explicitly tells the witness how to produce informative testimony (i.e., describe everything in as much detail as possible). The report-detail instruction produced more informative reports, suggesting that this explicit demand for detail will encourage a witness to reduce their report threshold. In other words, witnesses may be more willing to report additional information (i.e., Bekerian & Dennett, 1993) because the interviewer tells them to do so. Moreover, as evidenced by participants' belief ratings (i.e., collapsed across naivety, $U = 2027.50$, $n_1 = 65$, $n_2 = 66$, $z = -0.55$, $p = .59$, $CI_{95}[-1.00, 1.00]$, $r = -.05$), belief in the interviewer's naivety is an unnecessary condition to produce informative testimony. It appears that the mechanism that leads a witness to reduce their report threshold, is different to the naivety instruction. The implication this has for the type of information produced in testimony, is explored next.

### 4.3.3 The centrality of information and communication norms

It is often assumed that the communication norm applicable to producing an informative report, is governed by Grice's (1975) Maxim of Quantity (e.g., Ackerman & Goldsmith, 2008; Evans & Fisher, 2011; Milne, 2004; Yaniv & Foster, 1997). As highlighted in Chapter 2 (see Section 2.3.2, p. 32), this maxim guides the speaker to provide sufficient

information to make their point to their listener. In this section, I explore how the communication norm governed by the Maxim of Relation (Grice, 1975), might also be relevant to producing an informative report. The Maxim of Relation guides the speaker to focus on what is relevant to the topic of conversation. Specifically, this maxim might have implications for the centrality of information reported, and elucidate how the naivety and report-detail instructions produce informative testimony.

For a person to understand the basic story of an event, central information is considered more relevant than peripheral information (Heuer & Reisberg, 1990). Central information is also recalled more frequently (e.g., Migueles & García-Bajos, 1999; Shapiro, Blackford, & Chen, 2005; Wong & Read, 2011) and more accurately (Luna & Migueles, 2009; Paz-Alonso, Goodman, & Ibabe, 2013), although memory for peripheral details may be improved under some circumstances (e.g., when correct post-event information is provided to witnesses; Sutherland & Hayne, 2001). The implication that these findings have for witness narrative reports will depend on the criteria used to define centrality (i.e., Paz-Alonso et al., 2013). This is an important point because what the interviewer and witness consider central and peripheral information might be very different. For example, in describing what happened, the witness might focus on the basic story of the event and report central information. However, information peripheral to their basic story might not be peripheral in investigative value (e.g., Geiselman & Callot, 1990). Moreover, peripheral information can enhance witness credibility (Bell & Loftus, 1989). Therefore, the centrality of information might have important implications for narrative recall, and how the naivety and report-detail instructions produce witness reports.

It is speculated that the naivety instruction might produce finer-grained central information because central information is recalled more accurately than peripheral information (Paz-Alonso et al., 2013). Consistent with this, the naivety instruction produced

finer-grained reports that contained more correct details but not errors (see Section 3.2.3.2, p. 61; and, Section 4.2.3.1, p. 89). Moreover, the naivety instruction did not influence the number of grainsize chunks recalled. This suggests that additional statements of information, perhaps peripheral information, was not reported. Turning to the report-detail instruction, it is speculated that it might produce more central and peripheral information because errors may be committed with peripheral details (Luna & Migueles, 2009). Consistent with this, the report-detail instruction produced a greater number of grainsize chunks (i.e., more statements of information), that contained more correct details and errors.

To shed light on these speculations, supplementary analyses were conducted to examine the influence that the naivety and report-detail instructions had on report centrality (i.e., the centrality of information). Briefly, report centrality was assessed according to whether information was central or peripheral to investigative purposes (Appendix H details the methodology used and the inferential analyses conducted). Interestingly, the results suggest that the naivety instruction produced finer-grained central information, and the report-detail instruction produced more central and peripheral information. The findings are consistent with the idea that (a) the naivety statement leads a witness to produce finer-grained central information, and (b) the report-detail instruction leads a witness to produce a greater quantity of central and peripheral information. Importantly, these findings indicate that the mechanism, that leads a witness to reduce their report threshold, depends on the communication norm that is impacted by instructional demand. In other words, if an instruction (e.g. report-detail) encourages a witness to violate the communication norm guided by the Maxim of Quantity, they might produce more informative central and peripheral testimony. In contrast, if an instruction (e.g., naivety) encourages violation of the communication norm guided by the Maxim of Relation, the witness might produce more informative central testimony.

An area for future research is to replicate these findings because they have important implications for how an interview is conducted. For example, if the naivety statement reliably produces finer-grained, accurate central information (i.e., Paz-Alonso et al., 2013), the interviewer might focus on this instructional demand for time critical situations.

### 4.4   Summary of Key Findings

- Instructional demands made with the naivety and report-detail instructions produced more informative reports without compromising accuracy.

- The report-detail and naivety instructions produced more informative reports in both quantity and grainsize dimensions of report informativeness, however, the report-detail instruction moderated the impact of the naivety instruction.

- Narrative performance suggests that both instructions encouraged participants to violate communication norms and relax their report threshold. However, the mechanism that led participants to produce more informative reports, was different. Belief in the interviewer's naivety was a necessary condition for the efficacy of the naivety instruction but not the report-detail instruction. Moreover, the centrality of information defined how the naivety and report-detail instructions produced informative reports.

**CHAPTER 5**

**Experiment 3**

This chapter presents the last empirical study exploring how different instructional demands influence narrative recall, and what insights into metacognition this provides. The study applied Ackerman and Goldsmith's (2008) dual-criterion model of metacognitive control to naturalistic witness reports, and tested the theoretical relevance of the dual-criterion model to narrative recall. The metacognition model that I developed for the investigative interview setting (see Figure 1.1, p. 12), provided the overarching context of the study because Ackerman and Goldsmith's dual-criterion model is grounded in Koriat and Goldsmith's (1996) metacognitive framework.

The dual-criterion model (Ackerman & Goldsmith, 2008) is theoretically important because it expands our understanding of reporting behaviour twofold. First, the dual-criterion model demonstrates how a person's report threshold is influenced by two criteria: a minimum-confidence criterion and a minimum-informativeness criterion. These criteria interact to manage uncertainty in recall and, consequently, the decision to report (Ackerman & Goldsmith, 2008). Second, the dual-criterion model demonstrates how a person's knowledge state will influence their decision to report. Specifically, a person's knowledge state will govern the criterion used for their report threshold. This is reviewed in more detail in Section 5.1.

Ackerman and Goldsmith (2008) developed their dual-criterion model in the context of closed questioning. I was motivated to apply their model in the context of open questioning, to elucidate how instructional demand influences witness knowledge state, and how knowledge state manifests in narrative recall performance. To achieve this, I first developed a dual-criterion model for the investigative interview setting, by applying

Ackerman and Goldsmith's model to (a) narrative recall, and (b) instructional demands in an interview. I focused on the instructional demands made with the report-detail instruction (to apply an informativeness incentive) and the do-not-guess instruction (to apply an accuracy incentive) because informativeness and accuracy incentives play a central role in Ackerman and Goldsmith's dual-criterion model.

Further, the present study explored how linguistic qualifiers spontaneously communicate uncertainty in narrative reports. It is important to understand how these qualifiers relate to memory performance because conventional measures of metacognition (e.g., calibration and discrimination indices) are not suited to narrative recall. To recap, conventional indices require confidence judgements about individual pieces of information (Schraw, 2009), yet interruptions (to obtain confidence judgements) are at odds with best-practice interview techniques (i.e., Fisher & Geiselman, 1992). This is because interruptions interfere with memory retrieval and diminish report informativeness (Baddeley, 2001; Fisher & Geiselman, 1992; D. L. Nelson & Goodmon, 2003). However, if linguistic qualifiers express uncertainty, then they might reflect witness knowledge state. Linguistic qualifiers are, therefore, considered an important complement to the dual-criterion model. Moreover, linguistic qualifiers have typically been ignored in the Cognitive Interview literature. The gap in understanding how and why witnesses use qualifiers, is a major contribution that this study makes. This chapter will begin with an overview of the literature relevant to the study, and how I applied Ackerman and Goldsmith's (2008) dual-criterion model to the investigative interview setting.

## 5.1   The Dual-Criterion Model and Witness Knowledge State

Ackerman and Goldsmith (2008) defined two knowledge states that may exist for a person when they respond to questioning. The researchers called these states satisficing knowledge and unsatisficing knowledge. Importantly, their dual-criterion model demonstrates

how the informativeness and confidence criteria interact to define these two knowledge states. Figure 5.1 illustrates the relationship between the informativeness and confidence criteria in each knowledge state, and is an adaptation of Ackerman and Goldsmith's schematic confidence-informativeness trade-off diagrams (see their Figure 1 on p. 1228). I have adapted their figure by simplifying their diagrams whilst preserving the key theoretical elements to their dual-criterion model. These key elements are:

- The acceptable ranges for each of the confidence and informativeness criteria (solid black fill), represent feasible criterion settings for the report threshold. With respect to the confidence criterion, any setting beyond the acceptable confidence range, would produce grossly inaccurate responses. Thus, there is a minimum-confidence criterion that sets the boundary of the acceptable confidence range. Similarly for the informativeness criterion, any setting beyond the acceptable informativeness range, would produce overly coarse-grained responses that are uninformative. Hence, there is a minimum-informativeness criterion that sets the boundary of the acceptable informativeness range.

- The confidence and informativeness criteria ranges are presented on opposing scales. This is to represent the competing goals for accurate and informative answers (i.e., Grice, 1975), and that these competing goals manifest in recall as a grainsize-accuracy trade-off (e.g., Goldsmith & Koriat, 1999).

- The report threshold (Prc) is governed by (a) the minimum-confidence criterion for satisficing knowledge, and (b) the minimum-informativeness criterion for unsatisficing knowledge.

When a person responds to questioning with satisficing knowledge, their recall is both accurate and informative because their confidence and informativeness criteria are both satisfied. This is why a satisficing knowledge state is represented in Figure 5.1 ('Satisficing Knowledge'), as an overlap in the acceptable ranges for the confidence and informativeness

criteria. With satisficing knowledge, a person uses their minimum-confidence criterion in the decision to report. Consequently, as Figure 5.1 shows, recall will be moderately accurate and moderately informative.

## Satisficing Knowledge

### Confidence    Informativeness

| V. low | | | V. high |

| Low | | | High |

| Mod | | | Mod — **Prc** |

| High | | | Low |

| V. high | | | V. low |

## Unsatisficing Knowledge

### Confidence    Informativeness

| V. low | | | Mod |

| Low | | | Low — **Prc** |

| Mod | | | V. low |

| High | | | |

| V. high | | | |
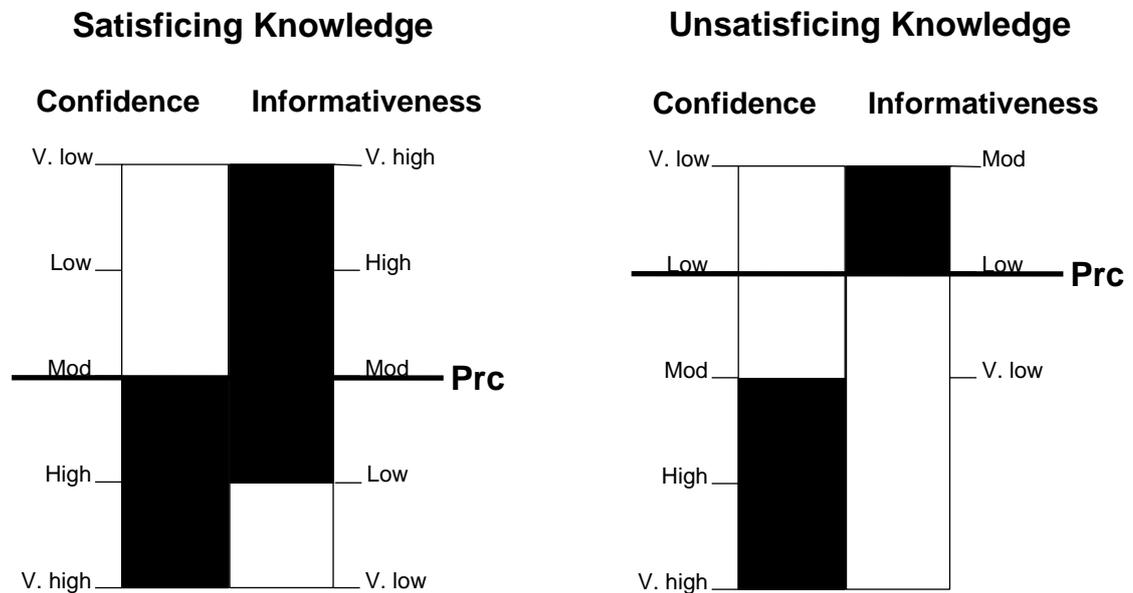
*Figure 5.1.* The dual-criterion model for satisficing and unsatisficing knowledge (adapted from Ackerman & Goldsmith, 2008). A satisficing knowledge state exists when there is overlap in the acceptable responses (solid black fill) that each criterion – confidence and informativeness – defines (left hand diagram). With satisficing knowledge, the minimum-confidence criterion is used for the report threshold so that information can be both accurate and informative. An unsatisficing knowledge state exists when there is a gap between the acceptable responses for each criterion (right hand diagram). With unsatisficing knowledge, the minimum-informativeness criterion is used for the report threshold so that information is somewhat informative (i.e., low informativeness) but it will be less accurate.

In contrast, when a person responds to questioning with unsatisficing knowledge, their recall is (somewhat) informative but less accurate because their confidence criterion is violated. This is why an unsatisficing state is represented in Figure 5.1 ('Unsatisficing Knowledge'), as a gap between the acceptable ranges for the confidence and informativeness criteria. With unsatisficing knowledge, a person uses their minimum-informativeness criterion in the decision to report and this is why, as Figure 5.1 shows, recall will be less accurate and less informative (i.e., coarser-grained). Notably, if a person were to use their minimum-

confidence criterion in their decision to report, recall would be grossly coarse-grained and uninformative (i.e., 'very low' informativeness in Figure 5.1). By using their minimum-informativeness criterion, this allows the person to report an answer of acceptable grainsize albeit coarser-grained than an answer produced with satisficing knowledge.

Ackerman and Goldsmith (2008) manipulated question difficulty to demonstrate the effect of knowledge state on recall. Easy questions were answered with satisficing knowledge and difficult questions were answered with unsatisficing knowledge. In the context of investigative interviewing, using best-practice techniques, it is relevant to consider how instructional demand influences knowledge state. This is conceptualised in Figure 5.2. I constructed the figure by applying Ackerman and Goldsmith's dual-criterion model to the investigative interview setting, and I considered the effect that three different instructions might have on witness knowledge state.

### 5.1.1 Proposed satisficing demand and satisficing witness knowledge

Ordinarily, when appropriate interview techniques are used (e.g., open questions and not interrupting the flow of information), a witness has full control over their narrative report (i.e., Fisher & Geiselman, 1992). In terms of the dual-criterion model, a witness who has full control over their narrative report is considered to have satisficing knowledge about the crime. In other words, if the witness' confidence and informativeness criteria are both satisfied, then their report should be both accurate and informative.

In the context of open questioning, one way the interviewer might produce a satisficing report is with a basic instruction that tells the witness to describe what happened (e.g. "Describe what you saw and heard"). This basic instruction is assumed to make a satisficing demand on the witness, and the witness uses satisficing knowledge to respond (see Figure 5.2 'Satisficing Demand'). Specifically, the witness is assumed to use their minimum-

confidence criterion in their decision to report and, as Figure 5.2 shows, their narrative report is assumed to be moderately accurate and moderately informative.

### 5.1.2    Proposed informativeness demand and unsatisficing witness knowledge

In contrast, the report-detail instruction – to describe everything in as much detail as possible – is assumed to make an informativeness demand on the witness. It is possible that the witness will respond with unsatisficing knowledge (see Figure 5.2 'Informativeness Demand'). This might occur because the witness, understanding the interviewer wants detailed descriptions, might set a finer-grained minimum-informativeness criterion for their report threshold, and violate their minimum-confidence criterion. If the witness uses unsatisficing knowledge then it is assumed that, as Figure 5.2 shows, their narrative report will be finer-grained but less accurate than a report produced with satisficing knowledge.

Notably, there is a fundamental difference between (a) Ackerman and Goldsmith's (2008) unsatisficing knowledge response to difficult questions, and (b) the proposed unsatisficing knowledge response to the report-detail instruction. This can be seen by comparing the informativeness criterion scales between Figure 5.1 and Figure 5.2. Critically, in Ackerman and Goldsmith's model, both satisficing and unsatisficing knowledge states use 'low' informativeness for the minimum-informativeness criterion (see Figure 5.1). Thereby, in demonstrating coarser-grained responses are produced with unsatisficing knowledge (in response to difficult questions), Ackerman and Goldsmith's informativeness scale is adjusted so that the minimum-confidence criterion may be violated (see Figure 5.1 'Unsatisficing knowledge'). In contrast, it is proposed that a witness produces finer-grained responses with unsatisficing knowledge (in response to the report-detail instruction) because they set a finer-grained minimum-informativeness criterion (e.g., 'high' informativeness). In this way, the minimum-confidence criterion may be violated without adjusting the informativeness scale (see Figure 5.2 'Informativeness Demand').
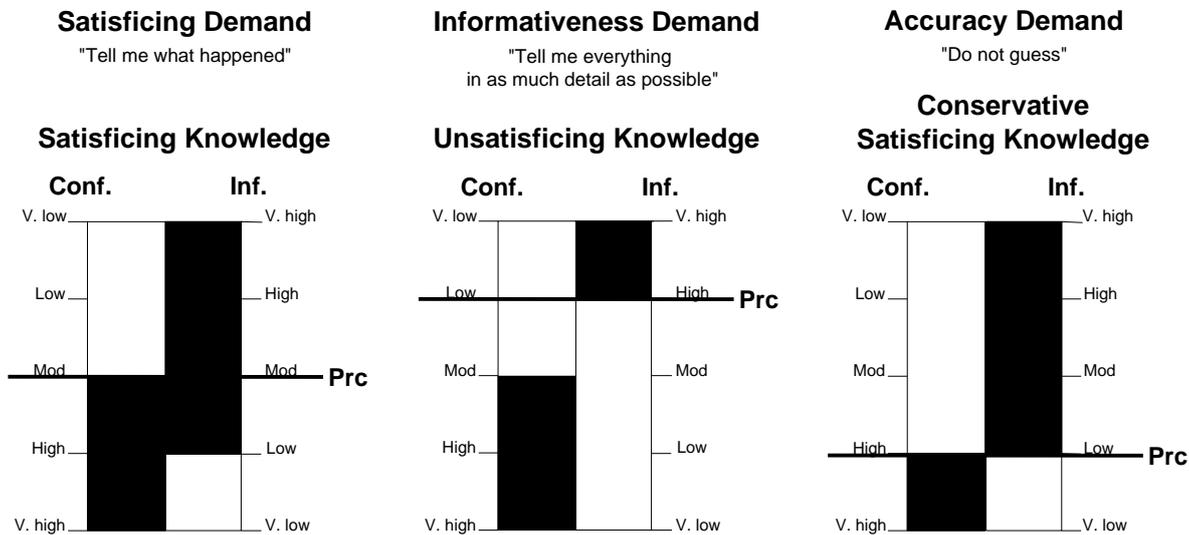
*Figure 5.2*. Ackerman and Goldsmith's (2008) dual-criterion model applied to narrative recall in the investigative interview setting. When the interviewer makes a satisficing demand (e.g. "Tell me what happened"), the witness responds with satisficing knowledge (left hand diagram) because their confidence (Conf.) and informativeness (Inf.) criteria are satisfied (i.e., the acceptable criterion ranges – solid black fill – overlap). With satisficing knowledge, the witness uses their minimum-confidence criterion for their report threshold (Prc). When an informativeness demand is made ("Tell me everything in as much detail as possible"), the witness responds with unsatisficing knowledge because a gap exists between their acceptable confidence and informativeness criteria ranges. This gap occurs because a finer-grained minimum-informativeness criterion is used for the report threshold but it violates the minimum-confidence criterion (centre diagram). When an accuracy demand is made ("Do not guess"), the witness responds with conservative satisficing knowledge. This occurs because a more stringent minimum-confidence criterion is used for the report threshold but it is limited by the minimum-informativeness criterion (right hand diagram).

### 5.1.3    Proposed accuracy demand and conservative satisficing witness knowledge

The do-not-guess instruction is assumed to make an accuracy demand on a witness, and the witness might use, for the purpose of definition, conservative satisficing knowledge to respond (see Figure 5.2 'Accuracy Demand'). It is assumed that the accuracy demand will influence the minimum-confidence criterion in the same way that an accuracy incentive does (i.e., penalty for errors; Goldsmith & Koriat, 2008; Koriat & Goldsmith, 1996). In other words, understanding that the interviewer wants an accurate report, the witness might use a more conservative minimum-confidence criterion for their report threshold (i.e., 'high' confidence in Figure 5.2). How conservative this is, might be restricted by the minimum-informativeness criterion (i.e., 'low' informativeness in Figure 5.2), otherwise the report will be overly coarse-grained and uninformative. If the witness uses conservative satisficing knowledge then it is assumed that, as Figure 5.2 shows, their narrative report will be coarser-grained and more accurate than a report produced with satisficing knowledge.

In summary, the application of Ackerman and Goldsmith's dual-criterion model to narrative recall in the investigative interview setting, predicts witness knowledge state when instructional demands are made. The proposed informativeness and accuracy performance outcomes might help elucidate why a witness decides to report or withhold information from narrative reports. Complementing this, is the question of how linguistic qualifiers communicate uncertainty in recall. It is important to understand how these qualifiers relate to memory performance because they might manifest knowledge state.

### 5.2    Communicating Uncertainty with Linguistic Qualifiers

Up until recently (i.e., Jack, Leov, & Zajac, 2014; Paulo, Albuquerque, & Bull, 2015), Cognitive Interview research had not examined how linguistic qualifiers communicate uncertainty in testimony (e.g., Aschermann et al., 1991; Dando et al., 2011; Davis et al., 2005; Fisher, Geiselman, Raymond, et al., 1987; Fisher et al., 1989; Jack et al., 2014; Mantwill,

Köhnken, & Aschermann, 1995; Mello & Fisher, 1996; Memon et al., 1994; Memon, Wark, Bull, et al., 1997). However, linguistic qualifiers might be an empirically insightful alternative to traditional indices of metacognition. As noted earlier, these indices require confidence judgements about individual pieces of information (Schraw, 2009) yet it is very poor practice to disrupt recall (e.g., Baddeley, 2001; Fisher & Geiselman, 1992; D. L. Nelson & Goodmon, 2003), to obtain these judgements. Consequently, linguistic qualifiers might produce a substitute measure of confidence, and provide an important link in applying metacognition theory to naturalistic witness behaviour.

Outside of the Cognitive Interview domain, it has been found that linguistic qualifiers express uncertainty (e.g., Budescu & Wallsten, 1995; Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986). Moreover, wide intra-individual variance exists in the specific qualifiers used (e.g., "possibly", "maybe", etc.), and the level of doubt that these communicate (Budescu & Wallsten, 1995). One reason this variance exists relates to interpersonal communication factors. For example, in an effort to avoid overusing any specific qualifier, there is a social norm that guides people to use a variety of synonyms to communicate doubt (Erev & Cohen, 1990). Nevertheless, as was found by Smith and Clark (1993), linguistic qualifiers can be broadly categorised as hedges (e.g., "I think", "I guess", "or something") or 'don't know' responses (e.g., "I'm not sure", "I don't know"). Importantly, both categories (hedges and 'don't know' responses) were found associated with lower cued-recall accuracy (i.e., in answers to general knowledge questions; Smith & Clark, 1993).

More recently, in the Cognitive Interview domain, a greater number of hedges were found to be associated with lower narrative accuracy (Jack et al., 2014) and, specifically, hedged details (e.g., "the jacket was black, I think") were found to be less accurate than un-hedged details (Paulo et al., 2015). This suggests a witness uses hedges to communicate doubt. However, research has not examined how witnesses use 'don't know' responses to

communicate doubt in narrative recall. Certainly, in response to closed questioning, 'don't know' responses are an important linguistic device to regulate accuracy (e.g., Evans & Fisher, 2011; Koriat & Goldsmith, 1994; Weber & Brewer, 2008)[21], and avoid overly coarse-grained uninformative responses (e.g., Ackerman & Goldsmith, 2008). Anecdotally, however, participants in Experiments 1 and 2 used 'don't know' responses in their narrative reports (e.g., "He was wearing jeans but I don't know").

The implication that the aforementioned findings have for narrative recall, is that <u>all</u> qualifying phrases (e.g., hedges and 'don't know' responses) are relevant to the possibility that linguistic qualifiers manifest witness knowledge state. The next section examines how linguistic qualifiers might be associated with narrative performance, in manifesting proposed knowledge state responses to different instructional demands.

### 5.3    Hypothesised Outcomes

The hypothesised outcomes are presented in diagrammatic form (Figure 5.3). The figure depicts the following key points:

- Ackerman and Goldsmith's (2008) dual-criterion model is applied to narrative recall in the investigative interview setting. Specifically, I considered the effect that instructional demands, made with three different instructions, might have on knowledge state. Of note, Ackerman and Goldsmith's (2008) informativeness criterion predicts grainsize precision because they used a closed questioning methodology that focused on response grainsize. In contrast, the informativeness criterion in Figure 5.3 predicts both the quantity of details reported and grainsize precision, because narrative informativeness comprises both dimensions.

- Knowledge state manifests in narrative performance and recall uncertainty.

---

[21] Further, 'don't know' was used more often to "yes"/"no" questions than closed questions because it was the only option to regulate accuracy (i.e., grainsize was also available for closed questions; Evans & Fisher, 2011)

- Narrative performance reflects the underlying assumptions of the dual-criterion model that (a) informativeness and accuracy are competing goals for memory reporting (i.e., Grice, 1975), and (b) these competing goals manifest quantity-accuracy (i.e., Koriat & Goldsmith, 1994) and grainsize-accuracy (i.e., Yaniv & Foster, 1995) trade-offs in recall. As Weber and Brewer state, "the quantity-accuracy trade-off in memory reporting is analogous to the accuracy-informativeness trade-off in grain size regulation" (p. 56; 2008). The inter-relationship between the three recall variables (quantity, grainsize and accuracy) is represented by the triangle in Figure 5.3.

- Recall uncertainty is reflected in linguistic qualifiers, and linguistic qualifiers include hedges and 'don't know' responses.

### 5.3.1   Satisficing knowledge: baseline narrative performance and recall uncertainty

Provided an interviewer demands testimony with a basic instruction to describe what happened (e.g., "Describe what you saw and heard"), the witness is assumed to have full control over narrative informativeness and accuracy (i.e., Fisher & Geiselman, 1992). It is proposed that the witness will respond with satisficing knowledge, and use their minimum-confidence criterion to produce a moderately informative and moderately accurate account (see Figure 5.3 'Satisficing Demand').

It is also proposed that the witness will communicate at least some information they are less certain is correct, and use qualifying phrases (e.g., hedges and/or 'don't know' responses) to communicate this doubt. The information that is qualified, is assumed to depend on the confidence held in the individual details reported. Moreover, details that might be qualified could be correct or erroneous, although it is assumed errors will be qualified more frequently (i.e., Paulo et al., 2015). Therefore, to assess recall uncertainty, I was interested in the number of qualifiers associated with errors, as a proportion of the total number of qualifiers associated with correct details and errors.

## Satisfising Demand  -  Satisficing Knowledge
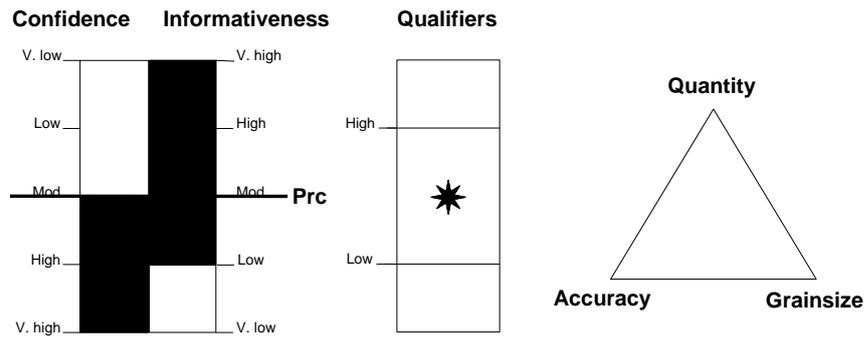


## Informativeness Demand  -  Unsatisficing Knowledge
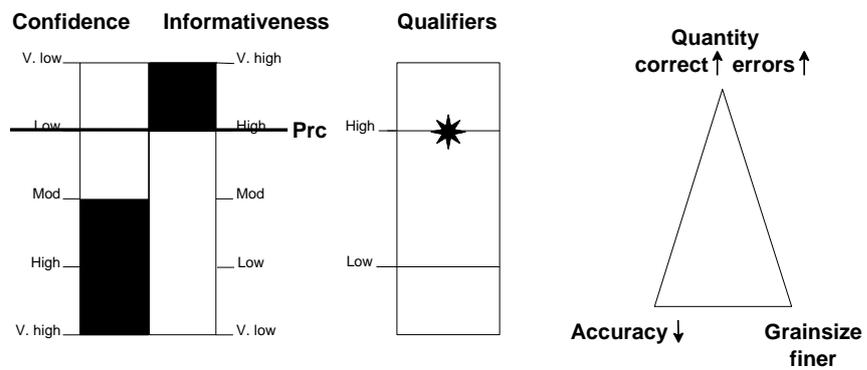


## Accuracy Demand  -  Conservative Satisficing Knowledge
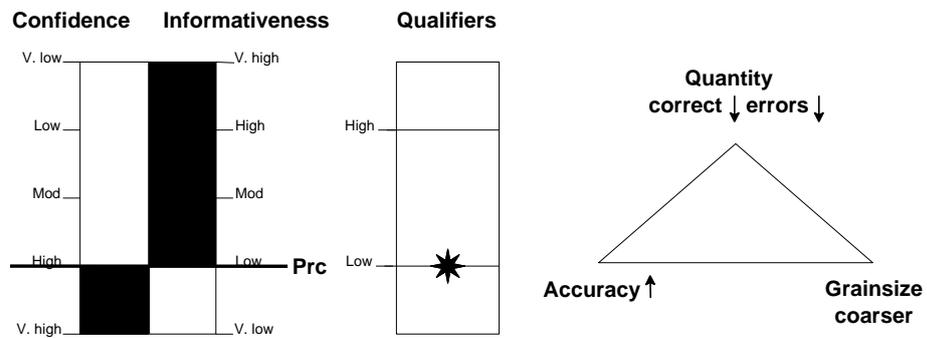


*Figure 5.3.* Ackerman and Goldsmith's (2008) dual-criterion model applied to narrative recall in the investigative interview setting, and hypothesised outcomes in recall performance and linguistic qualifiers. Compared with a satisficing knowledge response to a satisficing demand (top diagrams), unsatisficing knowledge – in response to an informativeness demand – will manifest a finer-grained, less accurate report that contains more correct details and errors qualified more frequently (middle diagrams). In response to an accuracy demand, conservative satisficing knowledge will manifest a coarser-grained, more accurate report that contains fewer correct details and errors qualified less frequently (bottom diagrams).

As Figure 5.3 (see 'Satisficing Demand') shows, narrative performance and recall uncertainty manifest satisficing knowledge with some arbitrary and unknown baseline values. These baseline vales were quantified by the experimental control condition, when the interviewer simply asked participants to describe what they saw and heard.

### 5.3.2 Unsatisficing knowledge: predicted narrative performance and recall uncertainty

When the interviewer demands informative testimony with the report-detail instruction, it is proposed that the witness will respond with unsatisficing knowledge. This occurs because the witness is assumed to violate their minimum-confidence criterion, to use a more informative minimum-informativeness criterion for their report threshold (see Figure 5.3 'Informativeness Demand').

As Figure 5.3 shows, unsatisficing knowledge is expected to produce a narrative account that is more informative (i.e., finer-grained and contains more details), and less accurate, than testimony produced with satisficing knowledge. Moreover, unsatisficing knowledge is expected to manifest greater recall uncertainty because the report contains more errors.

### 5.3.3 Conservative satisficing knowledge: predicted narrative performance and recall uncertainty

When the interviewer demands accurate testimony with the do-not-guess instruction, the witness is proposed to respond with conservative satisficing knowledge. This occurs because the witness is assumed to use a conservative minimum-confidence criterion – but restricted by their minimum-informativeness criterion – for their report threshold (see Figure 5.3 'Accuracy Demand').

As Figure 5.3 shows, conservative satisficing knowledge is expected to produce a narrative account that is less informative (i.e., coarser-grained and contains fewer details), and more accurate, than testimony produced with satisficing knowledge. Conservative satisficing knowledge is also expected to manifest lower recall uncertainty because the report contains fewer errors.

## 5.4   Method

### 5.4.1   Participants

The sample comprised 201 participants recruited from the general university community or from first year Psychology students. General university participants received AUD\$10 and first year Psychology students received course credit. Informed written consent, including permission to be video-recorded, was obtained from all participants. Nine participants were excluded from the study because the interviewer delivered instructions incorrectly (7 people), one participant replayed the film, and one participant did not follow instructions. The mean age of the remaining 192 participants was 21.69 years and ranged from 17 to 51 years. The sample comprised 63 males ($M = 21.52$ years, $SD = 5.80$) and 129 females ($M = 21.77$ years, $SD = 7.44$). All remaining participants had normal or corrected-to-normal vision and were fluent in English.

#### 5.4.1.1   Interviewers

Two interviewers were used and they were naive to the experimental aims of the research. Interviewer A (43-year-old male) conducted 108 interviews and he was the same person who interviewed the participants in Experiment 1. He volunteered his time for a small honorary payment per interview. Interviewer B (22-year-old female) conducted 84 interviews and she volunteered her time for free.

As with the first two studies, before the study commenced, the interviewers familiarised themselves with the interview instructions. These were printed on a separate card for each experimental condition. The interviewers each conducted several practice interviews over the course of a half-day, using each other and myself as mock witnesses. During these interviews, they practiced delivering instructions slowly and coherently, and restricted their use of non-verbal (i.e., head nods) and verbal (e.g., "uh huh", "yep") acknowledgements. This training helped to maintain experimental control.

### 5.4.2   Materials

All of the materials used were similar to the first two studies, including the film stimulus. However, the picture stimulus and post-interview questionnaire were omitted.

### 5.4.3   Procedure

The procedure outlined in Chapter 4 (see Section 4.1.3, p. 79) was followed except: there was no deception used, there was no restriction made on any interview condition (e.g., restricted to the morning), and a post-interview questionnaire was not administered. The deception, restriction and questionnaire used in Experiment 2, were all relevant to the naivety instruction, and therefore irrelevant to the present study.

Participants were interviewed with one of four interview conditions, randomly assigned to each participant. The report-detail (RDT) and do-not-guess (DNG) instructions were manipulated in a two-way factorial, between-subjects design to elicit a narrative report. The four interview conditions comprised: (a) RDT-absent, DNG-absent; (b) RDT-present, DNG-absent; (c) RDT-absent, DNG-present; and (d) RDT-present, DNG-present. These conditions are explained next.

Participants in the control condition (RDT-absent, DNG-absent) were asked to describe what they saw and heard on the film. When the report-detail instruction was present,

the interviewer stated: "Tell me everything that you saw and heard on the video and describe this to me in as much detail as you can". When the do-not-guess instruction was present, the interviewer stated: "If you cannot remember something do not guess".

After eliciting a response with one of the interview conditions, when the participant appeared to have finished volunteering information, the interviewer waited for a short period and then asked if there was anything else they could add. The participant's free-narrative response was concluded after additional information was volunteered or they declined to add anything else. The interviewer then asked a series of open questions. Participants were instructed to respond even if they had provided information in their free-narrative that was relevant to the questions. This encouraged participants not to omit any details. The open questions were used to elicit additional narrative responses about the film.

Although this thesis does not focus on question format – free narratives produce greater accuracy than cued responses (i.e., Evans & Fisher, 2011) – the purpose of the open questions was to avoid a ceiling effect on report accuracy, that was observed in the first two studies. Specifically, open questions were used to help observe a trade-off in accuracy when a more informative report is produced. This was an important observation to make, given that underlying assumptions of the dual-criterion model are: (a) informativeness and accuracy are competing goals for recall performance (i.e., Grice, 1975), and (b) these competing goals manifest as quantity-accuracy (i.e., Koriat & Goldsmith, 1994) and grainsize- accuracy (i.e., Yaniv & Foster, 1995) trade-offs in recall. Critically, however, open questions were used to invite narrative responses to ensure the present study addressed the overarching purpose of the thesis – to observe how instructional demands influence narrative recall.

The number of open questions was counterbalanced across interviews. This was done to follow-up on findings observed in Experiments 1 and 2, and explore the effect that question number (low, high) has on narrative performance. To recap, when data was collapsed across

interview condition, the series of open questions used in Experiment 2 produced less accurate reports ($N = 131$, $M = .92$, $SE < 0.01$, $CI_{95}[0.91, 0.93]$), than the reports produced without a series of open questions in Experiment 1 (i.e., film stimulus, $N = 81$, $M = .94$, $SE < 0.01$, $CI_{95}[0.94, 0.95]$). In the present study, three open questions (low-number condition) were used in half (i.e., 96) of the interviews, and nine open questions were used (high-number condition) in the remaining interviews.

The low-number open question condition comprised one question each from the three categories of information (Person, Surrounding, and Action). The order of these questions was pseudo-randomised and counterbalanced across the interview conditions. The high-number open question condition comprised three questions from each category of information. Questions were grouped by category. Question order within each category group, and category groups, were pseudo-randomised and counterbalanced across interviews.

The final narrative report comprised the free-narrative response and the additional narrative responses to the open questions. Participants were not interrupted during their reports and after interviewing was finished, each participant was debriefed.

The actual interview instructions and open questions that were used in the present study, can be found in Appendix B. These were read out verbatim to participants. Every interview was video-recorded for later transcription and scoring.

### 5.4.4 Dependent variables to assess report informativeness and accuracy

The same dependent variables outlined in Chapter 3 (see Section 3.1.4, p. 45) were used in the present study. To recap, report informativeness was assessed with quantity and grainsize variables. Quantity variables included the number of correct details and errors recalled. The grainsize variable was grainsize precision and it was calculated by dividing the number of fine-grained chunks by the total number of grainsize chunks recalled (fine-grained

+ coarse-grained). To recap, 'grainsize chunk' is a non-technical label applied to a statement of related information (e.g., a statement about the thief's upper-body clothing), and grainsize chunks could be either coarse-grained or fine-grained statements. Report accuracy was assessed as the proportion of details recalled that were correct (correct / correct + errors). How these variables were coded is described shortly.

### 5.4.5 Dependent variable used to assess recall uncertainty

Recall uncertainty was assessed with linguistic qualifiers that expressed doubt, and included hedges (e.g., "I think", "I guess") and 'don't know' responses (e.g., "I don't know", "I'm not sure"). Linguistic qualifiers that expressed certainty (e.g., "definitely", "I'm certain", etc.) were ignored.

For each participant, the number of qualifiers associated with correct details (qualified correct), and the number of qualifiers associated with errors (qualified errors), was calculated. The key variable of interest, recall uncertainty, was then calculated for each participant by dividing the number of qualifiers associated with errors, by the number of qualifiers associated with errors and correct details (qualified errors + qualified correct).

### 5.4.6 Scoring narrative recall and linguistic qualifiers

The approach taken to score narrative recall is outlined first in this section. Following this, the reader will be presented with the approach taken to score linguistic qualifiers.

### 5.4.6.1 *Scoring quantity, grainsize and accuracy*

The two dimensions of narrative informativeness, quantity and grainsize, were scored following the same procedure outlined in Chapter 3 for the free-narrative response (see Section 3.1.5, p. 46), and Chapter 4 for the narrative responses to the open questions (see Section 4.1.6, p. 82). Also following the procedure outlined in Section 4.1.6 (p. 82), the scores obtained from the free-narrative and open question responses were combined, to derive

narrative performance for each recall variable (i.e., correct, errors, accuracy, and grainsize precision). The reader is referred to Appendix C and E for further detail on the scoring rules, coding keys and examples.

### 5.4.6.2 Linguistic qualifiers coding key and scoring procedure

The coding key that I developed for scoring linguistic qualifiers, comprised two categories of qualifying phrases: hedges and 'don't know' responses. The full list of the qualifying phrases that were used to score the transcripts, can be found in Appendix I.

The hedges category included the qualifying phrases "I guess", "I think", and "or something". These phrases were cited by Smith and Clark (1993) as the most common hedges that their participants used to qualify answers to general knowledge questions. In the present study, the hedges category also included other similar qualifying phrases (e.g., "It might have been", "perhaps") and probability phrases (e.g., "maybe", "possibly").

The 'don't know' category included the qualifying phrases "I don't know", "I don't remember", and "I'm not sure". These phrases were cited by Smith and Clark (1993) as 'nonanswers' for the following reason. When their participants provided an answer (e.g., "Soccer" to the question "In which sport is the Stanley Cup awarded?") but followed their response with "I don't know" (i.e., "Soccer, I don't know"), the researchers ignored the answer regardless that it might have been correct or incorrect. I was interested in 'don't know' responses when they were associated with details that could be scored as correct or erroneous (e.g., "the jacket was black but I don't know") because, as qualifying phrases, 'don't know' responses communicate doubt. The 'don't know' category also included other similar qualifying phrases (e.g., "I forgot", "it escapes me", etc.).

It is possible that hedges and 'don't know' responses lie on a continuum of expressing recall uncertainty. For example, it might be the case that when a person qualifies a detail with

a 'don't know' response they might have greater doubt that the information is correct, than if they qualified the same detail with a hedge. I was motivated to explore this possibility, however, the number of 'don't know' responses was very low ($N = 192$, $M = 3.05$, $SE = 0.18$, $CI_{95}[2.69, 3.42]$) and so I did not pursue this line of thinking. As an aside, finding that 'don't know' responses were rarely used to qualify a piece of information, concurs with the finding by Evans and Fisher (2011) that their participants did not use "don't know" during narrative recall.

Free-narratives were scored for the number of linguistic qualifiers (hedges and 'don't know' responses) that were associated with details that had been scored with the quantity coding key. A linguistic qualifier was given a score of '1' and when a detail was qualified with more than one phrase, then each unique phrase was scored (e.g., "I think it was black or something"). Qualifiers associated with finer-grained details were scored in preference to qualifiers associated with coarser-grained details. For example, if a participant stated "I think it was a 4WD" and also stated "It was a Landcruiser or something", then the qualifier associated with '4WD' was ignored. This grainsize rule was used to match the grainsize rule used for quantity scoring (see Section 3.1.5.1, p. 47). Moreover, linguistic qualifiers were not scored if they were associated with suppositions that related to feeling (e.g., "He looked upset, I think") or thinking state (e.g., "I think he planned to steal the card"). Qualifiers were also not scored if they were associated with information that could not be verified from the film (e.g., "I think there was someone in the 4WD" but this could not be verified).

For narrative responses to the open questions, linguistic qualifiers were only scored when qualifying phrases were associated with new information that had not been stated in the free-narrative. For example, if "it was possibly a jacket" had been stated in the free-narrative, and "it was possibly a jacket, it was black I think" was stated in response to the question about the thief's appearance, only the hedge associated with the colour of the jacket was

scored. This is because the hedge associated with 'jacket' would have already been scored in the free-narrative. This scoring rule was used to match the rule for scoring new information in open question responses.

A separate tally was made for the number of linguistic qualifiers associated with correct details (i.e., qualified correct) and errors (i.e., qualified errors), reported across the free-narrative and open question responses. Recall uncertainty was calculated by dividing the qualified errors variable, by the sum of the qualified correct and qualified errors variables (qualified errors / qualified correct + qualified errors).

### 5.4.6.3 *Inter-rater reliability*

All interview transcripts were scored by myself, and thirty-nine transcripts (20%), randomly selected, were scored independently by a second coder. As outlined in Chapter 3 (Section 3.1.5, p. 46), the interview instructions were omitted from the transcripts during transcription, to ensure scoring was done blind to the interview condition. The inter-rater reliability was assessed by percentage agreement (see Section 3.1.5.3, p. 49, for why the kappa statistic was not used). My scores, except when they were changed during discussion with the second coder, were retained for the main analysis.

For the total quantity of details recalled (correct + errors), an inter-rater reliability of 80.4% was established for the free-narrative response (disagreements were not discussed) and 97.4% was established for the open question responses (disagreements were discussed). This derived an average inter-rater reliability of 84.4% for the total quantity of details recalled. Prior to discussing disagreements, the percentage agreement was approximately 75% for the open question responses. The majority of disagreements were related to details already provided in the free-narrative response and that should not have been scored in the open question responses.

For the total number of grainsize chunks recalled (coarse-grained + fine-grained), after disagreements were reviewed, an inter-rater reliability of 96.0% was established. The procedure outlined in Section 3.1.5.3 (see p. 49) was followed with grainsize disagreements.

For linguistic qualifiers, an inter-rater reliability of 92.1% was established after disagreements were reviewed. The majority of disagreements related to qualifiers simply being missed or miscategorised.

## 5.5    Results

Descriptive statistics for the different interview conditions are presented in Table 5.1. The table shows that most variables were influenced by interview condition, except accuracy.

### 5.5.1    Data screening

Parametric assumptions for each interview condition were checked using the Shapiro-Wilk normality test and the Levene's test for homogeneity of variance (using the median). Grainsize precision met parametric assumptions. Correct recall and accuracy were skewed (in some conditions) with equal variance. Errors and recall uncertainty were skewed in all conditions with equal variance.

Outliers were examined with standardized data for each interview condition and identified when $z$-scores were less than -2.50 or greater than 2.50. Outliers were found in all variables: correct recall (3 outliers), errors (6), accuracy (2), grainsize precision (2), and recall uncertainty (3).

#### 5.5.1.1    *Permutation ANOVA*

Permutation ANOVA was used to check parametric violations and outliers, when present, did not bias the statistical inference that was determined from parametric ANOVA. The same approach was taken as outlined in Chapter 3 (see Section 3.2.1.3, p. 52). However,

due to the larger sample size, 5,000 permutations were calculated instead of 2,000. For each dependent variable that showed a significant parametric ANOVA result, the permutation *p*-value that was obtained with permutation ANOVA, was less than .05. This suggested that the parametric ANOVA result was reliable and unlikely to be a Type I error.

Table 5.1

*Mean Recall [and 95% confidence intervals] by Interview Condition*

| Variable | Interview condition | | | |
|---|---|---|---|---|
| | RDT-a DNG-a (*n* = 48) | RDT-a DNG-p (*n* = 48) | RDT-p DNG-a (*n* = 48) | RDT-p DNG-p (*n* = 48) |
| | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* |
| Correct | 70.27 (34.30) [60.31, 80.23] | 88.83 (34.43) [78.84, 98.83] | 105.94 (35.42) [95.65, 116.22] | 105.23 (33.78) [95.42, 115.04] |
| Errors | 6.56 (4.22) [5.34, 7.79] | 7.48 (3.89) [6.35, 8.61] | 9.48 (4.92) [8.05, 10.91] | 8.90 (5.17) [7.39, 10.40] |
| Accuracy | .91 (.04) [.90, .93] | .92 (.04) [.91, .93] | .92 (.04) [.90, .93] | .92 (.04) [.91, .93] |
| GS Precision | .43 (.13) [.39, .47] | .46 (.11) [.43, .49] | .52 (.14) [.48, .56] | .49 (.13) [.46, .53] |
| Recall Uncert | .33 (.25) [.26, .40] | .37 (.24) [.30, .44] | .34 (.26) [.26, .41] | .30 (.24) [.23, .37] |

*Note.* Correct = Number of correct details; Errors = number of errors; Accuracy = proportion of correct details; GS Precision = proportion of fine-grained recall; Recall Uncert = proportion of qualifiers associated with errors. RDT-a = report-detail instruction absent; RDT-p = report-detail instruction present; DNG-a = do-not-guess instruction absent; DNG-p = do-not-guess instruction present.

### 5.5.1.2 *Influence of interviewer*

Group differences, collapsed across interview condition, were examined to assess if the person interviewing (interviewer A or B) influenced narrative performance and recall uncertainty. Student's *t*-tests are reported for accuracy and grainsize precision. Wilcoxon rank sum tests are reported (with $n_{interviewer\,A} = 108$ and $n_{interviewer\,B} = 84$) for correct recall, errors and recall uncertainty.

The person interviewing did not influence narrative performance: correct recall, $U = 4303.00$, $z = -0.61$, $p = .54$, CI$_{95}$[-14.00, 7.00], $r = -.04$; errors, $U = 4429.5$, $z = -0.28$, $p = .78$, CI$_{95}$[-1.00, 1.00], $r = -.02$; accuracy, $t(190) = -0.20$, $p = .84$, CI$_{95}$[-0.01, 0.01], $d = 0.03$; and grainsize precision, $t(190) = 0.68$, $p = .50$, CI$_{95}$[-0.02, 0.05], $d = 0.10$.

However, the person interviewing marginally influenced recall uncertainty, $U = 3799.50$, $z = -1.94$, $p = .05$, CI$_{95}$[-0.14, 0.00], $r = -.14$, suggesting participants' recall was more uncertain with interviewer A (*Mdn* = .38, *SE* = .02, CI$_{95}$[.36, .40]) than with interviewer B (*Mdn* = .32, *SE* = .04, CI$_{95}$[.23, .36]).

Closer scrutiny of the interview conditions, found that 57% of the interviews conducted by interviewer A (62 out of 108) used the high-number open question condition (i.e., nine open questions were asked). This contrasts with interviewer B, who conducted 40% of her interviews (34 out of 84) with the high-number open question condition. This disparity between the interviewers, is considered the most likely explanation for why interviewer A elicited greater recall uncertainty. This conclusion is supported by the observation that recall uncertainty was influenced by the number of open questions asked in the interviews (see Section 5.5.3, p. 134).

### 5.5.2 Narrative performance and recall uncertainty

In the following analyses, the focus was on testing the application of Ackerman and Goldsmith's (2008) dual-criterion model to narrative recall in the investigative interview setting. Specifically, analyses examined if different instructions impact witness knowledge, and produce reports that manifest knowledge state. The underlying assumption for these analyses is that the basic instruction (to simply describe what was seen and heard), used in the experimental control condition, produced baseline results that are consistent with satisficing knowledge. In detail, the anticipated main effects were:

- If the report-detail instruction induces unsatisficing knowledge, then reports were expected to be more informative, less accurate, and communicated with greater uncertainty.

- If the do-not-guess instruction induces conservative satisficing knowledge, then reports were expected to be less informative, more accurate, and communicated with lesser uncertainty.

Moreover, the underlying assumptions of the dual-criterion model are (a) informativeness and accuracy are competing performance goals (i.e., Grice, 1975), and (b) these competing goals manifest an informativeness-accuracy trade-off (i.e., Koriat & Goldsmith, 1994; Yaniv & Foster, 1995). Therefore, I was also interested in any significant interactions. In the presence of a significant interaction, I was specifically interested in the simple main effects of: (a) the report-detail instruction, at each level of the do-not-guess factor (i.e., absent, present); and (b) the do-not-guess instruction, at each level of the report-detail factor (i.e., absent, present).

Inferential analyses were calculated using univariate factorial ANOVA (alpha level set at .05) in IBM SPSS Statistics 21. Type III sum of squares was used due to the balanced design. Effect size, eta squared ($\eta^2$), was calculated using the *lsr* package (version 0.5;

Navarro, 2015) in R (version 3.2.2; R Core Team, 2014). For significant effects, point estimates of the mean are given with 95% confidence intervals. These were calculated with the bias-corrected and accelerated bootstrap method in SPSS, using 2000 permutations. Verbal description for eta squared adopted Cohen's (1988) recommendations (i.e., .01 = 'small', .06 = 'medium', .14 = 'large').

When interactions were present, a Bonferroni adjustment (alpha = 0.013) was applied to control for familywise error across the four pairwise comparisons used to explore simple main effects. In all group comparisons, $n_1 = n_2 = 48$. Simple main effects were examined with the Student's $t$ test when data met parametric assumptions (effect size was calculated with Cohen's $d$), and the Wilcoxon rank sum test when data was skewed with equal variance (effect size was calculated with $r$). Both tests were calculated using the *psych* package (version 1.4.8; Revelle, 2014) in R (version 3.2.2; R Core Team, 2014). For significant simple main effects, point estimates of the mean (or median) are given with 98.7% confidence intervals. These were calculated using the bias-corrected and accelerated bootstrap method in IBM SPSS Statistics 21, using 2000 permutations. Verbal description for effect size adopted Cohen's (1992) recommendations for $d$ (i.e., 0.20 = 'small', 0.50 = 'medium', 0.80 = 'large') and $r$ (i.e., .10 = 'small', .30 = 'medium', .50 = 'large').

### 5.5.2.1  *How the report-detail and do-not-guess instructions impacted report informativeness*

In this section, the results are presented separately for the quantity and grainsize dimensions of report informativeness. Analyses are collapsed across the low- and high-number open question conditions (the effect that the number of questions has on narrative performance and recall uncertainty, is examined in Section 5.5.3, p. 134).

### 5.5.2.1.1  Quantity dimension of report informativeness

As expected, the report-detail instruction increased correct recall, $F(1, 188) = 27.35$, $p < .01$, $\eta^2 = .12$ (see Figure 5.4), and increased recall errors, $F(1, 188) = 10.74$, $p < .01$, $\eta^2 = .05$ (see Figure 5.5). Interviews with the report-detail instruction, had a large sized effect on producing more correct details ($M = 105.58$, $SE = 3.51$, $CI_{95}[99.03, 112.21]$), and a medium sized effect on producing more errors ($M = 9.19$, $SE = 0.51$, $CI_{95}[8.25, 10.11]$), than interviews without the instruction (correct details, $M = 79.55$, $SE = 3.62$, $CI_{95}[73.03, 86.16]$; errors, $M = 7.02$, $SE = 0.41$, $CI_{95}[6.22, 7.80]$).
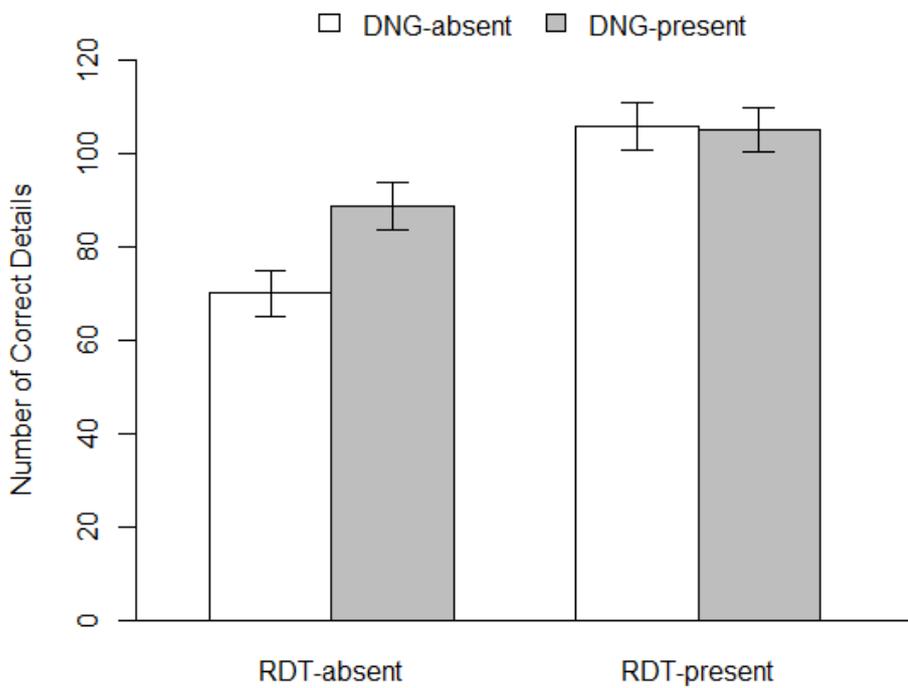


*Figure 5.4*. Mean correct recall (+SE) for each interview condition. RDT = report-detail instruction. DNG = do-not-guess instruction.

Unexpectedly, the do-not-guess instruction marginally increased correct recall, $F(1, 188) = 3.22$, $p = .08$, $\eta^2 = .01$ (see Figure 5.4), and did not influence recall errors, $F(1, 188) = 0.06$, $p = .80$, $\eta^2 < .01$ (see Figure 5.5). Interviews with the do-not-guess instruction, had a small sized effect on producing more correct details ($M = 97.03$, $SE = 3.56$, $CI_{95}[90.37, 104.06]$), than interviews without the instruction ($M = 88.10$, $SE = 3.98$, $CI_{95}[80.04, 94.96]$).

These results are unexpected because the do-not-guess instruction was expected to produce fewer correct details and fewer errors.
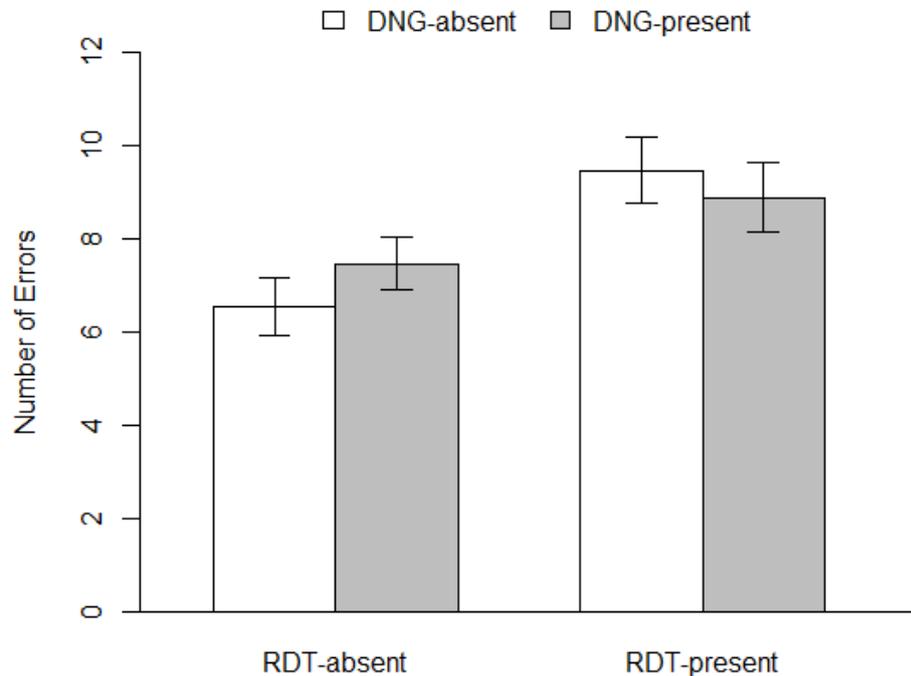


*Figure 5.5*. Mean errors (+SE) for each interview condition. RDT = report-detail instruction. DNG = do-not-guess instruction.

There was no interaction observed between the report-detail and do-not-guess instructions on recall errors, $F(1, 188) = 1.29$, $p = .26$, $\eta^2 = .01$ (see Figure 5.5), but a marginal interaction was observed on correct recall, $F(1, 188) = 3.75$, $p = .05$, $\eta^2 = .02$ (see Figure 5.4). This was explored with a simple effect analysis for: (a) the report-detail instruction, at each level of the do-not-guess instruction (i.e., absent, present); and (b) the do-not-guess instruction, at each level of the report-detail instruction (i.e., absent, present).

For interviews excluding the do-not-guess instruction, a large sized difference in correct recall, $U = 520.00$, $z = -4.63$, $p < .01$, $CI_{98.7}[-54.00, -19.00]$, $r = -.47$, suggested that more correct details were produced in the interviews with the report-detail instruction (*Mdn* = 102.00, *SE* = 5.14, $CI_{98.7}[94.00, 118.50]$), than the interviews without the report-detail instruction (*Mdn* = 69.00, *SE* = 4.68, $CI_{98.7}[54.00, 77.00]$). Moreover, for interviews including

the do-not-guess instruction, a medium sized difference in correct recall, $U = 816.50$, $z = -2.46$, $p = .01$, $CI_{98.7}[-34.00, 0.00]$, $r = -.25$, suggested that more correct details were produced in the interviews with the report-detail instruction ($Mdn = 100.50$, $SE = 8.47$, $CI_{98.7}[83.00, 119.00]$), than the interviews without the report-detail instruction ($Mdn = 80.50$, $SE = 3.90$, $CI_{98.7}[70.00, 95.00]$). These simple main effects suggest that the report-detail instruction increased correct recall, in spite of the do-not-guess instruction.

For interviews excluding the report-detail instruction, a medium sized difference in correct recall, $U = 773.00$, $z = -2.78$, $p = .01$, $CI_{98.7}[-34.00, -2.00]$, $r = -.28$, suggested that more correct details were produced in the interviews with the do-not-guess instruction ($Mdn = 80.50$, $SE = 3.90$, $CI_{98.7}[70.00, 95.00]$), than the interviews without the do-not-guess instruction ($Mdn = 69.00$, $SE = 4.68$, $CI_{98.7}[54.00, 77.00]$). In contrast, for interviews including the report-detail instruction, the do-not-guess instruction did not influence correct recall, $U = 1137.00$, $z = -0.11$, $p = .91$, $CI_{98.7}[-17.00, 19.00]$, $r = -.01$. These simple main effects suggest that the do-not-guess instruction only increased correct recall when the report-detail instruction was absent from interviews.

### 5.5.2.1.2 Grainsize dimension of report informativeness

As expected, the report-detail instruction influenced grainsize precision, $F(1, 188) = 12.03$, $p < .01$, $\eta^2 = .06$ (see Figure 5.6). Interviews with the report-detail instruction, had a medium sized effect on producing finer-grained reports ($M = 0.51$, $SE = 0.01$, $CI_{95}[0.48, 0.54]$), than interviews without the instruction ($M = 0.44$, $SE = 0.01$, $CI_{95}[0.42, 0.47]$).

Surprisingly, the do-not-guess instruction did not influence grainsize precision, $F(1, 188) < 0.01$, $p = .96$, $\eta^2 < .01$ (see Figure 5.6). This was surprising because it was expected that the do-not-guess instruction would produce coarser-grained reports. Moreover, there was no interaction observed between the report-detail and do-not-guess instructions on grainsize precision, $F(1, 188) = 2.36$, $p = .13$, $\eta^2 = .01$ (see Figure 5.6).
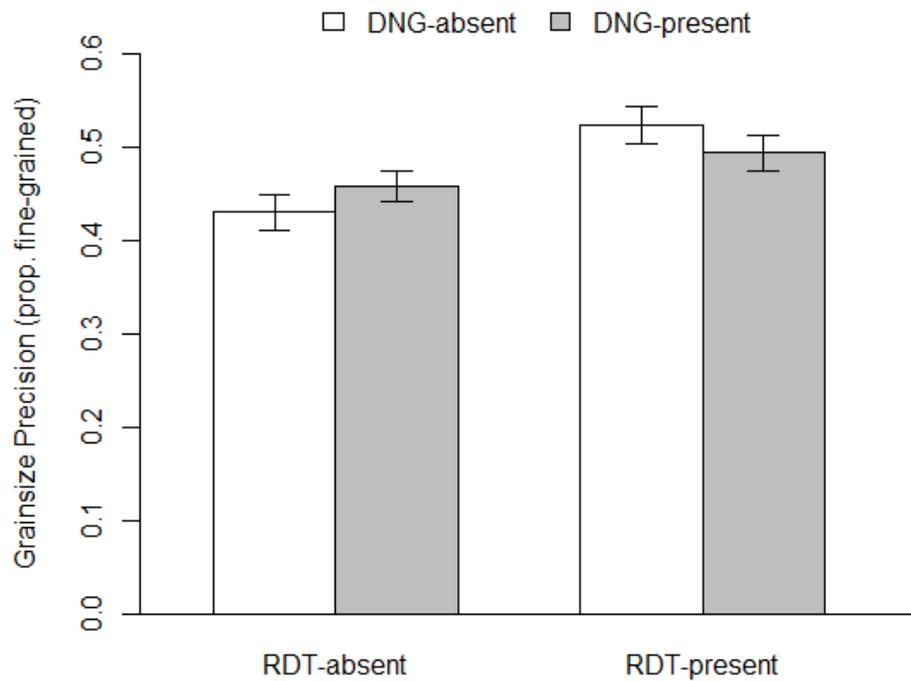
*Figure 5.6.* Mean grainsize precision (+SE) for each interview condition. RDT = report-detail instruction. DNG = do-not-guess instruction.

In summary, the effects observed on report informativeness (i.e., more correct details, more errors, and finer-grained reports) support the idea that participants responded to the report-detail instruction with unsatisficing knowledge. In contrast, the effects observed for the do-not-guess instruction (i.e., more correct detail, and no effect on errors or grainsize precision), do not support the idea that participants responded with conservative satisficing knowledge. Further, the interaction observed on correct recall suggests that the do-not-guess instruction was beneficial in producing more correct details, only when the report-detail instruction was absent from interviews. This indicates that an accuracy demand on narrative recall is potentially redundant in the presence of an informativeness demand.

### 5.5.2.2 *How the report-detail and do-not-guess instructions impacted report accuracy*

These analyses (collapsed across low- and high-number open question conditions) focused on a trade-off in report accuracy that is expected to occur with greater report informativeness, on either the quantity dimension (i.e., Koriat & Goldsmith, 1994) or grainsize dimension (i.e., Yaniv & Foster, 1995). Curiously, as Figure 5.7 shows, neither the

report-detail instruction, $F(1, 188) = 0.23$, $p = .64$, $\eta^2 < .01$, nor the do-not-guess instruction, $F(1, 188) = 1.42$, $p = .24$, $\eta^2 = .01$, influenced recall accuracy, and there was no interaction, $F(1, 188) = 0.06$, $p = .81$, $\eta^2 < .01$. These results suggest that the report-detail and do-not-guess instructions did not influence report accuracy.
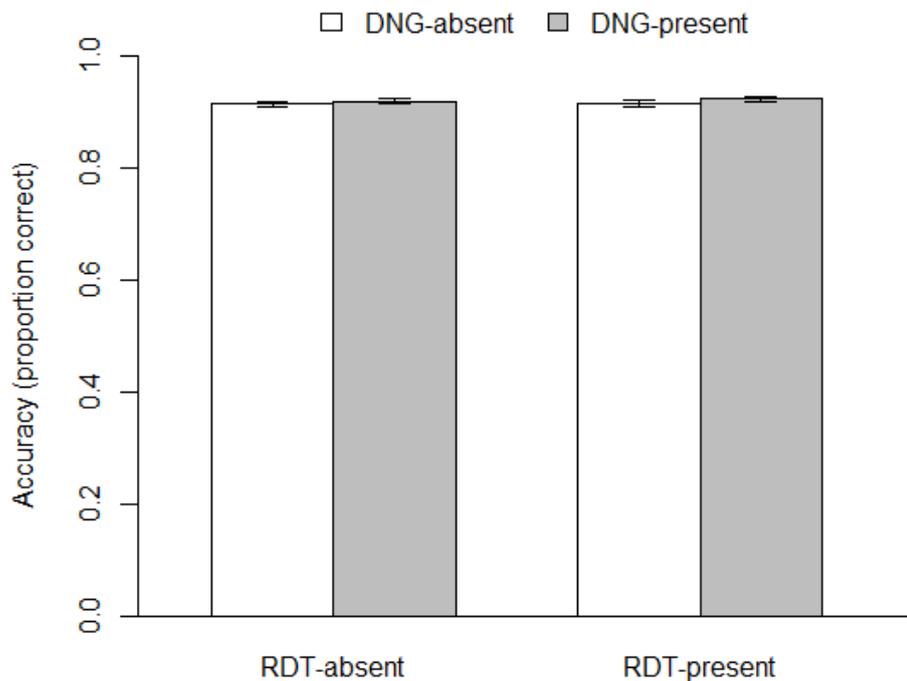


*Figure 5.7*. Mean recall accuracy (+SE) for each interview condition. RDT = report-detail instruction. DNG = do-not-guess instruction.

The results are curious because, as Figure 5.7 shows, a ceiling effect is once again observed, and it is similar to that found on recall accuracy in the previous experiments (see Section 3.2.2.2, p. 58; and Section 4.2.3.2, p. 95). This finding is interesting because it demonstrates how robust narrative accuracy is, regardless that a series of open questions were used (in the present study and Experiment 2) to avoid a ceiling effect on recall accuracy. Importantly, this robust accuracy suggests that a trade-off in narrative accuracy does not occur with greater narrative informativeness (cf. cued-recall performance; e.g., Goldsmith & Koriat, 1999, 2008; Koriat & Goldsmith, 1994, 1996; Weber & Brewer, 2008; Yaniv & Foster, 1995). This finding might be restricted to memory reports produced after a short temporal

delay (i.e., 30 min), however, the finding has broader theoretical and practical implications that will be addressed in the General Discussion (see Chapter 6).

In summary, the lack of effect on report accuracy does not support the idea that participants responded to the report-detail instruction with unsatisficing knowledge. This contrasts with the effects observed on report informativeness (i.e., effects are consistent with unsatisficing knowledge). How these disparate conclusions can be reconciled, will be examined in the Discussion (see Section 5.6.1, p. 138). For the do-not-guess instruction, the lack of effect on report accuracy does not support the idea that participants responded with conservative satisficing knowledge. This concurs with the conclusion drawn for the effects observed on report informativeness. The influence of the do-not-guess instruction on participants' accounts has implications for monitoring performance, and this will be discussed in Section 5.6.2 (p. 142).

### 5.5.2.3 *How the report-detail and do-not-guess instructions impacted recall uncertainty*

These analyses focused on finding evidence to support the idea that recall uncertainty is communicated with linguistic qualifiers. To recap, linguistic qualifiers are phrases (e.g., hedges and 'don't know' responses) that a witness might use to qualify details they are uncertain about. For example, a witness might be unsure about the colour of a thief's jacket, and express this doubt with a hedge (e.g., "The jacket was black, I think") or 'don't know' response (e.g., "The jacket was black but I don't know"). Importantly, the experience of doubt might occur with correct details and errors, although it is assumed errors will be qualified more frequently (i.e., Paulo et al., 2015). Therefore, to assess recall uncertainty, I was interested in the number of qualifiers associated with errors, as a proportion of the total number of qualifiers associated with correct details and errors.

As Figure 5.8 shows, neither the report-detail instruction, $F(1, 188) = 0.66$, $p = .42$, $\eta^2 < .01$, nor the do-not-guess instruction, $F(1, 188) = 0.01$, $p = .93$, $\eta^2 < .01$, influenced recall

uncertainty, and there was no interaction, $F(1, 188) = 1.04$, $p = .31$, $\eta^2 = .01$. These results do not support the hypothesised outcomes (i.e., the report-detail instruction would produce greater recall uncertainty, and the do-not-guess instruction would produce lesser recall uncertainty), nevertheless they are unsurprising in light of the report accuracy results. In other words, the lack of effect on recall uncertainty is consistent with the lack of effect on recall accuracy because qualifiers were expected to communicate a *change* in recall accuracy.
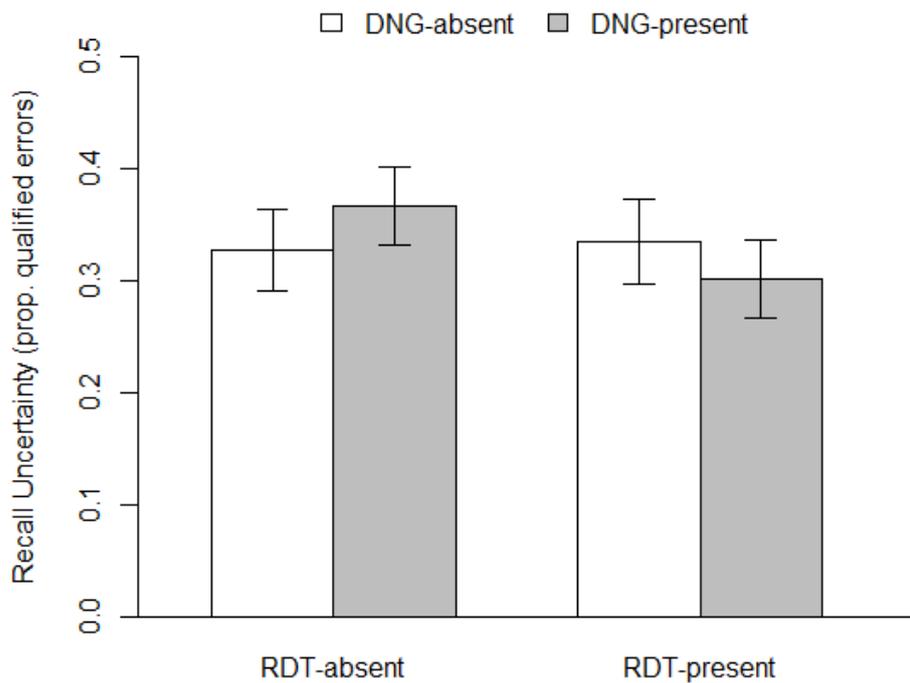


*Figure 5.8.* Mean recall uncertainty (+SE) for each interview condition. RDT = report-detail instruction. DNG = do-not-guess instruction.

### 5.5.3 How the number of open questions impacted narrative reports

After the second study, I grew interested in how the number of open questions used in an interview, might influence narrative recall. This interest arose from finding a ceiling effect on report accuracy, regardless that open questions were used in Experiment 2, to help observe a trade-off in accuracy that was expected to occur with greater report informativeness (i.e., Koriat & Goldsmith, 1994; Yaniv & Foster, 1995).

Certainly, it has been suggested that the number of questions used in an interview, will determine the quantity of details produced (Aschermann et al., 1991), and might influence report accuracy (Memon, Wark, Holley, et al., 1997). Yet, the effect of question number on narrative performance has not been systematically investigated (cf. Dando et al., 2011, who used question number as a covariate). Moreover, the literature commonly omits reporting the number of questions used in an interview (cf. Aschermann et al., 1991; Dietze et al., 2012). However, the number of questions might have important implications for meta-analytic research that reports on total interview performance (i.e., Köhnken et al., 1999; Memon et al., 2010) and, critically, the conclusions that can be drawn from the results.

Group differences were examined to elucidate how the number of open questions affected narrative performance. Differences between groups on recall accuracy, grainsize precision and recall uncertainty were assessed. The simple count variables (i.e., correct and errors) were not assessed because, intuitively, more information is produced when more questions are used. The Wicoxon rank sum test was used for recall accuracy (effect size was calculated with $r$) and the Student's $t$-test was used for grainsize precision (effect size was calculated with Cohen's $d$). Both tests were calculated using the *psych* package (version 1.4.8; Revelle, 2014) in R (version 3.2.2; R Core Team, 2014). A generalised Wilcoxon test was used for recall uncertainty, following Neuhäuser's (2010) approach, because data was skewed with unequal variance. Neuhäuser's approach uses the Brunner-Munzel test statistic, $W_{BF}$; effect size, $P$ (= P(X<Y) + 0.5*P(X=Y)); and a confidence interval for $P$. If the confidence interval contains the value '.5', then the result is not significant. The generalised Wilcoxon test was calculated using the *lawstat* package (version 2.4.1; Gastwirth et al., 2014) in R (version 3.2.2; R Core Team, 2014).

For significant group differences, point estimates of the mean (or median) are given with 95% confidence intervals. These were calculated using the bias-corrected and

accelerated bootstrap method in IBM SPSS Statistics 21, using 2000 permutations. Verbal description for effect size adopted Cohen's (1992) recommendations for $d$ (i.e., 0.20 = 'small', 0.50 = 'medium', 0.80 = 'large') and $r$ (i.e., .10 = 'small', .30 = 'medium', .50 = 'large'). For verbal description of the effect size $P$, $P$ was first converted to Somers' $D$ ($2*P-1$; Newson, 2002) to make it equivalent to effect size $r$ (Ferguson, 2009).

Question number had a medium sized effect on report accuracy, $U = 2980.00$, $n_1 = n_2 = 96$, $z = -4.23$, $p < .01$, $CI_{95}[-0.03, -0.01]$, $r = -.31$, a medium sized effect on grainsize precision, $t(190) = -3.72$, $p < .01$, $CI_{95}[-0.11, -0.03]$, $d = 0.54$, and a small to medium sized effect on recall uncertainty, $W_{BF}(146.07) = -2.44$, $n_1 = n_2 = 96$, $p = .02$, $CI_{95}[0.31, 0.48]$, $P = .40$.

Interviews that used nine open questions, produced reports that were less accurate ($Mdn = .91$, $SE < .01$, $CI_{95}[.90, .92]$), finer-grained ($M = .51$, $SE = .01$, $CI_{95}[.49, .54]$), and were communicated with greater recall uncertainty ($Mdn = .36$, $SE = .02$, $CI_{95}[.33, .40]$), than interviews that used three open questions (accuracy, $Mdn = .93$, $SE < .01$, $CI_{95}[.93, .94]$; grainsize precision, $M = .44$, $SE = .01$, $CI_{95}[.42, .47]$; recall uncertainty, $Mdn = .26$, $SE = .08$, $CI_{95}[.10, .38]$). These effects are interesting because open questions were used to invite narrative responses, thereby, allowing participants the freedom (i.e., report option) to select the memories to report, and to report only what they knew (i.e., Fisher, 2010).

Importantly, the effects suggest that a trade-off in accuracy occurred when participants produced finer-grained accounts (i.e., Yaniv & Foster, 1995). Moreover, the effects suggest that participants used linguistic qualifiers to express greater recall uncertainty in their diminished report accuracy. These findings are important because they suggest that it is the number of questions used in an interview – not different instructional demands – that will impact report accuracy (and the expression of uncertainty in the information communicated). Consequently, question number might account for disparate accuracy results that have been

reported in the literature (e.g., Bensi et al., 2011; Dando, Wilcock, & Milne, 2009; Memon, Wark, Holley, et al., 1997). For example, Memon et al. (1997) found interviewers, trained to use the Cognitive Interview, elicited less accurate reports (84%), than the reports (92% accuracy) elicited by interviewers who had no formal interview training. The researchers suggested that the number of questions used by trained and untrained interviewers might have influenced report accuracy, and the present findings support this. In future, literature should cite the number of questions used in interviews so that this influential variable can be examined.

## 5.6    Discussion

This study explored the theoretical relevance of Ackerman and Goldsmith's (2008) dual-criterion model to narrative recall in the investigative interview setting. The model was used in this context to determine if instructional demands, made with the report-detail and do-not guess instructions, impact witness knowledge state. Specifically, it was hypothesised that an informativeness demand applied with the report-detail instruction, would influence the minimum-informativeness criterion and induce unsatisficing knowledge (see Section 5.1.2, p. 107). In contrast, an accuracy demand applied with the do-not-guess instruction, was expected to influence the minimum-confidence criterion and induce conservative satisficing knowledge (see Section 5.1.3, p. 109). Accordingly, these knowledge states were expected to manifest an informativeness-accuracy trade-off in narrative performance.

The report-detail detail instruction produced more informative reports without compromising accuracy, suggesting that an informativeness demand does not induce unsatisficing knowledge. Contrary to expectations, the do-not-guess instruction produced more informative reports without affecting accuracy, suggesting that an accuracy demand does not induce conservative satisficing knowledge. The implications these findings have for witness metacognitive processes will be discussed.

### 5.6.1 Informativeness demand does not induce unsatisficing knowledge

To meet the informativeness demand of the report-detail instruction, the dual-criterion model predicted witnesses will use unsatisficing knowledge. This occurs because the witness uses a more informative minimum-informativeness criterion, to produce a more informative report. In doing this, however, the witness will violate their minimum-confidence criterion and produce a less accurate report. These hypothesised outcomes (see Section 5.3.2, p. 114) were partially supported.

Participants produced a greater quantity of information in finer-grained reports, suggesting they used a more informative minimum-informativeness criterion to report. However, a trade-off in report accuracy was not observed, suggesting that the participants did not violate their minimum-confidence criterion. In other words, it appears that the participants did not experience the criterion conflict that is a necessary condition for unsatisficing knowledge (i.e., Ackerman & Goldsmith, 2008). These findings suggest that the participants did not use unsatisficing knowledge to respond to the report-detail instruction, but maintained a satisficing knowledge state to produce their accounts. A possible explanation for why this occurred is the presence of the interviewer and, therefore, greater social consequences for reporting (e.g., McCallum et al., 2016). I expand on this point below.

#### 5.6.1.1 *Response privacy and knowledge state*

The term 'response privacy' refers to whether information is reported aloud in the presence of people, or privately and anonymously (Shaw, Appio, Zerr, & Pontoski, 2007). Recently, response privacy has been found to influence the reporting of fine-grained information by witnesses (McCallum et al., 2016). The researchers found that participants, who privately and anonymously answered 20 closed questions (about a mock bank robbery), showed greater bias towards reporting fine-grained information, than participants who thought they would vocalise their responses to an audience (Experiment 1; McCallum et al.,

2016). Importantly, the researchers established that this grainsize effect was not due to private responders reporting fine-grained information more accurately. In other words, the goal for informativeness, and not accuracy, was driving the grainsize effect. These findings have important implications for elucidating why the participants, in the present study, appeared to respond to the report-detail instruction with satisficing knowledge (i.e., a criterion conflict was not experienced).

Memory reporting might be considered private and anonymous in Ackerman and Goldsmith's (2008) study because questions and answers were computer administered. In other words, there were fewer social consequences for reporting (i.e., McCallum et al., 2016) because there was no interviewer present to scrutinize answers. Therefore, satisficing knowledge, as defined by Ackerman and Goldsmith, might be most relevant to private and anonymous recall. For argument sake, I define this as 'private satisficing knowledge'. Importantly, private satisficing knowledge might exist when a satisficing demand (e.g., the "Describe what happened") occurs privately and anonymously (e.g., if the instruction is computer administered; see Figure 5.5). That is, the minimum-confidence criterion might be used to report in private anonymity (i.e., Ackerman & Goldsmith, 2008). Consequently, when an informativeness demand is made (e.g., "Describe everything in as much detail as possible"), a person will respond with unsatisficing knowledge if they violate their minimum-confidence criterion to produce finer-grained responses (i.e., Ackerman & Goldsmith, 2008).

In contrast, in an interview setting, memory reporting might be considered public because the interviewer is present to scrutinise what is being said. Moreover, if testimony is recorded (e.g., video recorded), the witness might think this will be scrutinised at another time by additional people[22]. If the interviewer's presence leads a witness to produce coarser-

---

[22] Shaw, et al. (2007) found that witnesses public responses only differed to their private responses when there was more than one person present.

grained information (i.e, McCallum et al., 2016), this has implications for satisficing knowledge when a satisficing demand is made (e.g., "Tell me what happened"). For the purpose of definition, I define this knowledge state as 'public satisficing knowledge". Importantly, with public satisficing knowledge, the minimum-confidence criterion might not be used for reporting, because a coarser informativeness criterion is used instead (see Figure 5.5). Critically, knowledge is satisficing because the most precise *public* response can be volunteered, that satisfies both confidence and informativeness criterions. Consequently, when an informativeness demand is made (e.g., with the report-detail instruction), a witness may use a finer informativeness criterion to report, without violating their minimum-confidence criterion. In other words, the witness maintains satisficing knowledge to produce more informative testimony without compromising accuracy (see Figure 5.5). Thus, the presence of the interviewer in the present study, might explain why participants appeared to maintain satisficing knowledge when they responded to the report-detail instruction.

An area for future research is to examine the impact that response privacy has on witness knowledge state and narrative performance. One way to operationalise response privacy, so that it is relevant to real-world policing, is to compare narratives in person (public response) with narratives over the phone (private response). Witnesses are often initially interviewed by police emergency call handlers (Gabbert, Hope, Carter, Boon, & Fisher, 2016), therefore, phone interviews play a critical role in an investigation. It would be interesting to know if phone interviews give a sense of private anonymity to witnesses, to produce a more informative report. Moreover, a mobile phone application is under development to interview the phone user (C. J. Cunningham, personal communication, May 30, 2016). How this tech-savvy interviewing method impacts response privacy and narrative recall, might have important implications for producing real-time highly informative reports.

# PRIVATE RESPONSE

### Satisficing Demand
### Private Satisficing Knowledge

### Informativeness Demand
### Unsatisficing Knowledge



# PUBLIC RESPONSE

### Satisficing Demand
### Public Satisficing Knowledge
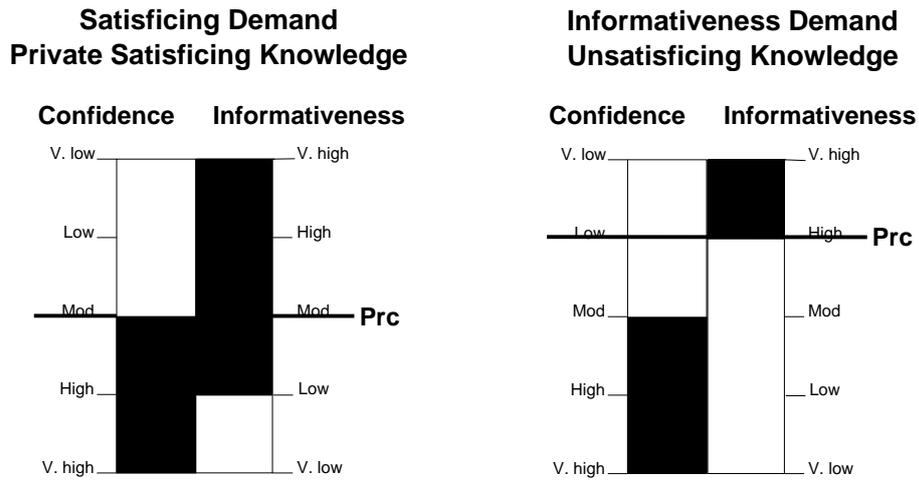
### Informativeness Demand
### Satisficing Knowledge
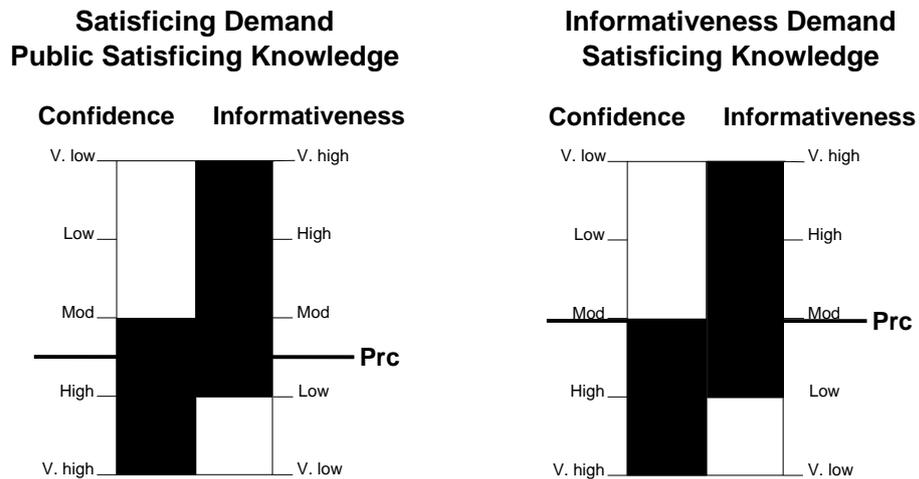


*Figure 5.5.* The effect that response privacy might have on knowledge state. In private, when a satisficing demand is made, a person responds with private satisficing knowledge and reports with their minimum-confidence criterion (Prc; top left diagram). Thus, when an informativeness demand is made, a person will respond with unsatisficing knowledge because they violate their minimum-confidence criterion to report with a more informative minimum-informativeness criterion (i.e., 'high' informativeness; top right figure). In contrast, when a satisficing demand is made in public, a person responds with public satisficing knowledge and reports with a more conservative confidence criterion (bottom left diagram). Hence, when an informativeness demand is made, a person can respond with a more informative minimum-informativeness criterion (i.e., 'moderate' informativeness; bottom right figure). That is, they do not need to violate their minimum-confidence criterion to report

### 5.6.2 Accuracy demand might impact monitoring performance

To meet the accuracy demand of the do-not-guess instruction, the dual-criterion model predicted witnesses will use conservative satisficing knowledge. The underlying assumption is that the do-not-guess instruction influences the confidence criterion, in the same way that an accuracy incentive does (i.e., penalty for errors; Goldsmith & Koriat, 2008; Koriat & Goldsmith, 1996). That is, to produce a more accurate report, the witness uses a more conservative minimum-confidence criterion. However, the minimum-informativeness criterion also plays a role in reporting and restricts the confidence criterion to some degree, to avoid an overly coarse-grained and uninformative report. None of the hypothesised outcomes (see Section 5.3.3, p. 114) were supported.

Contrary to expectations, participants produced *more* information, and the additional information comprised all correct details. Notably, grainsize precision was unaffected, suggesting that grainsize regulation did not produce the extra information. Theoretically, this is interesting because grainsize regulation is assumed to be a metacognitive control mechanism (i.e., Goldsmith et al., 2013). Hence, it appears that the do-not-guess instruction does did not impact participants' report threshold. In summary, it cannot be assumed that an accuracy demand influences narrative reporting similar to penalising errors (e.g., Goldsmith & Koriat, 2008; Koriat & Goldsmith, 1996). Instead, as will be discussed next, the data are consistent with improved monitoring performance.

#### 5.6.2.1 *Do-not-guess instruction and monitoring performance*

When people are rewarded for producing correct responses to closed questions (i.e., general knowledge), they (a) take longer to respond, and (b) recall more correct answers and not errors (i.e., Barnes, Nelson, Dunlosky, Mazzoni, & Narens, 1999). It was suggested that greater response latency showed people were more willing to search their memory for the correct response, and that they evaluated the information they knew best (Barnes et al., 1999).

Importantly, this willingness involves metacognitive monitoring because "if after a memory search, subjects either find no potential answer or are not confident enough in the correctness of a retrieved answer to output it [i.e., report it], then they must decide whether to continue searching for the answer" (p. 299, Barnes et al., 1999). In other words, people took the time to focus on the responses they were more sure were correct and to discriminate these from errors (i.e., their monitoring performance improved). It is possible that witnesses, when instructed not to guess, are more willing to search their memory for correct information and monitor the quality of their memory output more rigorously (i.e., Barnes et al., 1999). It is possible that these processes cause a witness to hesitate for longer (i.e., greater response latency), before giving their report.

In the present study, the do-not-guess instruction produced more correct details but not errors. These results are consistent with the idea that participants took the time to focus on the responses they were more sure were correct and to discriminate these from errors (i.e., Barnes et al., 1999). To add evidential weight to this interpretation, it was important to know if the participants hesitated for longer, before giving their reports. Subsequent analyses were performed to examine this possibility.

### 5.6.2.2 *How the do-not-guess instruction impacted response latency*

Response latency was measured with the time elapsed from when the interviewer spoke the last word of their instructions, to when the participant responded with the first word of their report. Interestingly, the do-not-guess instruction had a small to medium sized effect on response latency, $F(1, 188) = 5.45$, $p = .02$, $\eta^2 = .03$, with participants taking longer to respond in the interviews with the do-not-guess instruction ($M = 3.03$ ms, $SE = 0.16$, $CI_{95}[2.76, 3.32]$), than in the interviews without the instruction ($M = 2.54$ ms, $SE = 0.14$, $CI_{95}[2.28, 2.81]$). In conjunction with narrative performance (i.e., greater correct recall), this suggests that the participants took the time to focus on information they were more sure to be

correct, and improved their monitoring performance. In summary, the effects support the idea that the accuracy demand enhanced monitoring performance and did not, as predicted by the dual-criterion model, impact the decision to report. As an aside, it was found that the report-detail instruction did not influence response latency, $F(1, 188) = 0.94$, $p = .33$, $\eta^2 < .01$[23]. This suggests that the participants, in producing a greater quantity of information in finer-grained reports, did not take the time to search their memories more thoroughly for the extra information.

### 5.7    Summary of Key Findings

- Ackerman and Goldsmith's (2008) dual-criterion model was applied to narrative recall in the investigative interview setting, to predict how witness knowledge state manifests in narrative reports. This novel application of the model is theoretically important because it attempted to link our understanding of metacognition theory, established with closed question methodologies, to naturalistic witness behaviour.

- The dual-criterion model predicted that the report-detail instruction manifests unsatisficing knowledge in narrative reports. The instruction produced finer-grained reports with more detail (as predicted) without compromising accuracy (not predicted). These findings indicate that the report-detail instruction does not induce unsatisficing knowledge to produce a more informative report. This is a significant outcome, suggesting that the report-detail instruction is an important technique for interviewers to obtain informative and accurate reports.

- The dual-criterion model also predicted the do-not-guess instruction manifests conservative satisficing knowledge in narrative reports. The instruction produced more correct details (contrary to prediction) without affecting accuracy (not predicted). The findings indicate that the do-not-guess instruction does not induce unsatisficing

---

[23] Additionally, there was no interaction, $F(1, 188) = 0.94$, $p = .33$, $\eta^2 < .01$.

knowledge. Moreover, additional analyses found the do-not-guess instruction increased participant's response latency and, in conjunction with narrative performance (i.e., greater correct recall but not errors), this indicates that the instruction impacts monitoring performance.

- The number of questions used (in the interviews) influenced grainsize precision, accuracy, and the expression of recall uncertainty with linguistic qualifiers. These findings indicate that an informativeness-accuracy trade-off in narrative recall, is relevant to the number of questions used in an interview, and not the instructional demands made with different instructions.

**CHAPTER 6**

**General Discussion**

Criminal investigations rely on testimony that is both informative (to provide as many leads as possible to follow) and accurate (to avoid wasting police resources in chasing irrelevant leads). It is therefore a critical goal for an investigative interviewer to obtain an informative and accurate witness report. The Cognitive Interview is one of the best-practice methods to achieve this (Vrij et al., 2014) as evidenced by the robust finding that it produces a greater quantity of information without compromising testimonial accuracy (i.e., Memon et al., 2010). Whilst the Cognitive Interview comprises multiple techniques that are grounded in principles of memory and socio-communication practices (i.e., Fisher, Compo, Rivard, & Hirn, 2013; Fisher & Geiselman, 1992), it is less clear what role metacognition plays in producing testimony (cf. Granhag et al., 2004; Gwyer & Clifford, 1997; Mello & Fisher, 1996; Roberts & Higham, 2002). This is a key area to research because metacognitive processes have strategic importance in producing informative and accurate reports (Goldsmith et al., 2013). Notably, this understanding has derived from research methodologies focused on closed questions (cf. Evans & Fisher, 2011). This type of question has investigative value (e.g., N. Brewer et al., 2010), however, open prompts invite narrative responses that increase the likelihood of informative accurate recall and are of critical importance to use in an investigative interview (Fisher, 2010).

This thesis aimed to understand how metacognition mediates narrative recall when component techniques of the Cognitive Interview are used. If we can understand the metacognitive mechanisms that enhance recall performance, we can help investigative interviewers choose those techniques most suitable for producing informative and accurate testimony. To address the aim of this thesis, I developed a model of metacognition (see Figure 1.1, p. 12) to underpin the three empirical studies (Experiments 1 – 3). The model applies

Koriat and Goldsmith's (1996) framework of metacognition to narrative recall in witness interviews. This chapter presents an overview of the major findings (Section 6.1); the theoretical (Section 6.2) and practical (Section 6.3) contributions of the research; limitations and recommendations for future directions (Section 6.4); and conclusions (Section 6.5).

## 6.1 Overview of Major Findings

A summary of results is presented in Figure 6.1, according to the component interview techniques examined, namely (a) the mental-reinstatement-of-context instruction, (b) the naivety instruction, (c) the report-detail instruction, and (d) the do-not-guess instruction. Figure 6.1 highlights the two major outcomes of the research.

Report informativeness is a useful construct when taking a top-down approach in examining how metacognition mediates narrative recall. Importantly, the association between quantity and grainsize – the two dimensions of report informativeness – elucidates how instructional demands impact metacognition and narrative reports. In Figure 6.1, the size of the impact of instructional demands on narrative performance is listed from largest (i.e., report-detail and mental-reinstatement-of-context instructions) to smallest (i.e., do-not-guess instruction). The order of this listing is dependent on the sample size studied, however, it does highlight an important feature in participants' narrative performance – that is, grainsize precision did not follow the same trend as the quantity of recall. This is a significant observation because it highlights the value in evaluating both dimensions to report informativeness and, theoretically, it helps disentangle the metacognitive processes mediating narrative recall. Notably, as can be seen in Figure 6.1, grainsize precision is a determining indicator of the interview techniques that impact monitoring performance (i.e., mental-reinstatement-of-context and do-not-guess instructions) and the decision to report (i.e., report-detail and naivety instructions). The underlying assumption is that grainsize regulation is a metacognitive control mechanism (i.e., Goldsmith et al., 2013).

Second, participants responded to each technique and produced a more informative report without compromising accuracy. Notably, reports were highly accurate (92 – 94%) across all three studies. These are important findings because they shed light on why the Cognitive Interview does not produce a quantity-accuracy trade-off, and they suggest the prominent goal in narrative reporting is informativeness.

| INSTRUCTIONAL DEMAND | | |
|---|---|---|
| Mental-reinstatement-of-context (MRC) | | Report-detail (RDT) |
| | Do-not-guess (DNG) | Naivety (NVT) |

| | Best-candidate information retrieved | ▷ | Monitor | ▷ | Control | ▷ | Reported information |

| LONG – TERM MEMORY | METACOGNITION | NARRATIVE RECALL |

| Demand | Informativeness | | | Accuracy |
|---|---|---|---|---|
| | Correct | Errors | Precision | |
| RDT | ⇧⇧⇧ | ⇧⇧ | ⇧⇧ | ⇔ |
| MRC | ⇧⇧⇧ | ⇧⇧ | ⇔ | ⇔ |
| NVT | ⇧⇧ | ⇔ | ⇧⇧⇧ | ⇔ |
| DNG | ⇧ | ⇔ | ⇔ | ⇔ |
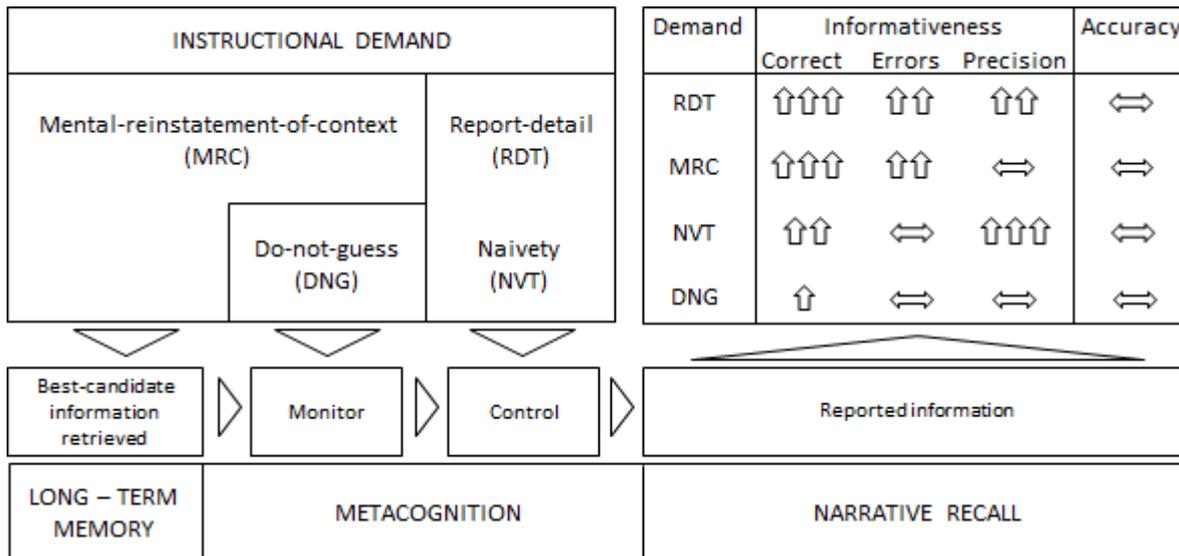
*Figure 6.1*: Research findings summarised on the model of metacognition used in this thesis. Narrative performance, especially grainsize precision, indicates: (a) the mental-reinstatement-of-context (MRC) instruction impacts memory retrieval and monitoring performance (Chapter 3), (b) the do-not-guess (DNG) instruction impacts monitoring performance (Chapter 5), and (c) the report-detail (RDT; Chapters 4 and 5) and naivety (NVT; Chapters 3 and 4) instructions impact the decision to report. Using Cohen's (1988) definitions, the number of upward-facing arrows reflects the size of the effect ($\eta^2$) observed (i.e., one arrow = small sized effect; two arrows = medium sized effect; three arrows = large sized effect). Horizontal arrows reflect no effect observed.

## 6.2   Theoretical Contributions

This thesis has made three main theoretical contributions. First, it helps clarify why the Cognitive Interview does not produce a quantity-accuracy trade-off (Section 6.2.1). Second, it helps explain the role that metacognition plays during cognitive interviewing

(Section 6.2.2). Third, it produced a model of narrative-report monitoring and control that is applicable to naturalistic witness behaviour (Section 6.2.3).

### 6.2.1   Why the Cognitive Interview does not produce an quantity-accuracy trade-off

All four interview techniques produced a more informative report without compromising accuracy (see Figure 6.1). This was an interesting finding because at the outset of this research, I speculated that the Cognitive Interview might not produce a quantity-accuracy trade-off if some techniques produce more information whilst other techniques improve accuracy. This speculation arose from existing research that suggests a trade-off in accuracy should occur with greater informativeness, be it in the quantity (i.e., Koriat & Goldsmith, 1994) or grainsize (i.e., Yaniv & Foster, 1995) dimensions.

It was especially curious to find that report accuracy was undiminished by the techniques that led participants to relax their report threshold [i.e., the report-detail (Chapters 4 and 5) and naivety (Chapters 3 and 4) instructions]. Certainly with cued-recall, it has been established that the report threshold is instrumental in producing performance trade-offs because the decision to report regulates informativeness and accuracy (i.e., Ackerman & Goldsmith, 2008; Evans & Fisher, 2011; Goldsmith et al., 2005; Goldsmith et al., 2002; Koriat & Goldsmith, 1996; Weber & Brewer, 2008). The present finding therefore suggests that the report threshold plays a crucial role in regulating *narrative* informativeness but not accuracy. This finding is significant for two reasons. Pragmatically, it demonstrates the importance of using open prompts to encourage accurate and informative testimony (i.e., Fisher, 2010). Theoretically, it supports the possibility that open prompts allow a witness to monitor information retrieved from memory, extremely well.

This thesis is not unique in finding that open prompts produce accurate recall (e.g., Evans & Fisher, 2011), however, it is unique in finding that testimonial accuracy is preserved when instructional demands produce more informative reports. This occurs if participants'

monitoring performance was exceptional. To explain, monitoring performance depends on the ability to discriminate correct information from errors (e.g., Koriat & Goldsmith, 1996, Experiment 2; Liberman & Tversky, 1993). However, there are two facets to discrimination ability. One relates to how well the monitoring assessments correspond to recall performance (i.e., calibration), and the other relates to how *polarised* these assessments are in their distribution[24] (the reader is referred to pp. 496-498 in Koriat & Goldsmith, 1996, for simulations showing the respective impact that these facets have on monitoring). To illustrate with cued-recall, Koriat and Goldsmith (1996) found that the test format of closed questioning – whether participants were required to recall or recognise answers – impacted monitoring performance. They found participants who recalled answers were more effective monitors because their monitoring assessments were more polarised than participants who recognised the answers. The implication for narrative recall is conjectural, however, it might be the case that open prompts produce more strongly polarised monitoring assessments. One line of evidence to support this conjecture comes from a recent study (Buratti et al., 2014) in the child witness literature.

Buratti et al. (2014) investigated the impact that question format has on children's (9 – 11 year-old) recall accuracy and confidence assessments. Specifically, the researchers compared responses to an instruction to report everything (about a film depicting a picnic) with responses to three open questions related to the film[25]. They found children's recall accuracy and confidence was higher for the instruction than open questioning. The researchers suggested that these findings can be explained if different question formats give a person

---

[24] On a calibration plot – a graph showing proportion correct recall against assessed probabilities – a dominance of responses at a particular confidence level (or levels) indicates assessments are polarised. Koriat and Goldsmith (1996) found that the calibration plot for their recognition participants showed a more equal distribution of assessments across all levels of confidence, whereas their recall participants showed stronger polarisation of assessments at the very high and very low levels of confidence.

[25] To obtain confidence accuracy data, the researchers first extracted units of information (e.g., "There was a black dog") from the children's interview transcripts, and then elicited confidence judgments for these units approximately 1 week later.

"various degrees of report option" (p. 190), with open instructions (e.g., "Report everything that you can remember") providing the greatest freedom, and closed questions (e.g., "What colour was his hair?") providing the least freedom, in deciding what to report. They suggested open questions (e.g., "What was he wearing") lie somewhere along the continuum between these end members. Therefore, the open questions – giving less report option than the instruction – may have pressured the children to respond with statements they were less certain about (Buratti et al., 2014). This conclusion is interesting and suggests that question format impacts the report threshold, however, a close inspection of the calibration curves presented by Buratti et al. (2014, p. 198-199) reveals important features consistent with effective monitoring performance: confidence assessments were strongly polarised for both question formats but they were more polarised for the instruction (to report everything). This suggests that the children were very effective in monitoring their responses to both question formats, but they were most effective in monitoring their responses to the instruction. This observation, in conjunction with Koriat et al.'s (1996) finding (that closed question recall is more polarised than closed question recognition), suggests question formats might lie on a continuum of monitoring polarization rather than report option. If open prompts (I use this term to include both question formats of Buratti et al.) produce strongly polarised monitoring, the potential impact on narrative recall performance is twofold.

First, strongly polarised monitoring assessments will produce reports heavily dominated by correct information. This might explain the persistent ceiling effect on narrative accuracy found in this thesis (Chapters 3 – 5) and other studies (e.g., Dando et al., 2011; Dando, Wilcock, & Milne, 2009; Mello & Fisher, 1996) that have cited narrative accuracy separately to total interview accuracy (i.e., comprising responses to open prompts and closed questions). Moreover, whilst some instructional demands [i.e., the mental-reinstatement-of-context (Chapter 3) and report-detail (Chapter 5) instructions] had a moderate impact on eliciting more errors (see Figure 6.1), strongly polarised monitoring may explain why there

was only a small absolute increase in the number of errors observed. Second, if monitoring assessments are strongly polarised, the decision to report more information (by relaxing the report threshold) will have very little impact on recall accuracy. This might explain why, in response to the report-detail instruction, participants appeared to maintain satisficing knowledge to produce more informative reports (Chapter 5).

In summary, the findings of this thesis support the idea that a quantity-accuracy trade-off is not produced with the Cognitive Interview because open prompts allow a witness to use strongly polarised monitoring assessments. Therefore, by virtue of excellent monitoring, the primary goal for narrative reporting appears to be informativeness and not accuracy.

### 6.2.2 The role that metacognition plays during a Cognitive Interview

According to the present research, the mental-reinstatement-of-context instruction impacts memory retrieval and the sensitivity of the monitoring mechanism to errors (Chapter 3); the do-not-guess instruction impacts monitoring (Chapter 5); and the report-detail (Chapters 4 and 5) and naivety instructions (Chapters 3 and 4) impact the decision to report. These are important findings because the literature reviewed in Chapter 2 focused on how the Cognitive Interview – as an omnibus method – influenced monitoring performance (i.e., Allwood et al., 2005; Granhag et al., 2004; Gwyer & Clifford, 1997; Mello & Fisher, 1996; Roberts & Higham, 2002). This literature suggests that the enhanced efficacy of the Cognitive Interview, against comparison interviews, is not attributable to metacognitive processes (i.e., Granhag et al., 2004; Gwyer & Clifford, 1997; Mello & Fisher, 1996). However, the present findings contradict this and suggest monitoring performance (Chapters 3 and 5) and the report threshold (Chapters 3 – 5) both play a role in the efficacy of the Cognitive Interview. These findings highlight the complexity in determining how multiple techniques combined, impact metacognition.

A surprise outcome of this thesis was that report centrality plays a role in leading a witness to relax their report threshold. Specifically, it was observed that the report-detail instruction produced more informative central and peripheral information, and the naivety instruction produced more informative central information (Chapter 4). Existing research suggests that cued-recall performance, whilst dependent on knowledge state (Ackerman & Goldsmith, 2008), response privacy (McCallum et al., 2016, Experiment 1) and memory quality (Sauer & Hope, in press), reflects competing goals for informative and accurate reports (i.e., Goldsmith et al., 2002; Koriat & Goldsmith, 1996; Weber & Brewer, 2008; Yaniv & Foster, 1995). Often, these authors note that these goals are guided by Grice's (1975) Maxim of Quantity (i.e., this guides the speaker to provide sufficient information to make their point) and Maxim of Quality (i.e., this guides the speaker to say only what they believe to be true). The present research suggests that a witness is also concerned with the centrality of information, and that this goal is guided by Grice's (1975) Maxim of Relation (i.e., this guides a person to focus on what is relevant to the topic of conversation).

Observing centrality in witness reports is not unique (e.g., Roberts & Higham, 2002; Wong & Read, 2011), however, it is significant to find evidence to suggest that report centrality may lead a witness to relax their report threshold. This indicates narrative performance reflect goals for informative <u>and</u> relevant reports. The point of difference between these findings, and existing research (i.e., Goldsmith et al., 2002; Koriat & Goldsmith, 1996; Weber & Brewer, 2008; Yaniv & Foster, 1995), is that the Maxim of Relation is likely irrelevant to closed questioning because this question format prompts specific information. That is, the specificity of a closed question (e.g., "What colour was the car?") will tell the witness what the relevant response should be (i.e., colour of the car). In contrast, open prompts (e.g., "Describe what happened?") seek a comprehensive account and so the witness must decide on appropriate information. Future research must continue to examine the impact that component techniques have on report centrality. The implication for

investigative interviewing is on producing good quality leads to follow, especially if central information is produced more accurately than peripheral information (i.e., Luna & Migueles, 2009; Paz-Alonso et al., 2013).

### 6.2.3   A model of narrative-report monitoring and control

The metacognitive model that underpinned this thesis (see Figure 1.1) applies Koriat and Goldsmith's (p. 494, 1996) framework of metacognition (see also p. 8 in Goldsmith & Koriat, 2008) to narrative recall in the investigative interview setting. Theoretically, the model makes a significant contribution because it links metacognition theory, established from closed question methodologies and traditional indices of metacognition (e.g., Ackerman & Goldsmith, 2008; Evans & Fisher, 2011; Goldsmith et al., 2005; Goldsmith et al., 2002; Koriat & Goldsmith, 1996; Weber & Brewer, 2008), to naturalistic witness behaviour.

Specifically, the model conceptualises the relationship between the interviewer, whose goal is to obtain informative and accurate testimony, and the witness, whose narrative report is mediated by monitoring and control processes. Importantly, the model highlights two features relevant to understanding witness metacognition in the context of investigative interviews.

First, the interviewer pursues their goal (for informative and accurate testimony) with instructional demands made with open prompts (e.g., "Describe what happened"). Open prompts are critical to an investigator because they increase the likelihood of informative and accurate recall (Fisher, 2010). As such, the investigator should take a funnel approach in their interview and use closed questions (e.g., "What colour was his hair?") to seek additional details, and clarity, about information elicited with open prompts (Fisher, 2010). Moreover, the majority of information (70 – 85%) is produced by a witness in the free-recall phase of an interview (e.g., Dando, Wilcock, Milne, et al., 2009; Roberts & Higham, 2002). Therefore,

the model makes a significant contribution because it is relevant to best-practice interviewing and the majority of witness recall.

Second, the witness' narrative report and, specifically, its informativeness, comprises both the quantity of details and grainsize precision. Examining both aspects of informativeness is an important contribution because existing research has focused on one aspect (cf. Evans & Fisher, 2011; Fisher, 1996), especially grainsize (e.g., Ackerman & Goldsmith, 2008; N. Brewer et al., 2010; Goldsmith & Koriat, 2008; Goldsmith et al., 2005; Goldsmith et al., 2002; Hope, Gabbert, & Brewer, 2011; Hope et al., 2010; Luna et al., 2011; Luna & Martin-Luengo, 2012; McCallum et al., 2016; Sauer & Hope, in press; Yaniv & Foster, 1995, 1997). However, a joint examination of quantity and grainsize is important because, although grainsize "will almost always be confounded with the quantity of details provided" (p. 508; Evans & Fisher, 2011), this thesis found that the quantity-grainsize relationship varies (see Figure 6.1) depending on the underlying metacognitive processes (see Figure 6.1).

### 6.3    Practical Contributions

This thesis has made two main practical contributions. For practitioners, it elucidates the impact that component techniques have on witness reports (Section 6.3.1). For researchers, it provides an approach to narrative grainsize analysis (Section 6.3.2).

### 6.3.1    Practical implications for practitioners

It is recommended that investigators consider the Cognitive Interview as a "toolbox of techniques, only some of which will be used in any specific interview" (p. 31, Fisher, 2010). Therefore, by understanding how component techniques influence narrative recall (see Figure 6.1), practitioners may choose the most appropriate technique(s) to use in their interview. The present research has made a significant contribution to this understanding because it

demonstrates how component techniques impact witness narratives. Moreover, the research

made a significant contribution with understanding how techniques impact report grainsize.

This is a significant outcome because published studies have not examined this, yet grainsize

has implications for producing helpful leads to follow in an investigation. For example, a

precise description (e.g., "He was wearing a long-sleeved black and white tracksuit jacket and

blue denim jeans") is more beneficial to help police apprehend a perpetrator than a coarse

description (e.g., "The person wore dark clothes"). Of note, it was encouraging to find that

none of the four techniques produced coarser-grained reports (see Figure 6.1). The following

summarises the implication that each technique has for investigative purposes.

### 6.3.1.1   *Mental-reinstatement-of-context instruction*

The mental-reinstatement-of-context instruction did not influence report grainsize,

however, in line with existing research (e.g., Dando, Wilcock, & Milne, 2009; Dietze et al.,

2012; Emmett et al., 2006), the instruction had a big impact on producing more correct

information (see Figure 6.1). This has implications for providing an investigation with a

potentially large number of good quality leads to follow, however, this is stated with a word

of warning. How important these potential leads are to investigators, depends on whether the

information is central or peripheral to an investigation (e.g., Roberts & Higham, 2002). There

is some evidence to suggest that the mental-reinstatement-of-context instruction produces

more central information but it has a stronger impact on producing more peripheral

information (Wong & Read, 2011). Therefore, if the best quality leads are related to central

information, then the mental-reinstatement-of-context instruction might have restricted

capacity in the number of good quality leads it produces. However, the investigative value of

leads will depend on the criteria used to measure the centrality of information (i.e., Geiselman

& Callot, 1990; Paz-Alonso et al., 2013). Of note, Wong and Read (2011) did not categorise

information according to its investigative value, hence their research findings (see above)

might have been different if they had done so. This thesis did not examine how the mental-reinstatement-of-context instruction affects report centrality and so this is an area for future research to pursue. Moreover, a specific line of enquiry is to examine if errors produced by the mental-reinstatement-of-context instruction (as observed in Chapter 3, see Figure 6.1) are related to peripheral details (i.e., Luna & Migueles, 2009).

### 6.3.1.2   Naivety instruction

An exciting outcome of this thesis was finding that the naivety instruction increased correct recall but not errors (see Figure 6.1) because this replicates what has been found with children (i.e., Brubacher et al., 2015). This is a significant observation because it suggests the efficacy in making a statement of naivety is robust across children and adults. Furthermore, the naivety instruction produced finer-grained reports (Chapters 3 and 4), and the finer-grained statements were centrally relevant to a criminal investigation (Chapter 4). This has important implications for how an investigation unfolds and its successful outcome. First, finer-grained information has the potential for producing better quality leads to follow, especially if the information is of central relevance to investigative purposes. Second, if central information is recalled more accurately than peripheral information (i.e., Paz-Alonso et al., 2013), the naivety instruction might help investigators elicit the best quality leads to follow. Because the naivety instruction produced such surprising results, it is strongly recommended that future research replicates these findings, especially in different population samples (e.g., the elderly, the intellectually challenged, etc.), and explores different ways to communicate naivety (e.g., "Imagine I am a movie producer but I have lost the movie script, so your account is the only one available to replace it").

### 6.3.1.3   Report-detail instruction

Curiously, as far as I know, the efficacy of the report-detail instruction has not been established to date. Yet, as highlighted in the Introduction (Chapter 1), police investigators

perceive the instruction[26] to be the most effective component Cognitive Interview technique to use, and they use it most frequently in interviews (Kebbell et al., 1999). Of the four instructions examined, especially when quantity and grainsize are conjointly considered, the report-detail instruction had the biggest impact on narrative informativeness (see Figure 6.1). This is a significant finding because it empirically supports practitioners' intuitions and interviewing habits. Moreover, the report-detail instruction produced more information of central relevance to an investigation but also more peripheral information (Chapter 4). The implications these findings have for investigative purposes is similar to the comments made above for the naivety instruction. That is, more details have the potential to produce more leads to follow, finer-grained statements have the potential for producing better quality leads, and central information has the potential to produce the best quality leads. However, peripheral information is also useful if it enhances witness credibility (i.e., Bell & Loftus, 1989).

### 6.3.1.4   *Do-not-guess instruction*

The do-not-guess instruction, although it produced more correct details, had the smallest impact on recall and it did not influence report grainsize (see Figure 6.1). Moreover, the instructional effect was moderated when the report-detail instruction was concurrently used by the interviewer (Chapter 5). The implication for investigators is that the do-not-guess instruction might be most useful to use during closed questioning because it appears to have the biggest impact on report accuracy, when this question format is used (i.e., report option; Koriat & Goldsmith, 1996)

---

[26] Specifically, practitioners rated the report-<u>everything</u> instruction but I consider the report-<u>detail</u> instruction to be operationally similar (see Section 3.3.2.2, p. 71, for my theoretical reasoning).

### 6.3.2    Practical implications for researchers

This thesis presents a novel methodology to code narrative grainsize (see Appendix E). Its methodological strength is that all reported information – about people, objects, surrounding and actions – is coded, including numerical (e.g., "about six foot" versus "six foot ten") and verbal descriptions (e.g., "top" versus "jacket"). Importantly, the same information is coded for quantity. Therefore, all information is examined on both dimensions (i.e., quantity and grainsize) of report informativeness (cf. Evans & Fisher, 2011, where it is not clear if all information, especially action descriptions, was assessed for grainsize). This methodological approach was instrumental in observing key differences in the relationship between quantity and grainsize. Critically, this relationship was of fundamental importance to the top-down approach taken to understand how component techniques impact narrative recall and metacognition. However, as noted by Evans and Fisher (p. 508), and keenly experienced by myself and the second coder, it is impossible to score narrative grainsize perfectly because there is no objective standard in the way people describe things in their own words. Nevertheless, the method used in this thesis is a novel contribution to research efforts in examining naturalistic witness behaviour. As a closing remark, I would like to echo Evans and Fisher's (2011) sentiment: "As more research is conducted researchers may define and measure precision in different ways. Should converging findings be reported despite the various imperfect measures of precision, this will provide strong evidence that our results are not a function of how precision was measured here" (p. 508).

### 6.4    Limitations and Directions for Future Research

Obtaining a witness report in a timely fashion is a challenge for real-life policing. For example, a witness might not come forward to police for several days, or a witness might have been injured and needs time to recover before giving their account. The risk these delays pose, and any delays in general, is that the witness is likely to forget some details about the

crime (e.g., Evans & Fisher, 2011). Importantly, research suggests that whilst report accuracy is unaffected, a temporal delay in interviewing produces a less informative report because it contains fewer details and is coarser-grained (Evans & Fisher, 2011). The implication for an investigation is that there will be fewer and poorer quality leads to follow. Therefore, it is critical that an investigator has techniques at their disposal to improve report informativeness after a temporal delay. One method – the SAI$^©$, a self-administered version of the Cognitive Interview – achieves this by helping to inoculate against the loss of information associated with a delayed interview (Gabbert et al., 2009). However, the efficacy of this method requires that a witness (a) has access to the SAI booklet, and (b) implements this self-directed activity in a timely fashion. Therefore, it is important that an investigator has other techniques to draw on when the SAI$^©$ is not used.

According to the present research, findings might only be relevant to testimony produced shortly after a crime is witnessed because the interviews were conducted soon after (i.e., 30 – 35mins) the stimulus event. So that the findings are more generalizable to real-world policing, it is relevant to understand how the loss of information associated with a delayed interview, impacts the efficacy of component instructions. There is evidence to suggest that even after a two week retention interval, the mental-reinstatement-of-context instruction increases the amount of correct information recalled by adults (Dietze et al., 2012). However, it would be good to know if the report-detail and naivety instructions also produce more informative reports (in both quantity and grainsize precision) after a temporal delay, and to what extent their impact is moderated by the degradation of memory that occurs over time. The implication for investigators is in deciding when to use a technique because, for example, there might be a temporal delay 'sweet spot' when the techniques have their biggest impact on informativeness.

In one of the original studies (Geiselman et al., 1985), investigating the efficacy of the Cognitive Interview, the researchers used four films depicting different crime scenarios (a bank robbery, a liquor store holdup, a family dispute, and a search through a warehouse). The researchers noticed that more informative reports – because they contained more correct details – were made about the bank and liquor store scenarios than for the family and warehouse scenarios. They examined this further with a post hoc analysis and found that the density of events was higher in the robbery and holdup scenarios because events happened at a faster pace, and several actions occurred simultaneously. In contrast, the density of events was lower in the family and warehouse scenarios because events happened more slowly, and they occurred in a sequential manner. The researchers suggested that the Cognitive Interview might produce more informative reports when a witness observes a greater density of events.

Surprisingly, there is no published study that has systematically investigated the impact that event density has on the efficacy of the Cognitive Interview (cf. the recall of complex events may be assisted by the timeline technique; Hope et al., 2013). Additionally, how grainsize relates to the association that Geiselman et al. (1985) observed between event density and correct recall, is unknown. However, event density is relevant to consider because a witness might observe a criminal activity involving a busy scene (e.g., multiple perpetrators, peak-hour traffic, etc.). Whilst it might be redundant for research efforts to examine this issue using the Cognitive Interview as an omnibus method – as Fisher (p. 36, 2010) states " we do not need any more validation tests of the [Cognitive Interview]" – it is relevant to consider how event density impacts the efficacy of component techniques. The implication for investigators is with deciding on the best technique(s) to produce a witness report. The crime scenario used in this thesis displayed a density of events that is arguably on the low side because the events happened at a relatively slow pace, and they occurred in a sequential manner. It would be interesting to know if the narrative recall effects observed for this event density are replicated, improved or diminished for a more complex event.

### 6.5    Conclusions

Since 1984, the Cognitive Interview literature has demonstrated the robust effects of cognitive interviewing, to comparison interviews, on producing a greater quantity of information without compromising accuracy. Further, the literature demonstrated that the enhanced efficacy of the Cognitive Interview is not attributable to metacognition. Scant attention, however, had been paid to the grainsize of the information nor the impact of component techniques on metacognition. This thesis contributed to the literature by (a) examining why a quantity-accuracy trade-off is not observed, and (b) by demonstrating how component techniques impact both the quantity and grainsize of testimony. Further, this thesis contributed by developing a model of narrative-recall monitoring and control, and demonstrating that metacognition plays an important role in explaining the efficacy of the Cognitive Interview. This present research suggests that the dominant goal in narrative reporting is informativeness, and some techniques have a greater impact on producing informative reports. This thesis therefore provides a guide for investigators regarding the best techniques to produce informative testimony. As investigators consider the Cognitive Interview as a toolbox of techniques, selecting the technique(s) most suited for their investigative purposes, further research needs to inform the limits of the conclusions that can be drawn about the effects that component techniques have on narrative recall and witness metacognition.

# APPENDIX A

## Picture Stimulus used in Experiment 1

# APPENDIX B

## Interview Protocols

### Experiment 1 Film Stimulus Interview Conditions

**Mental-reinstatement-of-context instruction absent, naivety instruction absent**

"I understand you have watched a video showing a crime. When you feel ready, tell me everything that you saw and heard on the video and describe this to me in as much detail as you can. If you cannot remember something do not guess."

"Is there anything else you would like to add?"

**Mental-reinstatement-of-context instruction present, naivety instruction absent**

"I understand you have watched a video showing a crime. I want you to describe to me what you saw and heard but first I'd like you to remember being seated to watch the video. Sometimes it helps us to remember things if our eyes are closed or we stare at a blank spot in front of us. So if you would like to, you can close your eyes or stare at the wall while you remember what happened on the video (pause). What I want you to do is create a picture in your mind of what you witnessed. What could you see? What could you hear? Think about how you were feeling while you were watching the video (pause). What thoughts you were having (pause). When you feel you have a really good picture in your mind of what happened, tell me everything that you saw and heard and describe this to me in as much detail as you can. If you cannot remember something do not guess."

"Is there anything else you would like to add?"

**Mental-reinstatement-of-context instruction absent, naivety instruction present**

"I understand you have watched a video showing a crime. I have not seen it myself so I do not know what you witnessed. You have all the information and I'd like you to share it with me. When you feel ready, tell me everything that you saw and heard on the video and describe this to me in as much detail as you can. If you cannot remember something do not guess."

"Is there anything else you would like to add?"

**Mental-reinstatement-of-context instruction present, naivety instruction present**

"I understand you have watched a video showing a crime. I have not seen the video myself so I do not know what you witnessed. You have all the information and I'd like you to share it with me (pause). I want you to describe to me what you saw and heard but first I'd like you to remember being seated to watch the video. Sometimes it helps us to remember things if our eyes are closed or we stare at a blank spot in front of us. So if you would like to, you can close your eyes or stare at the wall while you remember what happened on the video (pause). What I want you to do is create a picture in your mind of what you witnessed. What could you see? What could you hear? Think about how you were feeling while you were watching the video (pause). What thoughts you were having (pause). When you feel you have a really good picture in your mind of what happened, tell me everything that you saw and heard and describe this to me in as much detail as you can. If you cannot remember something do not guess."

"Is there anything else you would like to add?"

### Experiment 1 Picture Stimulus Interview Conditions

**Naivety instruction absent**

"Can you please describe what is on this card."

**Naivety instruction present**

"Can you please describe what is on this card. I do not know what is on the card so I only have your description to rely on."

## Experiment 2 Interview Conditions

**Report-detail instruction absent, naivety instruction absent**

"I understand you have watched a video showing a crime. When you feel ready, tell me what you saw and heard on the video."

"Is there anything else you can add?"

**Report-detail instruction present, naivety instruction absent**

"I understand you have watched a video showing a crime. When you feel ready, tell me everything that you saw and heard on the video and describe this to me in as much detail as you can."

"Is there anything else you can add?"

**Report-detail instruction present, weak naivety instruction present**

"I understand you have watched a video showing a crime. I have not seen it myself so I do not know what you witnessed. You have all the information so I only have your description to rely on. When you feel ready, tell me everything that you saw and heard on the video and describe this to me in as much detail as you can."

"Is there anything else you can add?"

**Report-detail instruction present, strong naivety instruction present**

"I understand you have watched one of several videos showing different crimes. I have not seen any of the videos myself and I don't know which one you have watched. I'm one of several interviewers and this is my first interview of the day so I have not heard from other people what they have watched. I do not know what you witnessed. You have all the information so I only have your description to rely on. When you feel ready, tell me everything that you saw and heard on the video and describe this to me in as much detail as you can."

"Is there anything else you can add?"

**Report-detail instruction absent, weak naivety instruction present**

"I understand you have watched a video showing a crime. I have not seen it myself so I do not know what you witnessed. You have all the information so I only have your description to rely on. When you feel ready, tell me what you saw and heard on the video."

"Is there anything else you can add?"

**Report-detail instruction absent, strong naivety instruction present**

"I understand you have watched one of several videos showing different crimes. I have not seen any of the videos myself and I don't know which one you have watched. I'm one of several interviewers and this is my first interview of the day so I have not heard from other people what they have watched. I do not know what you witnessed. You have all the information so I only have your description to rely on. When you feel ready, tell me what you saw and heard on the video."

"Is there anything else you can add?"

## Experiment 2 Open Questions

Questions were counter-balanced across interview conditions.

**Q1**: Please describe what object was stolen?

**Q2**: Please describe what the thief was wearing?

**Q3**: Please describe the vehicle parked in front of the restaurant?

## Experiment 3 Interview Conditions

**Report-detail instruction absent, do-not-guess instruction absent**

"I understand you have watched a video showing a crime. When you feel ready, tell me what you saw and heard on the video."

"Is there anything else you can add?"

**Report-detail instruction present, do-not-guess instruction absent**

"I understand you have watched a video showing a crime. When you feel ready, tell me everything that you saw and heard on the video and describe this to me in as much detail as you can."

"Is there anything else you can add?"

**Report-detail instruction absent, do-not-guess instruction present**

"I understand you have watched a video showing a crime. When you feel ready, tell me what you saw and heard on the video. If you cannot remember something do not guess."

"Is there anything else you can add?"

**Report-detail instruction present, do-not-guess instruction present**

"I understand you have watched a video showing a crime. When you feel ready, tell me everything that you saw and heard on the video and describe this to me in as much detail as you can. If you cannot remember something do not guess."

"Is there anything else you can add?"

## Experiment 3 Open Questions (Low Question-number)

Questions were psuedo-randomised and counter-balanced across half the interviews.

**Q1**: Please describe the object that was stolen?

**Q2**: Please tell me what the person, whose item was stolen, said to the waiter?

**Q3**: Please describe what the thief looked like?

**Experiment 3 Open Questions (High Question-number)**

Questions were psuedo-randomised and counter-balanced across half the interviews.

**Q1**: Please describe the object that was stolen?

**Q2**: Please describe where the waiter stood in the restaurant?

**Q3**: Please describe the vehicle parked in front of the restaurant?

**Q4**: Please describe what the thief looked like?

**Q5**: Please describe what the person, whose item was stolen, looked like?

**Q6**: Please describe what the waiter looked like?

**Q7**: Please tell me what the person, whose item was stolen, said to the waiter?

**Q8**: Please tell me what the thief wanted the waiter to do?

**Q9**: Please describe what the thief did when he left the restaurant?

**APPENDIX C**

**Film Stimulus Quantity Scoring Rules, Coding Key and Examples**

**GENERAL RULES FOR SCORING**

1. Details are categorised as Person, Action, or Surrounding information.

2. When scoring Action information, the basic rule is: who / did what / (to, with) what e.g.

   a. "The customer put the credit card onto the plate". Customer (who) put the (did what) credit card (what) onto the (did what) plate (what).

   b. In addition a person's location is scored under Action information e.g., "The waiter is behind the counter" is given 3 points for: he(1) is behind (1) the counter (1)

3. Score each detail as:

   a. Correct,

   b. Incorrect (e.g., saying the jacket was red when it was it was in fact black), or

   c. Confabulated (stating information that was not present or did not happen).

4. Only score the first time a detail is mentioned (e.g., do not score "counter" each time it is mentioned by the same participant. Alternatively, if the counter is referred to as a "bar" then as a "counter" then as a "bench" these alternative names are all correct for the same object. Therefore, only give a single correct score, not three).

   a. The exception is when a detail is part of the location of another detail. For example, "counter" may have been mentioned already such as in "the waiter was at the counter" and it is given 1 point under Surrounding but then the participant might then say "the cash register was on the counter". This statement would be given 3 points (i.e., for "cash register", "was on", "the counter") because it describes the location of the cash register.

5. Do not score suppositions because these include details that cannot be verified from the film (e.g., time of day, "He went back to his seat"). Do not score descriptions about feeling or thinking states of the people (e.g., "he looked nervous" or "he was clever"). Do not score interpretations about a person's actions (e.g., "to create a diversion") or opinions (e.g., "he was rugged looking")

6. Do not score digressions and vague answers.

7. If the participant corrects themselves, then only score the detail used to correct their statement (e.g., if the shop is referred to as a bar then this is incorrect but if the shop is later referred to as a restaurant then delete the first score and score "restaurant" instead).

8. If the participant retracts a detail then do not score it (e.g., "It was a restaurant, no, no, it wasn't").

9. If a participant gives an 'either/or' option (e.g., "It was either black or brown") without making a preference to one of the details (e.g., "It was either black or brown but I think it was black"), then score the correct detail. If both details are incorrect then score either one. If both details are correct then score either one.

10. Qualifiers that are used to express confidence (e.g., "probably", "not sure") are ignored (e.g., "I'm not sure but I think he was wearing a black jacket" is treated as "He was wearing a black jacket").

11. If an detail is referred to by more than one description that denotes a different grainsize, then score only the finer-grained response for correctness, e.g.,

    a. Toyota Land Cruiser is finer-grained than 4WD which is finer-grained than vehicle.

    b. Restaurant is finer-grained than building.

    c. American Express is finer-grained than credit card which is finer-grained than card.

    d. Indian is finer-grained than Asian which is finer-grained than non-Caucasian.

    e. Saturday is finer-grained than weekend.

    f. Rome is finer-grained than Italy which is finer-grained than Europe.

    g. White is finer-grained than light; black is finer-grained than dark. However, when light or dark are used in conjunction with a colour, the colour and tone are given a single combined scored (e.g., 'light blue' is 1 point, 'blue' or 'light' on its own is given 1 point).

    h. Nike sports jacket is finer-grained than bomber jacket which is finer-grained than jacket.

12. Finer-grained rule takes precedence over the 'either/or' rule (see point 9).

13. No points are awarded to the title given for the observable actors (e.g., "waiter", "worker", "thief", "customer") however 1 correct point is given for a person being identified.

14. When direct quotes are given, then score for the gist of what is stated and for any key specific words (e.g., he said "do you have any availability for Friday" would be given 1 correct point for asking about availability, 1 correct point for "for" because it is indicating

it is the upcoming Friday, but 1 incorrect point for "Friday" because he actually says "Saturday" in the film)

15. For scoring plurality then if the object is correct then a correct point is awarded but if the plurality is incorrect then an incorrect point is also awarded (e.g., "napkins" will be awarded 1 correct point as there is a napkin, but also 1 incorrect point is awarded because there is only <u>one</u> napkin).

16. Do not score any speculation made about how old the film is (i.e., when it was produced).

17. The free-narrative recall is scored according to the rules above.

18. The narrative responses to the open questions, are scored using the same general rules as above but only new information is scored (i.e., information that has not already been stated in the free-narrative recall). However, the following rules also apply to the open questions:

19. If a detail repeats information that was given in the free-narrative, it is scored only if it is now correct and it was incorrect in the free-narrative or vice versa, or it is now finer-grained (i.e. grainsize preference rule). Importantly, if this occurs, then to avoid double-dipping on scoring the detail, the free-narrative score must be removed (so only the open question score is retained).

20. Information has to be directly related to the open question and not repeat any information supplied by the question. For example, for the question "Describe the vehicle parked in front of the restaurant?", if the answer given is "A van was parked in front of the restaurant", then only score "van" in this instance (because the location of the van is supplied by the question).

21. Because each open question was designed to probe a specific category of information, the following is the relevant information to score for each question:

- "Please describe the object that was stolen?" (Surrounding)
    - Score details related to the description of the credit card and its location (i.e., on a napkin/plate/counter and additional locations).
- "Please describe where the waiter stood in the restaurant?" (Surrounding)
    - Score details related to the description of objects (e.g., cash register, telephone, etc.,) and their location (e.g., "The cash register was to the left-hand side of the waiter").
- "Please describe the vehicle parked in front of the restaurant?" (Surrounding)
    - Score details related to description of the vehicle and its location (other than "in front of the restaurant").

- "Please describe what the thief looked like?" (Person)
    - o Score details related to the clothing and physical attributes (e.g., hair colour, height, etc.) of the thief.
- "Please describe what the person, whose item was stolen, looked like?" (Person)
    - o Score details related to the clothing and physical attributes (e.g., hair colour, height, etc.) of the customer.
- "Please describe what the waiter looked like?" (Person)
    - o Score details related to the clothing and physical attributes (e.g., hair colour, height, etc.) of the waiter.
- "Please tell me what the person, whose item was stolen, said to the waiter?" (Action)
    - o Score details related to what the customer said to the waiter or what the customer wanted the waiter to do (e.g., "He wanted the waiter to start a tab").
- "Please tell me what the thief wanted the waiter to do?" (Action)
    - o Score details related to what the thief wanted and also what the waiter did in response to what the thief wanted (e.g., "He turned to check the diary").
- "Please describe what the thief did when he left the restaurant?" (Action)
    - o Score details related to leaving the restaurant, including exiting the door and running away.

## Conversation throughout Film

### (C = Customer, W = Waiter, P = Perpetrator)

*When waiter serves customer*

C: "Hi man how you going?"

W: "Hi."

C: "Just, ah, can you put my account on the card" (shows credit card)

W: "Ok, yep, not a problem, I'll bring it out to you."

C: "Cheers, ok, thanks."

*1st request for restaurant availability by perpetrator*

W: "Hi."

P: "How's it going there?"

P: "Yeah, I was just wondering, I've got a few friends coming from overseas and, um, I was thinking that since, that I've, I've heard from some friends that it's a really good restaurant here and we'd like to try it out.

W: "Ok."

P: "So, um, I was thinking what your availability is like on, ah, Saturday night?"

W: "Hold on a minute and I'll check for you."

W: "Yep, that's fine."

### 2nd request for restaurant availability by perpetrator

P: "Oh right, um, yeah, um, well I'm not too sure whether they'll be coming this Saturday. They might even be coming next Saturday. I think that would be the 22$^{nd}$. How are you looking for the 22$^{nd}$ there?"

W: "Mmm, I know we are busy for the next month. I can check that for you. One moment."

P: "Thanks."

W: "That's fine, yep. One moment" (*phone is ringing*)

P: nods

### When waiter answers phone

W: "Good afternoon, Capriccio's restaurant, this is Mario."

W: "Yes, he's here."

W: "Ok. Just one moment."

W: "Frank, there's someone heh, there's someone on the phone for you." (Frank responds "alright")

## PERSON INFORMATION (Perpetrator)

Person (1)

> *NB: any title given to the waiter is given 1 correct score for Person or generally stating there is a person*

Gender: Male (1). Also award 1 correct point for Person (if a title has not been given).

Age: Any mention of guy being in his twenties (1) e.g. any specific age given that is <25yo such as "about 23"; "early twenties"; "twenties"; "young adult"; "young man"; "young gentleman". 'Same age as others' is also correct.

> *NB: 'young' is too vague so it is not scored (e.g., as in "young person", "young guy")*

Build: Slim / medium / average / skinny / not fat (1)

> *NB" ' not very strong build is too vague to score'*

<u>Height</u>: Within 172 +/-3cm (5 foot 7 inches +/- 1 inch) or giving a relative height (shorter than customer and/or waiter; shorter than average Australian (178.4cm wikianswers.com)) (1)

    *NB: 'average height' / 'not too tall not too short' are too vague to score*

<u>Weight</u>: Within 75 kgs +/-4kg (1)

<u>Complexion</u>: olive / swarthy / dark (1) (i.e., not white)

<u>Facial</u>: Oblong face (1) long nose (1) straight nose (1) thick (1) dark / black eyebrows (1) dark / brown eyes (1) no glasses (1) clean shaven (1) looks like Harold Kumar (1)

<u>Ethnic appearance</u>: Non-Caucasian / Asian / Indian / Chinese / Australian (1)

<u>Hair</u>: Thick (1) dark / black (1) straight / flat (1) collar length / short / medium (1) part in middle (1) Paul McCarthy style / bowl cut / mop (1) came to his forehead / it's not to his eyes (1) longer than waiter's (1) no hat (1)

    *NB: shoulder length is 1 incorrect*

<u>Clothing</u>: dark / brown (1) ankle length (1) lace-up / closed (1) shoes / boots (1)

Light / blue / faded / acid wash (1) long (1) denim (1) pants / jeans / trousers (1) loose fitting / not tight (1)

Casual (1)

    *NB: well groomed / scruffier are suppositions*

    *NB: track pants (1 incorrect – do not score 'track' and 'pants' separately)*

    *NB: he is not wearing a belt (1 confabulated)*

    *NB: trousers is correct (google images displays jeans as well as more formal*

Dark or black (1) and light / white (1) long-sleeved (1) top / jacket / sports jacket / 'Nike' tracksuit top / jacket (1) large / puffy / loose (1) zipped up / mostly closed (1) with black (1) collar (1) and thick (1) vertical (1) stripes (1).

    *NB: if "tracksuit" then award just 1 correct point correct for 'tracksuit'*

    *NB: "sports wear" is 1 correct point.*

    *NB: 'hoodie', 'hooded' is confabulation – assume when this is stated it means the hood is on. 'bomber jacket' is 1 incorrect*

    *NB: 'high wasted' is supposition because the waist is not seen.*

Bottom half of jacket (1) is black (1) band (1) top half of jacket / down front (1) is white (1) and black (1).

Black (1) large (1) writing (1) 'NIKE' lettering (1: award point for each letter mentioned as either correct or incorrect) across back / shoulder height (1) made of tape (1) thick (1) vertical (1) stripes (1)

> *NB: 'K' / 'E' (1) is on RHS (1).*

White (1) sleeves / arms / arm pits (1) with thick (1) black (1) stripes (1) down both arms / on the sides / shoulders / on top (1) of back (1). Nike tick / symbol / logo / shapes / stylization (1) in centre (1) on back (1) is black (1) and large / 10-12cm in size (1) Nike tick / symbol / logo / design / shape / branded (1) on front / breast (1) is black (1) and smaller (1).

Jacket is made of lightweight or silky / parachute-type / shell-type ("shellcity") / polyester material / sporty material / water proofy / rain coat type material (1).

Jacket syle is irregular / unusual / patterned / distinctive (1).

> *NB: Age of jacket is supposition (can't verify).*

Two (1) colour (1) purple / lilac / mauve / dark (1) round neck (1) of top or T-shirt (1) with large (1) yellow (1) text (1) underneath (1) jacket.

No scarf (1)

Artefacts: Slight / thin (1) gold (1) necklace / chain (1) around neck (1) with a 10c or small sized (1) round (1) gold (1) pendant / medallion (1). No glasses (1). No watch (1). No scars (1)

Voice: Australian accent (1) fluent / perfect / first language (1) English (1)


**ACTION INFORMATION (Perpetrator)**

He (1) comes into the scene / in the street (1) on foot (1)

He (1) did not come (1) in a vehicle (1)

He (1) walks (1) across / on / in (1) the road (1) from the RHS (1) towards (1) the restaurant (1)

He (1) passes (1) behind (1) the 4WD / line of cars (1) and stoby pole (1) across (1) footpath (1) under (1) verandah (1)

He (1) looks (1) around / about / across / back (1) at the entrance / door / street / outside (1) the restaurant (1) four times (1)

He (1) opens / touched (1) the door (1)

He (1) walks (1) in / into (1) restaurant / foyer / reception / bar area (1) from LHS (1) about same time as customer (1).

He (1) walks (1) to / towards the (1) counter (1) and stands (1) in the background (1) *while the customer is being served*

> *NB: this action occurs prior to him being served by waiter so he does not approach waiter at this stage*

> *NB: He doesn't see the customer give over his card so "he saw him give his card" is he (1 confabulated) saw (1 confabulated) him (1 correct) give (1 correct) his (1 correct) card (1 correct)*

He (1) waits (1)

He (1) is behind (1) the customer (1) *or* Customer (1) is before (1) thief (1)

He (1) looks (1) around / about / down (1)

He (1) approaches / walks / moves (1) towards / comes up (1) to the counter / waiter / cash register (1) touches (1) counter (1)

He (1) is close (1) to credit card (1)

***1st request for restaurant availability***

He (1) speaks (1) to the waiter (1) "*How's it going there*"

He (1) stands (1) in front of / at / close (1) the counter / cash register / waiter / credit card (1)

He (1) looks / notices (1) down (1) the credit card (1) *on the counter*

He (1) watches / observes (1) the waiter (1) move (1) the plate *with the credit card on it* / credit card (1) down (1) behind / onto (1) the bar / counter / bench (1)

*He comments* he (1) has friends (1) coming / visiting (1) from overseas / not local (1) and he (1) has heard from some friends (1) that the restaurant is good (1) and they would like to try it out / go there or he (1) wants to take them there (1)

*NB: if it is stated something like "he has friends overseas who are coming" then this is the same as if it had been stated as above i.e., "he (1) has friends (1) coming (1) from overseas (1)*

*NB: "cousins" is incorrect but "people" is correct*

*NB: If it's stated that "he heard the restaurant is good" then he (1 correct) heard (1 correct) restaurant is good (1 correct)*

*NB: stating to have dinner/meal is incorrect since he only says he wants to try it out*

He (1) questions / asks / enquires (1) the waiter (1) about the availability / reservations / bookings (1) of the restaurant (1) for / on / this / upcoming / approaching (1) weekend or Saturday (1) night (1)

*NB: "a" or "the" or "next" Saturday is incorrect*

*NB: there is no gist that he is making or wants to make a booking (i.e., he does NOT make a booking) therefore score incorrect. The word "bookings" is ok if the other words stated with 'bookings' suggest he is enquiring about availability but NOT actually wanting to secure a booking (e.g., "he enquired about bookings" is not making a booking)*

*NB: he does NOT say his friends are coming on Saturday so any mention that this is when the friends are coming is given 1 incorrect point for the gist of when the friends are coming but correct points for the specific words e.g. "he has friends coming from overseas on Saturday and wanted to know the availability of the restaurant" is awarded 1 incorrect for "the friends coming on Saturday" but correct points for: he (1), friends (1), coming (1), overseas (1), availability (1) restaurant (1), on (1), Saturday (1) since the statement also implies he wants to know the availability of the restaurant on Saturday.*

He (1) waits / sees (1) until waiter turns (1)

He (1) reaches / reaching / leans (1) over / across / down / behind (1) the counter (1) to reach (1) with his (1) right (1) hand or arm (1) towards / to grab (1) the card or plate (1)

He (1) gets his (1) hand (1) close / near to the card (1) but he (1) quickly (1) withdraws / moves back (1) onto (1) counter (1) without it / does not grab (1) it / the card (1) *as the waiter turns around*

*NB: any mention of not enough time to grab the card is a supposition (perhaps the guy is too slow!)*

He (1) sees (1) waiter coming back (1)

He (1) attempts / tries to steal (1) the card (1) once / first time (1)

### 2<sup>nd</sup> request for restaurant availability

_2nd request for restaurant availability_

He (1) speaks or talks (1) again (1) and comments that he (1) is not sure if they are coming (1) this / on / approaching (1) Saturday/weekend (1) or next (1) Saturday/weekend (1) and he (1) asks / enquires (1) the waiter (1) for the availability / reservations / bookings (1) of the restaurant (1) for the following / next (1) weekend / Saturday (1) the 22nd (1) he (1) thinks (1)

> _NB: if he is described as just generally asking to check for another time (no specific days or dates) then award 1 correct point for 'asking' and 1 correct for the 'availability'._

_The waiter responds and turns to look at the book and is looking at the book_

He (1) looks or is looking (1) at the card or plate (1)

He (1) reaches / is reaching / leans (1) over / across / down / behind (1) the counter (1) again (1) with his (1) right (1) hand / arm (1) towards (1) the card (1) placing his (1) hand over (1) the plate (1) and grabs or steals or takes (1) the card (1) off / from / behind (1) the plate / counter (1) quickly (1)

He (1) puts / places / hides (1) it or the card (1) in his (1) right (1) back (1) pocket (1) of his jeans / pants (1)

> _NB: saying he has enough time to grab the card is a supposition._

> _NB: in reaching and grabbing the card on this second attempt this takes the same amount of time as the first time he reaches and withdraws without the card so any mention that grabbing the card takes longer or shorter is awarded 1 incorrect point._

> _NB: generally saying (without the detail of each attempt) he goes for the card a couple of times or twice give 1 correct point (plus the points for 'he' and 'card'). Any mention of it being more than twice award 1 incorrect point._

He (1) nods (1) at the waiter (1)

He (1) stands / stays in place (1) in front of (1) the counter / bar / bench / cash register (1) and looks (1) around (1)

He (1) hangs around / waits a while (1) leaning (1) and taps (1) his (1) right (1) fingers (1) on the (1) counter / bar / bench (1)

_Waiter turns back and answers the phone_

He (1) steps back / moves away / walks away (1) from the / to the (1) counter / waiter (1) turns (1) to his (1) left (1) and leans / stands (1) on / by the (1) counter *away from the waiter* (1)

> *NB: The time between the waiter picking up the phone and the thief exiting is 10s. All time estimates in "seconds" or using verbal qualifiers such as "really quickly" or "a little while" give 1 correct point.*

> *NB: a comment such as "he was wandering around" is too vague to score*

> *NB: a comment such as "he moved back and forth" is more specific than 'wandering around' and can be given 1 correct point for 'moving'*

He (1) looks (1) at the (1) waiter (1) *on the phone* and walks (1) to the (1) front / door / entrance (1) without speaking (1)

He (1) doesn't wait (1) for waiter (1) to get back to him (1)

He (1) opens (1) the door (1) and leaves / exits (1) the restaurant / shop / door / front (1) towards RHS / same direction he came in from (1) with the card (1) on foot (1)

Outside (1) he (1) looks (1) around / behind him / left & right / both ways (1) at the (1) front / door / entrance (1) walks (1) few steps / seconds (1) and runs / bolts / sprints (1) off / away (1) towards / from (1) the RHS *of the scene* / restaurant (1) up / onto / down (1) the road / empty parking space (1) and disappears / is gone (1)

> *NB: he does not run across the road if it's stated 'across' then 1 incorrect point. 'Crossing' is correct*

He (1) didn't call (1) anyone (1)

He (1) didn't get into (1) a vehicle (1)

He (1) touches (1) the verandah post (1) with his (1) left (1) hand (1) *on running away*

**NOTE:**

Passport = card for Action scoring

Calendar = diary for Action

Computer ≠ diary for Action

Computer ≠ diary for Action

**PERSON INFORMATION (Waiter)**

Person (1)

> *NB: any title given to the waiter is given 1 correct score for Person or generally stating there is a person*

Name: Mario / Alario (1).

> *NB: if the name is given then also award 1 correct for Person and 1 correct point for gender if these havn't already been identified*

Gender: Male (1)

> *NB: Also award 1 correct for Person (if a title has not been given or the name Mario has not been stated)*

Age: Any mention of guy being in his twenties (1) e.g. any specific age given that is <25yo such as "about 23"; "early twenties"; "twenties"; "young adult"; "young man"; "young gentleman". Same age as others is correct.

> *NB: 'young' is too vague so it is not scored (e.g., as in "young person", "young guy")*

Build: Slim / medium / average / normal / not fat (1)

Height: Within 180 +/-3cm (5 foot 10 inches + 2/- 1 inch) or relative height (taller than perpetrator or customer) (1)

> *NB: 'tall' is too vague to score*

Weight: Within 83 kgs +/-4kg (1)

Complexion: Olive / light complexion (1)

> *NB: 'white' is incorrect*

Facial: Slim face (1) dark eyebrows (1) square jaw (1) square head (1)

Ethnic appearance: Caucasian / European / Mediterranean / Italian (1)

> *NB: Anglo-saxon is not Italian (it's Germanic in origin) therefor incorrect*

Hair: Dark / black (1) straight (1) short (1). Clean shaven (1)

> *NB: shiny hair is not verifiable (supposition)*

Clothing: Dark / black (1) clacky / clicky / loud (1) shoes / dress shoes (1) heeled (1)

Dark / black (1) short-sleeve (1) button-up (1) collared (1) top / shirt (1) restaurant / waiter attire (1)

*NB: 'polo shirt' is incorrect*

Dark / black (1) pants / trousers (1)

*NB: 'belt' is supposition (can't verify)*

*NB: waiter is well groomed for the job (correct)*

Artefacts: Ring (1) on one of the fingers (1) of the right (1) hand (1) and a gold (1) watch (1) on his left (1) arm / wrist (1)

Voice: Australian accent (1) and fluent / perfect (1) English (1)


## ACTION INFORMATION (Waiter)

He (1) stands / waits (1) behind / at (1) the counter / bar / bench / front desk / desk (1) at the (1) cash register (1)

He (1) looks (1) at perp (1)

He (1) serves the / speaks / talks to (1) the customer (1) first (1)

He (1) says (1) to the customer (1) he will bring it (card or bill) out to him (1)

He (1) had an account to put through (1) with the card (1)

He (1) looks (1) at the perp (1)

He (1) left the plate / card (1) on the plate / desk / counter (1)

## 1<sup>st</sup> checking of diary

He *then* (1) serves the or speaks/talks to (1) the perpetrator (1)

*While serving the perpetrator* he (1) picks up / moves (1) the plate *with the credit card* (1) off the (1) counter (1) and puts (1) it / plate / card (1) down (1) behind / onto (1) the counter / bar / bench (1) near the / front of (1) till / phone / waiter (1) away (1) from the perpetrator (1)

*NB: the exception is if it is stated that "he picks up the credit card" then 1 correct score for "he", 1 correct for "picks up" BUT 1 incorrect for "credit card" (he actually picks up the plate and not the credit card, however he does 'move' the credit card so "he moves the credit card" is 3 correct points – 1 for 'he', 1 for 'moves', 1 for 'credit card')*

*NB: he pulls the credit card off the plate is confabulated*

*NB:  a statement like "he leaves the plate there" is unscored*

He (1) says (1) to the perpetrator (1) to hold on (1) he (1) will check (1) the availability / bookings (1) *of the restaurant*

PLEASE NOTE THE FOLLOWING IS A GENERAL STATEMENT:

He (1) checks *i.e., physically checks* (1) the diary (1) twice / two times (1)

*NB: any reference to >2 times in checking the diary or a "few" times then award 1 incorrect point*

**THE FOLLOWING IS THE DETAILS OF CHECKING THE BOOK (participants tend to make a general or detailed statement of checking the book):**

He (1) moves away / steps away / turns (1) away / around / behind (1) him / his (1) back (1) from the cash register / counter / perpetrator / plate / card (1) to the table / diary (1) reaches (1) to look at / check *i.e., physically checks* (1) the diary / date (1) for availability / bookings (1)

*e.g., "he moves away" is scored he (1) moves away (1)*

*e.g., "he steps away" is scored he (1) steps away (1)*

*e.g., "he turns around" is scored he (1) turns (1) around (1)*

*NB" sometimes 'it' = date but it depends on context of descriptions.*

He (1) turns (1) back / around (1) from the book (1) to the counter / cash register / perpetrator (1) too soon / quickly (1) before (1) the perpetrator could take the card (1)

*NB: if it is stated more generally that he "comes back" then award just 1 correct point.*

*NB: it takes him 4s to turn to check the book and turn back.*

He (1) doesn't notice / doesn't see (1) card (1) is missing / gone (1)

He (1) says / replies (1) to the perpetrator (1) that they are available / have availabilities / have bookings (1)

**2ⁿᵈ checking of diary**

He (1) *responds to the second enquiry by* saying / speaking / responding / goes (1) they are busy (1) for the next (1) month (1) but he (1) will check (1) the availability / bookings (1) for the perpetrator (1)

*NB: e.g. He (1) said / goes / like (1) "I'll (1) need to check" (1)*

He (1) moves away / steps away / turns (1) away / around / behind (1) him / his (1) back (1) again (1) from the cash register / counter / perpetrator / plate / card (1) to the table / diary (1) to look at / check *i.e., physically checks* (1) the diary / date (1) for availability / bookings (1)

> *NB: he does not leave the scene so any mention of this is a confabulation (leaving the scene does not happen).*

> *NB: a general statement like "he went back" or "he did" is given 1 correct point for implying he is checking or looking at the diary again*

> *NB: the time it takes to check the diary the second time is longer so score this as correct if it is stated. It takes 4s to check the first time and 15s to check the second time so if any estimate is given in "seconds" or use of a verbal qualifier such as "a bit longer" then award 1 correct point*

He (1) flips / turns (1) pages (1) of the diary (1) *and the phone rings*

He (1) turns (1) back / around (1) to the counter / cash register / perpetrator (1) and says (1) to the perpetrator (1) it is fine / available (1) and he says to hold for one moment / he'll be right with him (1)

He (1) picks up (1) and answers (1) the phone (1)

> *NB: if it is generally stated that he got a call then award 1 correct point for implying he answers the phone*

> *NB: it takes him 8s to answer the phone*

*While on the phone* he (1) looks at / notices (1) the perpetrator (1) leaving (1)

He (1) does not hang up (1) the phone (1)

He (1) leaves / leaves off the hook / puts down / finishes with (1) the phone (1) onto / down (1) the plate *where the credit card was* / counter (1)

He (1) turns (1) around / to the back room (1) to call out / says / speaks / tells (1) Frank / a guy / someone (1) that the phone is for him / the phone is for someone else (1).

He (1) doesn't leave (1) the room (1)

> *NB: waiter does not leave the scene to call out to Frank. If this is said then it is a confabulation (i.e., it does not happen)*

*NB: if it is stated that the phone is handed to Frank then this is a confabulation (i.e., it does not happen)*

He (1) waits (1) for Frank (1) to come to (1) the phone (1)

He (1) went to / goes to process bill / card / payment / wanted to sort bill out / put order through (1)

He (1) turns (1) back / goes back (1) to the counter / bar / bench / cash register (1)

He (1) checks / looks (1) down (1) receipts / bills / papers (1) and takes / lifts / picks *the paper / bill / receipt* (1) off / up (1) the spike / pin / nail (1) and puts it / places it back (1) on the nail / pin / spike (1)

*NB: if it is generally stated that he "does something else" do not score because it is too vague*

*NB: if it is stated - he (1) goes to make payment / wanted to sort the bill out / put the order through (1)*

He (1) picks up / lifts / moves (1) the phone (1) from the (1) plate (1) and puts it (1) onto (1) the counter / bar / bench (1)

He (1) looks (1) down (1) reaches / went for (1) lifts / picks up (1) the plate / credit card (1)

He (1) sees / notices / realizes / recognises (1) the card (1) is missing / gone / stolen (1) by the perpetrator / off the tray (1)

*NB: if it's stated the "waiter realised what happened" this is non-scorable without some indication of what he realised.*

*NB: The time between picking up the plate and realising the card is missing is 3s.*

*NB: The time from putting down the phone and realising the card is missing is 14s.*

*NB: The time between putting down the phone and exiting the restaurant is 16s.*

*NB: the time between the perpetrator leaving the restaurant and the waiter realizing the card is missing is 21s.*

*All time estimates given in "seconds" or using verbal qualifiers such as "really quickly", "a little while" are awarded 1 correct point.*

He (1) moves (1) the plate (1) towards the / from the (1) top counter (1) but he (1) puts or moves (1) it / plate(1) down onto (1) the lower counter / counter / bar / bench (1)

He (1) makes / says (1) a 'mmm' sound / grunt / huffed (1) looks up (1) in direction of door (1) and walks (1) briskly (1) towards (1) the front / door / entrance (1) opens (1) it / the door (1) and exits / went out / leaves / gives chase (1) the restaurant / door / front (1)

> *NB: "goes to door and looks about" is scored as goes to (1) door (1) looks (1 incorrect) about (1 incorrect) because he exits the door*

> *NB: "he looks about" is he (1) looks (1) about (1 confabulated)*

He (1) walks (1) across (1) footpath (1) to the edge / onto / into (1) the street (1) behind (1) the 4WD (1)

He (1) stands (1) and looks (1) up and down / on / around / left and right (1) the street (1) for the perpetrator (1)

He (1) doesn't see / can't see (1) the perpetrator (1)

He (1) turns (1) around (1) and walks (1) back / goes back inside (1) the restaurant (1)

## PERSON INFORMATION (Customer)

Person (1)

> *NB: any title given to the waiter is given 1 correct score for Person or generally stating there is a person*

Gender: Male (1). Also award 1 correct for Person (if a title has not been given or it is generally stated "person")

Age: Any mention of guy being in his twenties (1) e.g. any specific age given that is <25yo such as "about 23"; "early twenties"; "twenties"; "young adult"; "young man"; "young gentleman". Same age as others is correct.

> *NB: 'young' is too vague so it is not scored (e.g., as in "young person", "young guy")*

Build: Slim / average / normal / medium (1)

> *NB: 'strong' build is too vague to score*

Height: Within 177 +/- 3cm (5 foot 9 inches +1/-2 inches) / relative height (taller than perpetrator / shorter than waiter) (1)

Weight: Within 78 kgs +/-4kg (1)

Complexion: Fair / white (1)

Facial: Long / oval face (1) red / reddish (1) beard / goatee (1). No moustache (1). Dark eyebrows (1) sharp nose (1)

Ethnic appearance: Caucasian / Australian (1)

Hair: Balding / shaved / 'follically challenged' (1)

*NB: hair / hair + colour of hair is 1 confabulated. If it's stated 'long blonde hair' then x2 confabulated (1 for colour, 1 for length)*

Clothing: Dark / blue (1) pants / trousers (1).

Light / white / cream (1) cotton (1) long-sleeved (1) open / unzipped / unbuttoned (1) long (1) top / shirt / jacket (1)

Dark / blue (1) buttoned (1) top / shirt (1) with a white collar (1) underneath (1) *the long-sleeved shirt/jacket* and not tucked in (1)

Casual (1)

*NB: well-groomed is supposition (cf. waiter is well groomed for the job)*

Artefacts: None observed (1)

*NB: 'watch' is supposition (can't verify)*

Voice: Australian accent (1) and fluent (1) English (1)

**ACTION INFORMATION (Customer)**

He (1) walks (1) in / into (1) the foyer / reception / bar area (1) entering from the LHS / another part of the restaurant (1).

*NB: if it is just generally stated that he "comes in" then award 1 correct point.*

He (1) approaches / goes to (1) the waiter / counter / cash register (1)

He (1) stands (1) at the / by the (1) counter / bar / bench / cash register (1) opposite / in front of (1) the waiter (1)

He (1) mumbles / little bit inaudible (1) speaks to / asks the / makes enquiries (1) the waiter (1) that he (1) wants to pay his account / meal / drinks (1) with his (1) credit card (1) waving / flashing / presents (1)

*NB: he is not making a booking or enquiring about bookings or ordering drinks or food. These are all incorrect.*

*NB: he does not say "take care of this", this is incorrect (e.g. "here is my credit card please take care of it for payment later" = my (1 correct) credit card (1 correct) take care of it (1 incorrect) for payment (1 correct)*

*NB: he said "he'd pick it up later" = He'd pick…up later (1 incorrect) it (1 correct)*

*NB: "put it on my tab" = 1 correct (it's not clear that 'it' = credit card so don't score separately)*

He (1) relinquishes / leaves / hands over (1) his (1) card (1) by putting / placing (1) it / card (1) down (1) on the / at the serviette / plate / counter / bench / bar (1) with his (1) right (1) hand (1) down (1)

*NB: if the participant drills down and states something like "he put it onto the napkin on a plate on the counter" then scoring for the part "on a plate on the counter" gets captured under Surroundings only*

*NB: "he leaves the card with the waiter" is all correct and is scored: he (1) leaves (1) the card (1) with the waiter (1). If it is stated "he hands over his credit card to the waiter" then the scoring is: he (1) hands over (1) his (1) credit card (1) to the (1 incorrect for action of actually giving it to the waiter) waiter (1)*

He (1) walks (1) off / away / out / leaves (1) the scene / foyer / server / counter (1) for rest of video (1) goes into restaurant (1)

*NB: e.g. he (1) goes away (1)*

*NB: When he leaves the scene it is unverifiable what he then does or where he goes so any comment about this is a supposition and is not scored*

He (1) does not (1) return / come back (1)

## PERSON INFORMATION (Frank & General)

Frank

Person (1)

Gender: male (1).

*NB: Also award 1 correct for Person if not already stated by more another way (e.g., "person out back")*

Name: Frank (1).

*NB: if the name is given then also award 1 correct for Person and 1 correct point for gender if these havn't already been identified*

Frank (1) is not visible / off camera (1)

He / Frank (1) is in the room / restaurant (1) out back (1)

Italian (1)

<u>General</u>

Three (1) people (1) in view or observable or in the scene inside the restaurant (1)

## ACTION INFORMATION (Frank)

Frank or person or male (1) speaks / voice / responds (1) and says "alright" (1) in the background (1)

## SURROUNDING INFORMATION (Outside)

View from outside (1)

Australia / South Australia (1)

City or suburban Adelaide or urban / western Adelaide / Glenelg (1)

Two lane (1) side (1) streetscape / road / street (1) quiet (1)

*NB: main (1 incorrect) road (1); single lane (1 incorrect) road (1)*

Straight (1) *road*

Footpath (1)

Residential / buildings / houses (1) either side (1) of street (1)

Picket fences (1)

Little or not much (1) pedestrian traffic (1)

*NB: no people apart from thief can be seen crossing the road – if mentioned then confabulation*

Little / no (1) vehicular traffic (1)

Day light (1) overcast (1) not raining (1)

*NB: mentioning a time of day such as lunchtime or midday is a supposition because it cannot be verified from the video.*

Two (1) people (1) walking (1) on other side (1) of the road / in background (1)

A person (1) with a camera (1) tripod (1) on the LHS (1) of the verandah (1) in the corner (1)

*NB: any mention of the perp behind a car is a confabulation – no one is there.*

Man's (1) voice (1) in the background (1)

*NB: cannot hear cars (incorrect)*

Wooden (1) post (1) on LHS (1) of scene/door (1)

Stobey or electricity pole (1)

Little / small (1) boxy (1) building / shop / restaurant (1) on LHS (1)

*NB: describing the place as a 'pub' or 'bar' is incorrect*

Italian (1) *restaurant*

Name is 'Capriccio' or very similar such as 'Copiccios' or it starts with a 'C' or it has a double 'c' in the middle (1)

White (1) red (1) writing (1) sign (1) above it / out the front (1)

Brown (1) and cream (1) brick (1) front of facade (1) with two (1) windows (1) glass (1) door / entrance (1)

No balcony (1)

Red (1) and cream (1) verandah (1)

No tables (1) out front (1)

A few / some / several / seven (1) cars (1) parked (1) on the street (1)

Three (1) cars (1) parked (1) out front (1) opposite (1)

Car (1) parked (1) in driveway (1) opposite (1) the restaurant (1)

Empty parking space (1) behind (1) 4WD (1)

One (1) large / bulky (1) dark / red / maroon (1) vehicle / car / 4WD / SUV / Toyota Landcruiser (1) parked (1) outside / directly / closest to (1) the restaurant / door / line (1) with black (1) roof racks (1) on LHS (1) of street/road (1)

View of the rear / back (1) orange (1) indicators (1)

Silver (1) black (1) bumper (1)

*NB: mention of a spare tyre on the back is x2 confab – there isn't one on the back*

South Australian (1) white (1) number plate (1) with 3 (1) numbers (1) and 3 (1) letters (1). Alpha (1) numeric (1)

VGP 291 (1 correct for each letter/number identified correctly and 1 correct for correct location of letter/number e.g. "first letter was 'V'" is 2 correct points)

*NB: saying the number plate has 6 numbers is incorrect – it has 3*

*NB: blue thing at top is confabulated*

Old / older / oldish model (1)

One (1) oldish (1) dark / black / dark blue (1) vehicle / car / station wagon / Holden Commodore (1) parked (1) opposite / other side of road (1) to the restaurant (1)

One (1) white (1) vehicle / car / station wagon (1) parked (1) to the RHS (1) of the Commodore (1) opposite or other side of road (1) to the restaurant (1)

One (1) white (1) vehicle / van (1) parked (1) same side (1) further along / LHS (1) street *from the 4WD*

## SURROUNDING INFORMATION (Inside)

Doorbell (1)

View from behind (1) the counter / waiter (1)

Front of shop / room / foyer / waiting area / bar area / serving area (1) tiled (1) floor (1) walls (1) roof (1)

Granite (1) brown (1) reception bench / counter / bar / front desk / desk (1) L-shaped / perpendicular (1) higher (1) and lower (1) levels or counters (1) in front / opposite / slightly to left (1) door / front of shop (1) open to one side (1) in front / back (1) of wall (1)

*NB: if it is mentioned that the counter has two levels award 1 correct point for implying 'higher' level and 1 correct point for implying 'lower' level)*

*NB: saying the counter is wooden is incorrect*

Bar / alcohol / bottles (1) to RHS / next to (1) of waiter / counter (1) (or LHS (1) of entrance / door (1)) on the (1) back wall / side of bar (1)

Mirrors (1) on the / RHS (1) wall (1)

Rows (1) of wine (1) glasses (1) hanging (1) upside down / inverted (1) over bar (1) to RHS (1) of waiter (1)

White (1) cloth (1) napkin / serviette (1) on (1) a small / little (1) white (1) round (1) plate (1) on top of / at the (1) counter / bar / bench (1) in front of (1) the waiter / customer (1) folded (1) into square (1)

*NB: "napkins" is given 1 correct score for identifying the object as a napkin and 1 incorrect score for plurality of the napkins*

*Plate / Card is also located* near / next / RHS / front of (1) cash register / waiter / phone (1) on RHS (1) on lower / underneath / down behind (1) counter (1)

*Card also located* in his (1) pocket / jean pocket (1)

*NB: if it's stated that the card is in the jacket pocket then this is incorrect*

One (1) light / gold / brown / orange / amber (1) plastic (1) regular / standard size / rectangular / square (1) card / bank card / credit card / American Express / AMEX card (1) with dark / black (1) symbol / face / logo (1) in middle *of card* (1) and black (1) stripes / border (1) magnetic strip (1) numbers / lettering (1) no photo ID (1) face-up (1) on a (1) serviette / plate / counter / bar / bench (1) in front (1) of perpetrator / waiter (1)

*NB: if participant drills down and says something like "the card was on the serviette on a plate on the counter" then scoring for the part "on a plate on the bench" gets captured above for the serviette*

*NB: the image in middle of card in not of the World so this is incorrect if stated*

*NB: Debit card is incorrect.*

Close-up (1) of credit card (1)

Diary / book / bookings book / reservations book / schedule (1) behind (1) the waiter / counter (1) on a (1) table / desk (1)

*NB: "calendar" is incorrect as it implies a month to a page which the book isn't. "booking sheet" and "log book" are incorrect*

Pages (1) *in the diary*

Table / desk (1) behind (1) counter / waiter (1) *that has diary on it*

White / cream (1) cash register / till (1) on LHS / at the (1) of waiter (1) on the / behind the (1) counter / bar / bench (1) with white (1) paper (1) on it (1) next to / by the (1) telephone / waiter (1) or in front of (1) wall (1)

White (1) telephone (1) rings / call (1) loud (1) rings about 3-4 times (1) on the / behind (1) counter / bar / bench (1) to the LHS (1) of the waiter (1) next to / by the (1) cash register / waiter (1)

Papers (1) on the / behind (1) counter / bar / bench (1) next to / by the (1) cash register / waiter (1)

Payment machine (1) on the (1) counter / bar / bench / register (1) next to / by the (1) cash register (1) in front (1) of the waiter (1)

Blue pen (1) next to (1) plate (1)

*NB: it is a single pen, therefore plural is incorrect but identifying the object is correct*

Smokes (1) next to (1) plate (1)

Spike / pin / nail (1) on the (1) counter / bar / bench (1) with receipt / paper / bills (1) on it (1) next to / by the / behind (1) cash register / bar (1) in front (1) of the waiter (1)

European / Italian (1) music playing (1) *in the background*

General colour (1) of bar / reception / foyer area / waiting area (1) is red / reddish / maroon (1)

Cannot hear (1) other people *except Frank* (1) talking (1) in the restaurant (1)

Room (1) out back (1)

Dimly lit (1)

## APPENDIX D

## Picture Stimulus Quantity Scoring Rules, Coding Key and Examples

### GENERAL RULES FOR SCORING

1. Information is only categorised into Surrounding and Person since it is a still shot.
2. Only score the first time a detail is mentioned.
3. If the participant corrects themselves, then only score the detail used to correct their statement.
4. If a detail is referred to by more than one description that denotes a different grainsize, then score only the finer-grained response for correctness, e.g.,
    a. Europe is coarser than Bulgaria which is coarser than Plovdiv.
    b. Vehicle / cars is coarser than hatchback which is coarser than Volkswagon / 4WD.
5. If any qualifiers are used to express confidence (e.g., "probably", "not sure") ignore the qualification.
6. Confabulations are things that have been added by the participant (e.g., "there was a moped on the left side of the street") and do not exist on the picture.
7. Suppositions are things that cannot be verified (e.g., "Winter" or "Autumn"; "it's a quiet street")
8. Unscorable if cannot assign to any category of scoring.

### PERSON INFORMATION

Single (1) person (1) woman (+1)

On LHS (1) of street / picture (1)

Standing (1) on the (1) footpath / curb / side of street (1)

Outside (1) in front (1) of door / house / store / building (1) in foreground (1)

Middle aged / older (1)

*NB: "old" is incorrect*

White complexion (1)

Brown hair (1)

Dark / black (1) clothes (1)

Dark / blue / dark blue / black (1) and grey (1) coat / cardigan (1)

Dark / black (1) pants (1)

Dark / black (1) shoes (1)

Holding / has / had (1) a broom (1)

*NB: any item mentioned that is held by woman is correct but if it's not a broom it gets scored as incorrect in the Surrounding*

Sweeping / cleaning (1) the footpath / pavement / side of street (1)

Bent over / hunched / stooped (1)


## SURROUNDING INFORMATION

Outside Australia / European / Bulgaria / Plovdiv (1)

*NB: stating any country of Europe that is not "Bulgaria", and any city that isn't "Plovdiv" is incorrect*

Old / 18th-19th century (1) town (1)

*NB: "village" is incorrect*

Residential / urban / precinct / suburban (1)

Street / street scene / alleyway / road / pathway (1)

Junction (1) with 3 (1) streets (1)

Y- or T- intersection (1) at far end (1)

Street (1) is wider (1) at bottom / front (1) of picture (1)

*NB: "Courtyard" / "little square" is not verifiable so this is a supposition*

Street (1) is narrower (1) in middle of / going away in (1) picture (1)

Paved / cobbled (1)

Rocky (1) uneven (1) surface (1)

Irregular (1) large (1) light / pale (1) grey (1) shale / slate (1) stone (1) pavers (1)

Footpaths / sidewalks / pavement / curb (1) against (1) buildings (1) either side / both sides (1) of street / road / picture (1)

No front yards (1)

Power lines (1) across (1) street / road (1)

4 / a few (1) tall (1) trees (1) on both sides (1) of street / road (1) along / on (1) footpaths (1) branches (1) hanging over (1) street (1)

Two (1) trees (1) on LHS (1) of street / road (1)

Two (1) trees (1) on RHS (1) of street / road (1)

One (1) tree (1) on RHS (1) of 4WD (1)

Deciduous (1) trees (1) few leaves (1)

   *NB: "no leaves" is incorrect*

Daylight / daytime (1)

Overcast / no shadows / gloomy (1)

Dry / no rain / not snowing (1)

Clearway (1) sign (1) at end (1) of street (1).

   *NB: it is not a stop sign*

Row (1) houses (1)

3 (1) blocks (1) buildings / apartments / houses (1) on both / either side (1) of street / road / picture (1)

2 (1) blocks (1) buildings / apartments / houses (1) in foreground (1)

1 (1) block (1) buildings / apartments / houses (1) in background / at the back / straight on (1)

8 (1) buildings / houses / apartments (1) *in total*

4 (1) buildings / houses / apartments (1) on LHS (1) of street / road / picture (1)

3 (1) buildings / houses / apartments (1) on RHS (1) of street / road / picture (1)

1 (1) building / house / apartment (1) at far end / background (1) of street / road / picture (1)

Multi-storey / 2- / 3-storey / tall (1) buildings (1)

   *NB: stating buildings are >3 storey is incorrect*

House / apartments / building (1) at far end of street / in background / straight on (1) is 3-storey (1) and upper storeys (1) overhang (1) ground storey / level (1)

Colourful (1) houses / buildings / apartments (1)

<u>Additional points</u> if mention any of the following specific colours and/or location of the buildings/houses/apartments and their colours according to:

- One / 1st (1) building (1) on RHS (1) is orange / beige / mustard / gold / yellow / peach (1) with white (1) trim (1) around (1) windows (1)
- One / 2nd (1) building (1) on RHS (1) is pink or red or marone (1) with white (1) trim (1) around (1) windows or door (1)
- One / 1st (1) building (1) on LHS (1) is pale / olive / cream (1) with brown (1) trim (1) around (1) windows (1)
- One / 2nd (1) building (1) on LHS (1) is pink / red (1) with brown (1) and white / cream (1) trim (1) around (1) windows (1)
- One / 3rd (1) building (1) on LHS (1) is grey (1) with white (1) trim (1) around (1) windows (1)
- Building (1) at far end / in background / straight on (1) is light / white / cream (1) with blue (1) trim (1) around (1) windows (1)

Rendered (1) walls (1)

*NB: mentioning there are bricks is incorrect as none can be seen as part of the buildings*

Tile (1) rooves (1)

Bars / mesh (1) curtains (1) shutters (1) windows (1)

Doors / doorways (1) on either side / both sides (1) of the street / road / picture (1)

*NB: if it is stated "doors down one side and down the other side" then consider this to mean doors down both sides*

White (1) sign (1) on wall / building (1) on LHS (1) of door / building / picture (1) near (1) woman / person (1)

Yellow (1) sign (1) on wall / building (1) above (1) vehicles / vehicle / cars / car (1) on RHS (1) of street / road / picture (1)

Garbage can / rubbish bin (1) up / along (1) the street / footpath (1) just past (1) the woman / person (1) on LHS (1) of street / road / picture (1)

Red (1) broom (1) *held by woman*

Two (1) vehicles / cars (1) parked (1) in the street / road (1) in foreground (1) on the RHS / near centre (1) of street / road / picture (1) opposite (1) woman / person (1) in front (1) of ochre / orange / yellow / gold (1) building / house / apartment (1) with back to camera / facing forward (1) on the curb / side of road (1)

> *NB: "a few" / "some" is incorrect*

> *NB: if either of the two cars is specifically located in the scene then score with respect to parked (1) in the street/road (1) in foreground (1) on the RHS / near centre (1) of street / road / picture (1) opposite (1) woman / person (1) in front (1) of ochre / orange / yellow / gold (1) building / house / apartment (1) with back to camera / facing forward (1) on the curb / side of road (1)*

Parked (1) next to each other / side by side (1)

One (1) small (1) red (1) vehicle / car / hatchback / Volkswagon (1) parked (1) on LHS / next to (1) of the 4WD (1) European make (1)

> Number plate (1) is PB7330XH (any mention of a correct letter / number award (1) point for each correct)

One (1) dusty / dirty (1) dark / blue / dark blue (1) vehicle / car / 4WD / Range Rover (1) parked (1) on the RHS / next to (1) of car (1)

> Number plate (1) is yellow (1) is B6723CM (any mention of a correct letter / number award (1) point for each correct)

> The 4WD (1) has a boxy (1) shape (1)

Recognising plates are not Australian (1) **if havn't already mentioned outside of Australia or Europe**

# APPENDIX E

**Film Stimulus Grainsize Scoring Rules, Coding Key and Examples**

## GENERAL RULES FOR SCORING

1. Arbitrary demarcation of grainsize chunks into fine- or coarse-grained examples but essentially the more detail given for a chunk then it will be deemed a fine-grained response. Paraphrasing or attempting to quote dialogue will be scored as fine-grained.
2. The information is not scored according to whether it is correct or erroneous.
3. If a grainsize chunk is referred to more than once, then only score the finer-grained example (if it is finer-grained).

## PERSON GRAINSIZE CHUNKS AND EXAMPLES

| Person Chunk | Coarse-grained examples / rules | Fine-grained examples / rules |
|---|---|---|
| 1. Thief head and facial features | x1 hair attribute (e.g., dark/light, length, style) Or one facial feature | Specific hair colour (e.g. black) +/- 1 or more attributes Or dark hair + one or more attributes Or x1 hair attribute (e.g., colour – dark or specific – length, style) + any facial feature (e.g. brown eyes, large nose) Or two or more hair attributes – not colour (e.g., length, style) Or two or more facial features |
| 2. Thief clothing – upper body | Top, jacket, hoodie, jumper, T-shirt, Nike jacket +/- one attribute (e.g. colour, material, feature) Or casual clothing Or beanie Or top / jacket + 1 artefact (no attributes, e.g., necklace) | Top / jacket / etc. + two or more attributes (e.g., colour, material, feature) +/- mentioning T-shirt underneath, artifacts (e.g., necklace, sunnies, glasses) Or jacket + beanie Or jacket + T-shirt underneath |

| | Or one artifact | Or two items of clothing +/- one or more attributes<br><br>Or jacket + 1 attribute + 1 artifact |
|---|---|---|
| 3. Thief clothing – lower body | Pants, tracksuit pants, jeans<br><br>Or Shoes, sneakers, boots | Pants / tracksuit pants / jeans + one or more attributes (e.g. colour, material, length)<br><br>+/- footwear<br><br>Or footwear + one or more attributes (e.g., colour, shape)<br><br>Or pants / jeans / jumper/etc. + footwear |
| 4. Thief ethnicity | Region (e.g. Asian, Middle Eastern)<br><br>Or Caucasian | Any specific country (e.g. Indian) |
| 5. Thief skin colour | Dark skin | Tanned, olive skin, white |
| 6. Thief build and height | 'average' build<br><br>Or relative height (e.g., shorter than waiter) | Skinny, not fat, slim, not too thin not too fat<br><br>And / or specific height (e.g., five foot eight, 173cm)<br><br>'Average' build + relative height |
| 7. Thief age | Young man, teenager, twenties<br><br>Or relative age (e.g. younger than waiter)<br><br>Or age range >=10years (numerical or qualitative e.g., late twenties to early thirties) | Any exact age (e.g. 20), specified age range <= 5years (e.g., 18 – 23), early / mid / late twenties, early / mid / late teens, mid to late teens. |
| 8. Customer head and facial features | No hair, bald, balding<br><br>Or not much hair (no attributes)<br><br>Or one facial feature | Any mention of hair + one or more attributes (e.g., colour, length, patches of hair, receding hairline)<br><br>+/- facial features (e.g., long nose, beard). |

| | | Or bald + one or more facial features |
|---|---|---|
| 9. Customer clothing – whole body | Top, shirt, jacket<br>Or pants, trousers<br>Or 'dark clothes'<br>Or casual | Top / shirt / jacket + one or more attributes (e.g. colour, feature, material)<br>Or pants / trousers + one or more attributes<br>Or top / shirt + pants/trousers +/- artefacts (e.g., ring, toothpick, glasses, no glasses)<br>Or one item clothing + one artifact |
| 10. Customer ethnicity | Region (e.g. European)<br>Or Caucasian<br>Or Anglo Saxon | Any specific country (e.g. Australia) |
| 11. Customer skin colour | Pale | White, tanned |
| 12. Customer build and height | 'average' build or<br>Relative height (e.g., shorter than waiter) | Slim, broad shoulders<br>And / or specific height (e.g., five foot eight, 173cm)<br>'Average' build + relative height |
| 13. Customer age | Young man, twenties<br>Relative age (e.g. older than waiter) | Any exact age (e.g. 20), specified age range <= 5years (e.g., 20 – 25), early / mid / late twenties |
| 14. Waiter head and facial features | x1 hair attribute (e.g., colour, length, style)<br>Or one facial feature | Specific hair colour (e.g. black) +/- 1 or more attributes<br>Or dark hair + one or more attributes<br>Or x1 hair attribute (e.g., colour – dark or specific – length, style) + any facial feature (e.g. brown eyes, large nose)<br>Or two or more hair attributes – not colour (e.g., length, style) |

| | | Or two or more facial features |
|---|---|---|
| 15. Waiter clothing – whole body | Dressed all in black<br>Or top / shirt +/- 1 attribute (e.g. colour, feature)<br>Or pants +/- 1 attribute (e.g. colour, feature)<br>Or 'dark clothes'<br>Or shoes +/- 1 attribute (e.g., colour, sound)<br>Or waiter attire | Top / shirt + 2 or more attributes<br>Or pants + 2 or more attributes<br>Or any 2 items of clothing or more +/- 1 or more attributes (e.g. colour, feature)<br>Or any of above +/- shoes (+/- attribute (e.g. colour, sound))<br>Or shoes + 1 attribute + dressed all in black<br>Or shoes + 2 attributes (features, sound) |
| 16. Waiter ethnicity | Region (e.g. European)<br>Or Caucasian<br>Or Anglo Saxon | Any specific country (e.g. Italian) |
| 17. Waiter skin colour | Dark, pale | Olive, tanned, white |
| 18. Waiter build height and height | 'average' build<br>Or relative height (e.g., taller than thief) | Slim, broad shoulders<br>And / or specific height (e.g., five foot eight, 173cm)<br>'Average' build + relative height |
| 19. Waiter age | Young man, twenties<br>Relative age (e.g. older than thief) | Any exact age (e.g. 20), specified age range <= 5years (e.g., 20 – 25), early/mid/late twenties |
| 20. Frank | A person | Frank +/- 'out the back'<br>Or a person + 'out the back'<br>Or person not visible |
| 21. General | No other people | Three people in view |

**ACTION GRAINSIZE CHUNKS AND EXAMPLES**

| Action Chunk | Coarse-grained examples | Fine-grained examples / key words |
|---|---|---|
| 1. Perp approaches restaurant | Referring simply to approaching restaurant or front door without any other descriptors.<br>A fellow came into the scene.<br>He crossed the road.<br>(x2 elements) | He walked across the street.<br>Man sort of run across the road.<br>He parked his car outside.<br>Walking down the street.<br>Sort of semi-jogged in three-quarter pace.<br>Looked around street / café. |
| 2. Perp enters restaurant | Came into / went into restaurant.<br>Entered restaurant.<br>Came in / going inside.<br>Young man enters restaurant. | Walked in.<br>Walked into restaurant.<br>Walks into a bar.<br>Skipped inside the restaurant. |
| 3. Customer enters reception | Gentleman went in.<br>He comes in. | He walks in. |
| 4. Customer is served | The person at the counter was talking to another man at first.<br>He was paying.<br>To pay for something.<br>The guy (waiter) serving.<br>He was standing at the counter.<br>To order something.<br>A tab.<br>Make a reservation.<br>Mumbles.<br>Came up to pay.<br>Approached counter to make payment.<br>Came up to counter and mumbled something. | He wanted a drink or something and ordered a drink.<br>He was paying his bill.<br>To pay the bill.<br>Came up / walks up to pay for his meal.<br>Customer wanted to pay the bill.<br>There is a man standing at the counter before him (perp) who has gone to pay for his meal.<br>Man in there who had his credit card waiting to pay for some food or something or make a booking.<br>He wants to pay and waiter says he'll bring it out to him. |

| | | Approached counter and was speaking to guy to make payment. Talked to person to make payment. Asked for something to be paid with his card. |
|---|---|---|
| 5. Perp waits | Younger man standing behind him (customer). A younger boy behind him. There was a person in front of him. He stood beside the counter. He was looking around a lot. He waited. | He was waiting at the counter. He was in the background for a little while. He was holding back a little bit until the other customer left. Sees customer give his card because he's cued behind him. (generally need some indication of time) |
| 6. Customer leaves card | Give him his credit card. Gave his credit card. He put his card down. | He gave his card by popping it on a tray. The waiter asked him to put it on a plate. The first guy put his credit card on the plate which was on the table in front of the cashier. |
| 7. Customer leaves | He went away. He left. | Customer walked off. He left to go back to his table. He went away from the picture. He leaves the shop. |
| 8. Perp is served (asks about restaurant availability) | Wondering if he could check the availability. The boy came up and enquired about dinner reservations / availability on the Saturday night. Not sure whether the young guy made a, or wanted to make a booking. | He said his cousins were gonna come over and they wanted to eat at the restaurant. I've got friend coming from overseas, talked to his friends, thought it was a good restaurant, could he enquire about a booking. He went up to the counter and sees the card and enquires about |

| | He walked up and he asked the fellow about booking dates, he told him to check for Saturday. Noticed the credit card.<br>He came up and asked for a booking.<br>(approx. <= 5 elements: asks, booking / availability, Saturday, night, friends, sees card, etc.) | booking for the weekend / Saturday night. |
|---|---|---|
| 9. Waiter moves card | He moved the card.<br>He moved the card away from the perp | He put it behind the counter.<br>He moved the card / plate down behind the counter away from the perp.<br>He took the plate put it down to the counter and there's a till there and there's a sort of it raises up a bit. |
| 10. Waiter checks availability | He checked.<br>He checked availability.<br>He checked the books.<br>He turned away / around.<br>He turned to check.<br>His back is turned.<br>He turned around and checked / looked.<br>He said he'll check.<br>He wasn't looking.<br>(approx. <= 3 elements: turned, around, looks / checks, date / diary, availability, etc.) | He turned around to check the bookings books / diary / calendar / date.<br>He said "I'll check for you" and turned around and checked availability.<br>Turned his back towards him to check availability. |
| 11. Perp attempts to steal card | He tried to take the card.<br>He tried the first time.<br>He tries to grab the credit card.<br>He tries but isn't quick enough. | He leaned over the counter and tried to take the card but wasn't quick enough. |

| | He tries to take the card but doesn't. | The guy leans over the counter almost takes the card but doesn't. The younger gentleman reached over to try and grab the gold card but just missed grabbing it and moved back. Tries to grab the card off the table / plate. |
|---|---|---|
| 12. Waiter turns back | Waiter turns back. Waiter says it's fine. Came back too soon. | Waiter turns / comes back tells him Saturday night's fine Waiter turns his face and body back to the perp. Turned around before he could steal the card. He turned back too quickly. |
| 13. Perp requests second time for availability | He asked the waiter for the availability of another date. He kept talking a little bit. Asked about next weekend. Asked about Saturday afterwards. He said it might not be this weekend, it might be next. (approx. <= 3 elements: asks, looking / checks, booking / availability, weekend / Saturday, night, not sure when friends coming, etc.) | He tried by asking him to check the availability for the following Saturday / weekend. He said oh can you check the following weekend, Saturday I think it was the 22$^{nd}$. He says I'm not sure when my friends are coming so can you check the next Saturday +/- 22$^{nd}$. He said it might not be this weekend, it might be next so he asked him to check again / next weekend. Not sure when my friends are coming, could be this Saturday or next Saturday, the 22$^{nd}$. |
| 14. Waiter checks again | Waiter checks again. He turned around +/- again. | So he checked again, turned his back towards him. |

| | (approx. <= 2 elements: turned, around, looks/checks, date/diary, availability, etc.) | He turned around and checked the book again.<br>He turned around to look at the date.<br>The waiter I think we're pretty busy this month but I'll check for you and he checks.<br>He turned around for longer.<br>He went to look for another time / date. |
|---|---|---|
| 15. Perp steals card | He steals the card.<br>He takes the card.<br>He grabs the card.<br>He managed to garb the credit card. | He leans over / reaches for the card and takes it and put it in his pocket.<br>He was a lot quicker and a lot smoother and picked it up and put it in his back jean pocket.<br>He reached across the counter and grabbed the credit card.<br>He quickly grabbed the card.<br>He takes the card quickly.<br>He takes the card in his hand. |
| 16. Waiter turns back | Waiter turns back.<br>Says it's fine.<br>Told him to wait a moment. | Waiter turns back and says that's fine, one moment please.<br>Turns back and didn't notice card was missing. |
| 17. Waiter answers phone | He answered / talking the phone.<br>While he was on the phone.<br>Bartender took the call.<br>Picked up the phone.<br>He sees guy leave. | Waiter picks up and answers the phone and has a conversation on the phone.<br>He spoke on the phone for a minute.<br>He picked up the phone and started talking to the person on the phone.<br>Picked up the phone and didn't notice card was missing. |

| | | He answers phone and sees guy leaving. |
|---|---|---|
| 18. Perp waits | Perp waits. He stood there. He backed / walked away. Hung around a bit. | He stands there for a bit. Perp stepped back from the counter. He stood back a bit. He walked around/away for a bit. He looked around (+ some element of time) Waits for a few seconds. (any time estimates = fine-grained) |
| 19. Perp exits restaurant | He left / leaves. He escaped He left the door / shop. Went out with card. He slipped out the door. | He walked slowly to the door and then once outside….. He walked slowly out the door. He walked / walks out. He left the store quickly. He leaves the shop with the card. He runs out with the card. |
| 20. Perp runs away | He ran away. He ran away with the card. < 3 elements (unless specified which direction he ran) | Once he was outside he ran off to the right. Outside the door he made a run for it. He looked around and ran off. He ran across the street. Outside / as soon as he was out in the open he ran away. He ran away from the restaurant. He ran to the right. He ran and he didn't leave in a car. |
| 21. Waiter puts phone down | The waiter put the phone down. He hangs the phone up. He finished the phone call. | The waiter put the phone down on the plate / desk / counter. The waiter put the phone down on the plate where the credit card was. |

| 22. Waiter calls to another person | To call his manager / boss / colleague.<br>The phone was for someone else / Frank. | Waiter says "Frank the phones for you"<br>Called his manager to come and receive the phone call.<br>Employee called another employee to the phone.<br>Told his boss that there was someone on the phone for him |
|---|---|---|
| 23. Waiter checks docket | He checked the receipt / bill.<br>He picked up receipt.<br>He was going to process the order. | He lifted the receipt off the pin, checked it and put it back on it. And then picked up a cash register receipt and was going to put it on, you know stack it on a pin. |
| 24. Waiter realises card is gone | He realised the card had gone.<br>Realised the credit card was missing.<br>Realised kid had taken the card. | As he went to grab the phone he realised that the credit card had been taken.<br>He recognised that the credit card was not on the plate.<br>He goes to look at the credit card, picks up a napkin, or the tray with the napkin on it, looks at it and goes oh no.<br>He realised card was gone after a few seconds / minutes.<br>He recognised card had gone and kid had taken it. |
| 25. Waiter exits restaurant | He ran / walked outside.<br>He went outside.<br>He went out the shop.<br>He went out after him. | He ran out to the front.<br>He ran out the door.<br>He ran out the store and ran onto the street.<br>He ran out after him.<br>He went out to the road.<br>Went out the door quickly and out onto the road. |

| | | He ran out quickly. |
|---|---|---|
| 26. Waiter looks for perp but can't see him | To see if he could find the guy. To find who stole the card but he can't really find him. He had gone. To look for him but can't see him/he was gone. He looked but didn't see him. Looked left and right. Looked up and down the street. To look for the guy. Found no one on the street. | Had a quick look, didn't go across the road, he just stood in the footpath, couldn't see him. He looked right and left and sort of shook his head. He looked around outside but couldn't find him. |
| 27. Waiter goes back inside | He went back inside. | And walked back in. He went back into the restaurant. Turned around and went back in. |

**SURROUNDING GRAINSIZE CHUNKS AND EXAMPLES**

| Surrounding Chunk | Coarse-grained examples / rules | Fine-grained examples / rules |
|---|---|---|
| 1. Street scene | One of the following: Street, road, street scene, footpath, climate conditions, daytime, locality (Glenelg, country town, etc.), other buildings, people/sounds in background, etc. | Two or more of the following: Street, road, street scene, footpath, climate conditions, daytime, locality (Glenelg, country town, etc.), other buildings, people/sounds in background, etc. |
| 2. Cars | 'There were cars' Or 'few cars' Or car +/- one attribute (e.g. colour) +/- 'parked (in the street) Or '4WD(s)' (no attributes) | Two or more attributes (e.g. colour; features; type / make / model, i.e., sedan, station wagon, 4WD; location, i.e., parked out the front) |
| 3. Restaurant (outside) | Building, Shop, Restaurant, Café Pub | Building / shop / restaurant / café / pub + one or more additional descriptions (e.g., name incl. 'it started with a C', type, features, colour) |
| 4. Reception area | Reception area, foyer, serving area Or door / entrance Or one of the following: Room small wall ceiling doorbell music in the background dimly lit general colour (e.g., red, maroon) | Reception area, foyer Or door / entrance + one or more of the following: Room small wall ceiling doorbell music in the background dimly lit general colour glasses over bar wine bottles on wall mirrors |

| | | |
|---|---|---|
| | glasses over bar<br>wine bottles on wall<br>mirrors<br>etc. | etc.<br>If reception / foyer / door is not mentioned then x2 of the above<br>Or reception area / foyer / serving area + door |
| 5. Reception desk (incl. desk behind waiter) | Counter, desk, bar, table | Counter / desk / bar + one or more attributes (e.g., material, reference to levels, opening, desk behind waiter, view from behind waiter, L-shaped, bar to the right, opposite door, etc.) |
| 6. Objects on desk (incl. diary on desk behind waiter) | Card / credit card / AMEX on the napkin/plate/counter<br>+/- phone call<br>+/- diary/calendar<br>Or phone call<br>Or diary<br>Or phone call + diary | Card / credit card/AMEX on the plate / counter + additional location (e.g., lower bench, behind counter, near waiter)<br>Or two or more additional objects or additional attributes of objects (e.g. receipts, receipt pin, pages in the diary, location of diary on the desk, cash register, white telephone, barrel / keg, plate, etc.)<br>Or card / credit card/AMEX on the napkin/plate/counter (+/- phone call +/- diary/calendar) + one or more additional objects / attributes of objects (as above) |
| 7. Credit card | Card, credit card, bank card | AMEX / VISA +/- one attribute (e.g. colour, shape)<br>Or card / credit card + one or more attributes (e.g. colour, shape) |
| 8. Plate | Plate, tray | Plate / tray + one or more attributes (e.g. colour, size, shape, any reference to cloth / napkin) |

# APPENDIX F

## Picture Stimulus Grainsize Scoring Rules, Coding Key and Examples

### GENERAL RULES FOR SCORING

4. Arbitrary demarcation of statements into fine- or coarse-grained examples but essentially the more detail given for a grainsize chunk then it will be deemed a fine-grained response. Paraphrasing or attempting to quote dialogue will be scored as fine-grained.

5. The information is not scored according to whether it is correct or erroneous.

6. If a grainsize chunk is referred to more than once, then only score the finer-grained example (if it is finer-grained).

### PERSON AND SURROUNDING GRAINSIZE CHUNKS AND EXAMPLES

| Person Chunk | Coarse-grained examples / rules | Fine-grained examples / rules |
|---|---|---|
| 1. Clothing | Two or less descriptors (e.g. tone, colour, type) | Three or more descriptors (e.g., tone, colour, type (i.e., shawl)) |
| 2. Age | Old | Middle-age, forties |
| 3. General | Person/lady <= 2 attribute (e.g., on LHS, footpath / side of street / outside, Sweeping / standing / stooping / walking, with broom / cane) | Person / lady + 3 or more attributes (e.g., on LHS, footpath / side of street / outside, sweeping/standing/stooping/walking, with broom/cane) |

| Surrounding Chunk | Coarse-grained examples / rules | Fine-grained examples / rules |
|---|---|---|
| 1. Buildings | Buildings <= 2 attributes (e.g., colourful, specific colours, multi-storey, size, age, features, type (houses / apartments / residential), locations) | Buildings + 3 or more attributes (e.g., colourful, specific colours, multi-storey, size, age, features , type(houses / apartments / residential), locations) |

| 2. Street | Street/road/alley way <= 2 attributes (e.g., size, location, courtyard, material, sign, power lines, garbage can, footpath) | Street / road / alley way / footpath + 3 or more attributes (e.g., size, location, courtyard, material, sign, power lines, garbage can, footpath) |
|---|---|---|
| 3. Trees | Trees +/- 1 attribute (e.g., no leaves, location, number) | Tree / trees + 2 or more attributes (e.g., no leaves, location, number) |
| 4. Vehicles | Car/cars/moped <= 2 attributes (e.g., colour, type, number, parked in street, location) | Car / cars / moped + 3 or more attributes (e.g., colour, type, number, parked in street, location) |
| 5. Location | Region (e.g., Europe) | Region + town / village<br>Specific country +/- town / village |

**APPENDIX G**

**R Code used to Calculate Permutation ANOVA (adapted from Howell, 2009)**

The independent variables shown in the following code are related to Experiment 1, however, the same code was used to calculate permutation ANOVA for Experiments 2 and 3 (using the relevant independent variables and sample size). The same code was also used to calculate permutation ANOVA for the picture stimulus in Experiment 1 (removing lines of code so that only the one-way naivety effect was assessed).

```
# First step is to run a standard ANOVA on the data (dataset table structure comprised
column-oriented variables)

#FMRC is the F-statistic for the mental-reinstatement-of-context main effect

#FNVT is the F-statistic for the naivety instruction main effect

#Finteract is the F-statistic for the MRCxNVT interaction

mod1<-lm(variable~MRC*NVT)        #variable, e.g., correct recall (NOTE: for the picture
                                  stimulus, the model was run as a one-way ANOVA)

ANOVA<-Anova(mod1, type=2)        #use Type II sum of squares (NOTE: Type III was used
                                  for Experiment 3)

cat(" The standard ANOVA for these data follows","\n")

FMRC<-ANOVA[1,3]                  # Saves the MRC main effect value

FNVT<-ANOVA[2,3]                  # Saves the NVT main effect value

Finteract<-ANOVA[3,3]            # Saves the MRCxNVT interaction value

print(ANOVA)

etaSquared(mod1, type=2)         #calculates effect size using lsr package (Navarro,
                                  2015). NOTE: Type III was used for Experiment 3.

cat("\n")

cat("\n")


cat( "Resampling as in Manly with unrestricted sampling of observations..","\n")


# Now start resampling
```

```
# FMrc is the permuted F-statistic of the mental-reinstatement-of-context main effect

# FNvt is the permuted F-statistic of the naivety instruction main effect

# FMrcNvt is the permuted F-statistic of the MRCxNVT interaction

nreps <- 2000                #number of permutations

FMrc <- numeric(nreps)       #Set up space to store permuted F values as each is calculated

FNvt <- numeric(nreps)

FMrcNvt <- numeric(nreps)

FMrc[1] <- FMRC              #the first F value of the 2000 permutation

FNvt[1] <- FNVT

FMrcNvt[1] <- Finteract

for (i in 2:nreps) {

        newvariable<-sample(variable,81)  #the raw data of the 81 participants is sampled
                                          (NOTE: this number is '61' for the sub-set of
                                          data used for the picture stimulus, '131' for
                                          Experiment 2, and '192' for Experiment 3)

        mod2<-lm(newvariable~MRC*NVT)

        b <- Anova(mod2, type=2)          # Type III sum of squares was used for Expt 3

        FMrc[i] <- b[1,3]

        FNvt[i] <- b[2,3]

        FMrcNvt[i] <- b[3,3]

        }

        probMrc <- length(FMrc[FMrc >= FMRC+.Machine$double.eps^0.5])/nreps
        #calculates

        probNvt <- length(FNvt[FNvt >= FNVT+.Machine$double.eps^0.5])/nreps

        probMrcNvt  <-  length(FMrcNvt[FMrcNvt >=
Finteract+.Machine$double.eps^0.5])/nreps

### The addition of "+ .Machine$double.eps" is an aid against two numbers that differ only
by floating point computer calculations at the extreme.

        cat(" The probability value for the interaction is ",probMrcNvt,"\n")

#returns the permutation p-value (i.e., a value <.05 suggests the MRCxNVT F-statistic is
unlikely to be a Type I error)

        cat(" The probability value for MRC is ", probMrc,"\n")

        cat(" The probability value for NVT is ", probNvt,"\n")
```

# APPENDIX H

## Supplementary Analyses for Report Centrality

These supplementary analyses explored the idea that the report-detail instruction produces more informative central and peripheral information – by demanding everything is described in as much detail as possible – while the naivety instruction produces more informative central information. Specifically, the analyses explored the idea that the naivety instruction produces finer-grained central information because it has been observed that central information is recalled more accurately than peripheral information (Paz-Alonso et al., 2013). Thus, report centrality might explain why the naivety instruction produced finer-grained reports with more correct details but not errors, in Experiment 1 (see Section 3.2.3.2, p. 61) and Experiment 2 (see Section 4.2.3.1, p. 89).

## Method

The key concept for these supplementary analyses is report centrality. This was operationally defined by the centrality of the grainsize chunks reported by participants. To determine the information centrality, I adopted the thematic approach of Roberts and Higham (2002), who used four police officers and a crown counsel to judge if information was relevant or peripheral to an investigation and/or court proceeding. I used two police officers (Chief Inspector and Senior Sergeant level) with the South Australian Police. The officers independently judged each grainsize chunk on the grainsize coding key (that was used to score participants' grainsize precision; see Appendix E) as relevant or peripheral to an investigation. They did this by determining if the information related to the grainsize chunk would be treated as evidence-in-chief (i.e., critical or best evidence to the outcome of the investigation) or secondary evidence (i.e., supporting evidence that is helpful but not essential to the outcome of the investigation). Each grainsize chunk was categorised according to this determination and labelled 'central' for evidence-in-chief, and 'peripheral' for secondary

evidence. An inter-rater reliability of 78.6%, assessed by percentage agreement, was established for the officer's categorisation. They discussed disagreements. The senior (Chief Inspector) officer's categorization, except where it was changed in discussion with the second officer, was retained for coding the centrality of grainsize chunks. Because the participants grainsize chunks had previously been coded as coarse-grained or fine-grained, it was a simple process (albeit time consuming) to categorize each grainsize chunk as central or peripheral information.

**Dependent variables to assess report centrality**

Three variables were used to measure report centrality. The first variable, central grainsize chunks, was calculated for each participant by tallying the number of coarse-grained and fine-grained central chunks recalled. The second variable, peripheral grainsize chunks, was calculated by tallying the number of coarse-grained and fine-grained peripheral chunks recalled. I was interested in these two variables to explore if the report-detail instruction produced more central and peripheral information, and if the naivety instruction produced more central information.

The third variable, central grainsize precision, was calculated for each participant by dividing the number of fine-grained central chunks recalled, by the total number of central grainsize chunks recalled (coarse-grained central chunks + fine-grained central chunks). I was interested in this variable to examine if the naivety instruction produced finer-grained central information.

<div align="center">

**Results**

</div>

Descriptive statistics for the different interview conditions are presented in Table A.1. The table shows that the three report centrality variables were influenced by interview condition.

Table A.1

*Mean Centrality [and 95% confidence intervals] by Interview Condition*

| | Interview Condition | | | | | |
|---|---|---|---|---|---|---|
| | RDT-a NVT-a | RDT-a NVT-w | RDT-a NVT-s | RDT-p NVT-a | RDT-p NVT-w | RDT-p NVT-s |
| | ($n$ = 22) | ($n$ = 21) | ($n$ = 22) | ($n$ = 22) | ($n$ = 22) | ($n$ = 22) |
| Variable | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* |
| Central | 16.36 (5.39) | 19.43 (5.47) | 18.36 (3.90) | 19.68 (3.73) | 19.45 (4.55) | 20.50 (2.81) |
| | [13.98, 18.75] | [16.94, 21.92] | [16.64, 20.09] | [18.03, 21.34] | [17.44, 21.47] | [19.26, 21.74] |
| Periph | 7.95 (3.39) | 9.24 (3.08) | 8.95 (3.11) | 9.95 (3.05) | 9.73 (4.51) | 9.82 (2.06) |
| | [6.45, 9.46] | [7.84, 10.64] | [7.58, 10.33] | [8.60, 11.31] | [7.73, 11.73] | [8.90, 10.73] |
| Cent Prec | .47 (.17) | .58 (.17) | .55 (.13) | .51 (.14) | .56 (.13) | .57 (.13) |
| | [.40, .55] | [.51, .66] | [.49, .61] | [.45, .58] | [.50, .61] | [.51, .63] |

Note. Central = number of central grainsize chunks; Periph = number of peripheral grainsize chunks; Cent Prec = proportion of fine-grained central recall.

RDT-a = report-detail instruction absent; RDT-p = report-detail instruction present; NVT-a = naivety instruction absent; NVT-w = weak naivety instruction; NVT-s = strong naivety instruction.

## Data screening

Parametric assumptions for each interview condition were checked using the Shapiro-Wilk normality test and the Levene's test for homogeneity of variance (using the median). Central grainsize precision met parametric assumptions. Central grainsize chunks was skewed (in one condition) with equal variance. Peripheral grainsize chunks was normally distributed with unequal variance.

Outliers were examined with standardized data for each interview condition and identified when $z$-scores were less than -2.50 or greater than 2.50. No outliers were found in central grainsize precision or peripheral grainsize chunks but they were found in central grainsize chunks (3 outliers).

Permutation ANOVA was used to check if parametric violations and outliers, when present, did not bias the statistical inference that was determined from parametric ANOVA. The same approach was taken as outlined in Chapter 3 (see Section 3.2.1.3, p. 52). For each dependent variable that showed a significant parametric ANOVA result, the permutation *p*-value that was obtained with permutation ANOVA, was less than .05. This suggested that the parametric ANOVA result was reliable.

**How the report-detail and naivety instructions impacted report centrality**

These analyses examined how the report-detail and naivety instructions influenced report centrality according to (a) the number of central and peripheral grainsize chunks recalled, and (2) the grainsize precision of central information.

Inferential analyses were calculated using univariate factorial ANOVA (alpha level set at .05) in IBM SPSS Statistics 21. A Type II sum of squares calculation was used due to the unbalanced design albeit the interview conditions only differed in size by one person. Effect size, eta squared ($\eta^2$), was calculated using the *lsr* package (version 0.5; Navarro, 2015) in R (version 3.2.2; R Core Team, 2014). For significant effects, point estimates of the mean are given with 95% confidence intervals. These were calculated with the bias-corrected and accelerated bootstrap method in SPSS, using 2000 permutations. Verbal description for eta squared adopt Cohen's (1988) recommendations (i.e., .01 = 'small', .06 = 'medium', .14 = 'large').

In presenting the results, I focus on the main effects that the report-detail and naivety instructions had on report centrality (see Figures A.1 to A.3) because there was no interaction observed on any dependent variable: central grainsize chunks, $F(2, 125) = 1.56$, $p = .21$, $\eta^2 = .02$; peripheral grainsize chunks, $F(2, 125) = 0.63$, $p = .54$, $\eta^2 = .01$; and central grainsize precision, $F(2, 125) = 0.60$, $p = .55$, $\eta^2 = .01$.

As expected, the report-detail instruction increased the number of central grainsize chunks recalled, $F(1, 125) = 5.73$, $p = .02$, $\eta^2 = .04$ (see Figure A.1), and marginally, the number of peripheral grainsize chunks recalled, $F(1, 125) = 3.84$, $p = .05$, $\eta^2 = .03$ (see Figure A.2). Interviews with the report-detail instruction, had a small to medium sized effect on producing more central information ($M = 19.88$, $SE = 0.46$, $CI_{95}[18.96, 20.75]$) and more peripheral information ($M = 9.83$, $SE = 0.41$, $CI_{95}[9.07, 10.62]$), than interviews without the instruction (central, $M = 18.03$, $SE = 0.63$, $CI_{95}[16.75, 19.26]$; peripheral, $M = 8.71$, $SE = 0.40$, $CI_{95}[7.93, 9.50]$).
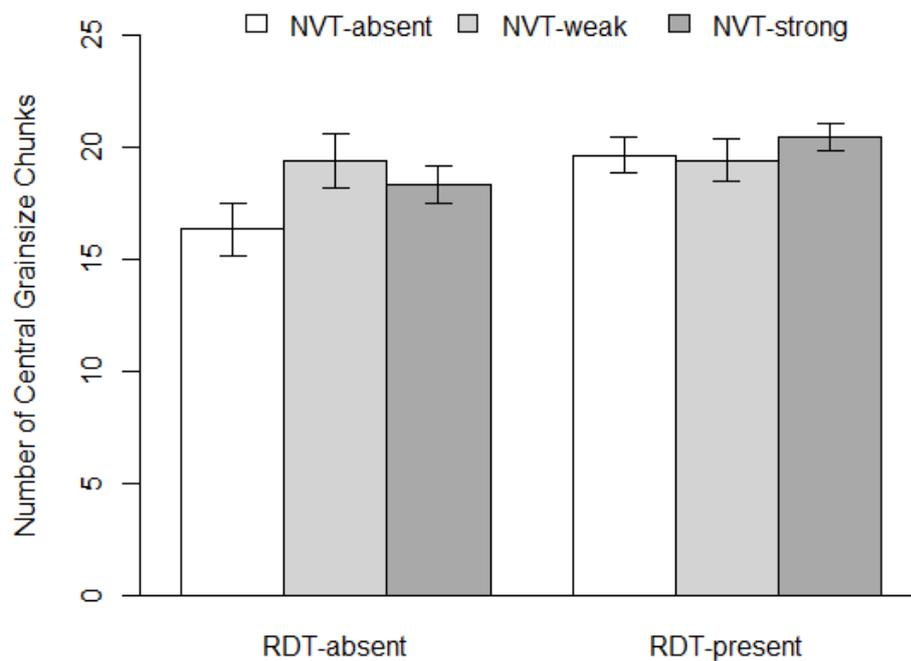


*Figure A.1*. Mean central grainsize chunks (+SE) for each interview condition. RDT = report-detail instruction. NVT = naivety instruction.
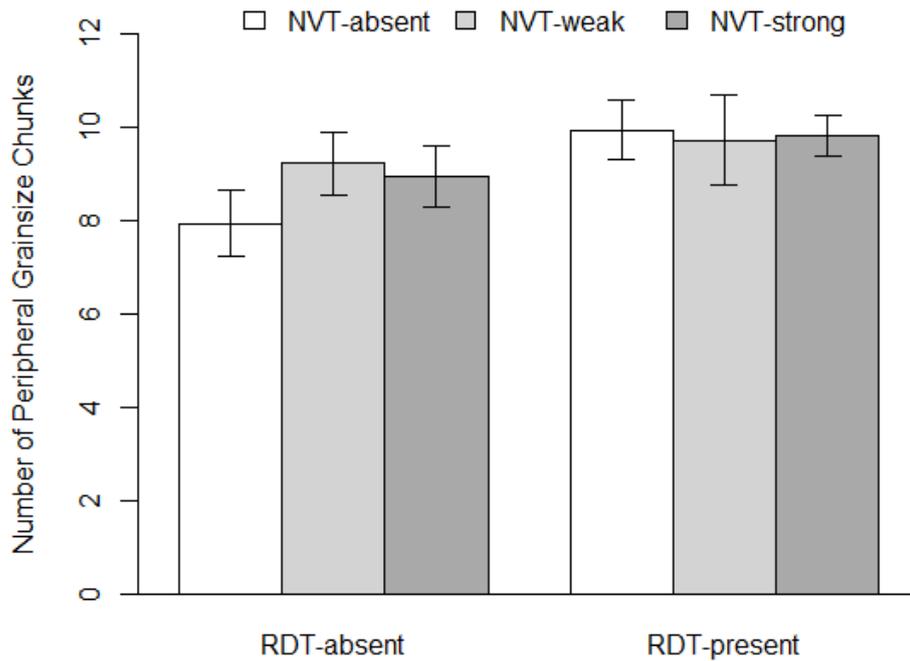
*Figure A.2.* Mean peripheral grainsize chunks (+SE) for each interview condition. RDT = report-detail instruction. NVT = naivety instruction.

Notably, the report-detail instruction did not influence the grainsize precision of central information, $F(1, 125) = 0.20$, $p = .65$, $\eta^2 < .01$ (see Figure A.3). This result might reflect a lack of power with the study because the report-detail instruction had an observed power of .07.

Turning to the naivety instruction, it was curious to find that it did not influence the number of central grainsize chunks recalled, $F(2, 125) = 1.49$, $p = .23$, $\eta^2 = .02$ (see Figure A.1), or the number of peripheral grainsize chunks recalled, $F(2, 125) = 0.32$, $p = .73$, $\eta^2 < .01$ (see Figure A.2). However, as expected, the naivety instruction influenced the grainsize precision of central information, $F(2, 125) = 3.48$, $p = .03$, $\eta^2 = .05$ (see Figure A.3). A Tuckey post hoc test revealed a difference in central grainsize precision ($p = .04$, $d = .50$), suggesting that interviews with the weak naivety statement had a medium sized effect on producing finer-grained central information ($M = 0.57$, $SE = 0.02$, CI$_{95}$[0.53, 0.61]), than interviews without a naivety statement ($M = 0.49$, $SE = 0.02$, CI$_{95}$[0.44, 0.54]). There was also a marginal difference in central grainsize precision ($p = .10$, $d = .45$), suggesting that

interviews with the strong naivety statement had a medium-sized effect on producing finer-grained central information ($M = 0.56$, $SE = 0.02$, $CI_{95}[0.52, 0.59]$), than interviews without a naivety statement ($M = 0.49$, $SE = 0.02$, $CI_{95}[0.44, 0.54]$). There was no difference in central grainsize precision between the strong and weak naivety statements ($p = .92$, $d = .09$).
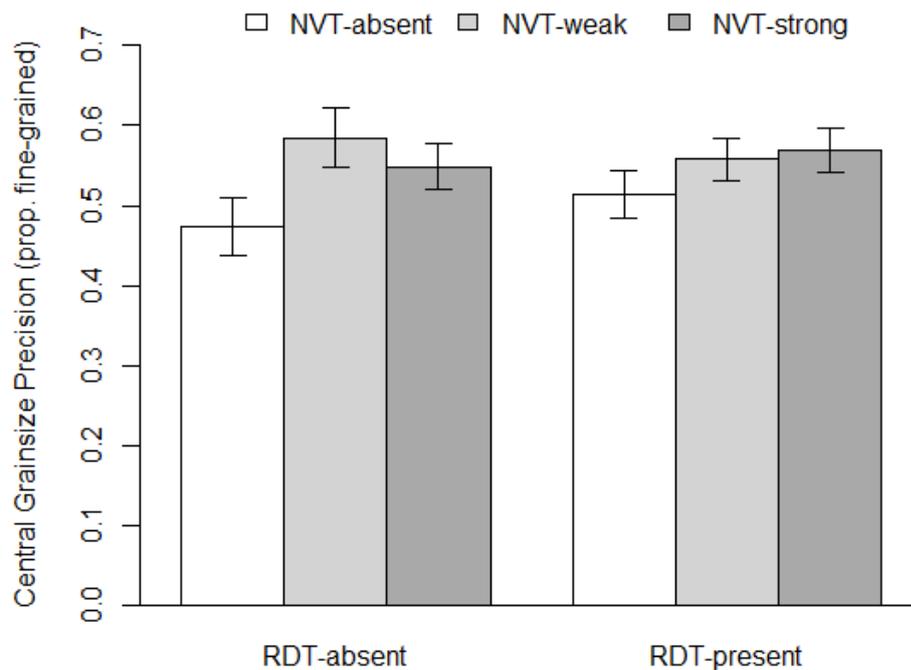


*Figure A.3.* Mean central grainsize precision (+SE) for each interview condition. RDT = report-detail instruction. NVT = naivety instruction.

In summary, the results are consistent with the idea that the report-detail instruction, by demanding a witness describe everything in as much detail as possible, produces more informative central and peripheral information. The results also support the idea that the naivety instruction produces finer-grained central information. Moreover, although the effect was marginal for the strong naivety statement, the results also suggest that finer-grained central information will be produced regardless of the form of the naivety statement.

**Discussion**

The results are discussed in Section 4.3.3 in Chapter 4

# APPENDIX I

## Linguistic Qualifiers Scoring Rules, Coding Key and Examples

### GENERAL RULES FOR SCORING

1. Linguistic qualifiers are scored as hedges or 'don't know' responses (see examples below).

2. Only score qualifiers associated with details that have been scored according to the quantity coding key.

3. When synonyms are used, count all qualifiers related to the detail of information (e.g., "I think it was a 4WD I'm pretty sure").

4. Do not score repeated qualifiers (e.g., "I'm pretty sure it was a van, I'm pretty sure it was a van").

5. Do not score qualifiers when they are associated with suppositions (because suppositions are not scored with the quantity coding key).

6. Do not score qualifiers when they are associated with any titles given for the actors (e.g., I'm not sure if he was the waiter or bartender) because titles were not scored with the quantity coding key.

7. Do not score qualifiers of the 'sure bet' kind (e.g., "I'm certain", "I'm sure", "I recall", etc.).

8. Do not score "either" and "or". Although these suggests participants are using a plurality option in their uncertainty (i.e., Luna et al., 2011), the quantity coding key only scored one of the details given.

9. Do not score statements that refer to the relevance of information (e.g., "I'm not sure if that's relevant to the police").

10. Categorise qualifiers into Person, Surrounding and Action categories (e.g., "he put it, I don't know, in his pocket or something", qualifiers are related to location therefore categorised as Surrounding)

11. Do not score concluding statements (e.g., "… and that's about all").

12. For the narrative responses to the open questions, because only new and/or finer-grained details are scored with the quantity coding key, then only the linguistic qualifiers associated with this information are scored. However, any additional unique qualifiers observed with any 'old' information are also scored.

# HEDGES EXAMPLES

| | | |
|---|---|---|
| "I guess" | "I'm guessing" | "I think" |
| "I don't think" | "I thought" | |
| "I'm pretty sure" | "I'm fairly sure" | |
| "possibly" | "around" | "about" |
| "couldv'e been" | "approximately" | "maybe" |
| "possibly" | "perhaps" | "roughly" |
| "It might have been" | "may have been" | "hopefully" |
| "I reckon" | "something like" | "or something" |
| "Or something like that" | "or someone" | "or anything" |
| "or whatever" | "or what" | "if anything" |
| "almost a" | "sort of" | "some sort of" |
| "some sort" | "kind of" | "type of thing" |
| "it was like" | "like a" | "It didn't look like" |
| "If I observed correctly" | "from what I could see" | "that's what it looked like" |
| "it sounded" | "sounds" | "sounding" |
| "of some description" | "something along those lines" | "along the lines of" |
| "It seemed" | "it didn't seem" | "looked like" |
| "looked" | "looking" | "appears to be" |
| "it would appear" | "appeared to be" | "in appearance" |
| "would have" | "must have" | "might have" |
| "might not have" | "whether" | "as far as I can tell" |
| "as far as I know" | "I suppose" | "presumably" |
| "I assume" | "I believe" | "as I recall" |
| "from what I remember" | "if I remember correctly" | "if I'm not mistaken" |
| "I can see it somewhere" | "I'd say" | "I would say" |
| "I want to say" | "I wouldn't say" | "I take it to be" |
| "It was harder to say" | "harder to explain it" | "I'm fading here a bit" |
| "not totally clear" | "I tried to concentrate" | "fuzzy recollection" |
| "I could be wrong" | "I'm not too good with" | "I could've created that" |
| "could be making it up" | "I'm making that up" | "just creating things" |
| "I would hazard a guess" | "that would be my guess" | "I vaguely remember" |
| "unless I've forgotten" | "I'm not too familiar" | "basically" |
| "don't remember too much" | "I don't recall much but" | "not much more" |

"couldn't really give you"     "not really"

## DON'T KNOW EXAMPLES

| | | |
|---|---|---|
| "I don't know" | "I don't really know" | "dunno" |
| "I can't remember" | "completely forgotten that" | "I really can't remember" |
| "I don't recall" | "it'll come to mind" | "escapes me" |
| "I forgot" | "got a bit of a blank there" | |
| "I can't exactly remember" | | |
| "can't remember specifically" | "I can't really remember" | "I don't really recall" |
| "I can't quite recall" | "Not off the top of my head" | "I can't pinpoint" |
| "I wouldn't put any money on that" | "I couldn't swear on it" | "could not guarantee" |
| "don't quote me on that" | "not confident with that" | "I'm not sure" |
| "I'm not entirely sure" | "not certain" | "not too sure" |
| "I'm not 100% sure" | "I can't say for sure" | "couldn't really tell" |
| "it's hard to tell" | "I couldn't pick the type" | "not totally clear" |
| "I couldn't make that out for certain" | | "I didn't really take note" |
| "I didn't really hear" | | |
| "couldn't see it clearly" | "I didn't really see" | "didn't really get a look" |
| "didn't pay much attention" | "didn't quite pay attention" | |
| "I didn't notice too much" | | |
| "I can't hardly help you" | "I can't exactly help you" | |
| "can't tell you much of that" | "all I've really got" | |
| "I don't quite remember what" | | |

## NON-SCORABLE EXAMPLES

Smith and Clark (1993) found the phrase "I have no idea" was associated with recognition performance that was not better than guessing. Therefore, this phrase and similar phrases were treated as reflecting the limit of memory (i.e., purely guessing) and were not scored.

| | | |
|---|---|---|
| "No idea" | "who knows" | |
| "I wouldn't know at all" | | |
| "I don't know whatsoever" | "I didn't particularly notice" | "I didn't notice" |
| "I can't honestly say I took much notice" | | |
| "I didn't take note" | | |

"I really didn't pay attention"

"I didn't pay attention"          "that's all I can recall"

"don't remember anything else"

"that's as much as I can remember"     "that's all I can think of"

"can't think of anything else"     "that's all I remember"

"couldn't tell you that"          "I honestly couldn't tell you"

"completely missed that"     "I can't see"          "I didn't see"

"I didn't catch"          "I didn't get"          "I didn't hear"

"other than that I can't say"     "I can't say"          "I really didn't catch"

"not that I can think of"     "I couldn't make out the detail"

"cannot give you the name"

"I would not be able to identify it"

"couldn't be more specific than that"

"that's as much detail as I can give"

**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*(5), 1224–1245. doi: 10.1037/a0012938

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

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a
metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235. doi:
10.1177/1088868309341564

Anderson, M. J. (2001). Permutation tests for univariate or multivariate analysis of variance
and regression. *Canadian Journal of Fisheries and Aquatic Sciences, 58*(3), 626–639.
doi: 10.1139/cjfas-58-3-626

Aschermann, E., Mantwill, M., & Köhnken, G. (1991). An independent replication of the
effectiveness of the Cognitive Interview. *Applied Cognitive Psychology, 5*(6), 489–
495. doi: 10.1002/acp.2350050604

Baddeley, A. D. (2001). Is working memory still working? *American Psychologist, 56*(11),
851–864. doi: 10.1037/0003-066X.56.11.851

Barnes, A. E., Nelson, T. O., Dunlosky, J., Mazzoni, G., & Narens, L. (1999). An integrative
system of metamemory components involved in retrieval. In A. E. Barnes, T. O.
Nelson, J. Dunlosky, G. Mazzoni, & L. Narens (Eds.), *Attention and performance
XVII* (pp. 285–313). Cambridge, MA: MIT Press.

Bekerian, D. A., & Dennett, J. L. (1993). The Cognitive Interview technique: Reviving the
issues. *Applied Cognitive Psychology, 7*(4), 275–297. doi: 10.1002/acp.2350070403

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*(5), 669–679. doi: 10.1037/0022-3514.56.5.669

Bensi, L., Nori, R., Gambetti, E., & Giusberti, F. (2011). The Enhanced Cognitive Interview: A study on the efficacy of shortened variants and single techniques. *Journal of Cognitive Psychology, 23*(3), 311–321. doi: 10.1080/20445911.2011.497485

Blank, H. (2009). Remembering: A theoretical interface between memory and social psychology. *Social Psychology, 40*(3), 164–175. doi: 10.1027/1864-9335.40.3.164

Boon, J., & Noon, E. (1994). Changing perspectives in cognitive interviewing. *Psychology, Crime & Law, 1*(1), 59–69. doi: 10.1080/10683169408411936

Bower, G. (1967). A multicomponent theory of the memory trace. *Psychology of Learning and Motivation, 1*, 229–325. doi: 10.1016/S0079-7421(08)60515-0

Brewer, N., Muller, T., Nagesh, A., Hope, L., & Gabbert, F. (2010, March). *Interviewing eyewitnesses: Enhancing outputs by controlling the grain size of reports.* Paper presented at the American Psychology-Law Society Annual Conference, Vancouver, BC, Canada. Retrieved from http://convention2.allacademic.com/one/apls/apls10/index.php?click_key=1&cmd=Multi+Search+Search+Load+Publication&publication_id=399082&PHPSESSID=biq91dtitoqhf9313h9jb5l387

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13*(2), 207–230. doi: 10.1016/0010-0285(81)90008-6

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

Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. *Psychology of Learning and Motivation, 32*, 275–318. doi: 10.1016/S0079-7421(08)60313-8

Buratti, S., MacLeod, S., & Allwood, C. M. (2014). The effects of question format and co-witness peer discussion on the confidence accuracy of children's testimonies. *Social Influence, 9*(3), 189–205. doi: 10.1080/15534510.2013.804434

Campos, L., & Alonso-quecuty, M. L. (1999). The Cognitive Interview: Much more than simply "try again". *Psychology, Crime & Law, 5*(1), 47–59. doi: 10.1080/10683169908414993

Clarke, J., Prescott, K., & Milne, R. (2013). How effective is the Cognitive Interview when used with adults with intellectual disabilities specifically with conversation recall? *Journal of Applied Research in Intellectual Disabilities, 26*(6), 546–556. doi: 10.1111/jar.12049

Clifford, B. R., & George, R. (1996). A field evaluation of training in three methods of witness/victim investigative interviewing. *Psychology, Crime and Law, 2*(3), 231–248. doi: 10.1080/10683169608409780

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. doi: 10.1177/001316446002000104

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. doi: 10.1037/0033-2909.112.1.155

Compo, N. S., Gregory, A. H., & Fisher, R. (2012). Interviewing behaviors in police investigators: A field study of a current US sample. *Psychology, Crime & Law, 18*(4), 359–375. doi: 10.1080/1068316X2010.494604

Coyne, A. C. (1985). Adult age, presentation time, and memory performance. *Experimental Aging Research, 11*(3–4), 147–149. doi: 10.1080/03610738508259178

Dando, C. J., Geiselman, R., MacLeod, N., & Griffiths, A. (2016). Interviewing adult witnesses and victims. 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* (pp. 79–106). Chichester, England: Wiley-Blackwell.

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*(1), 59–70. doi: 10.1348/135532506X162498

Dando, C. J., Wilcock, R., & Milne, R. (2009). The Cognitive Interview: The efficacy of a modified mental reinstatement of context procedure for frontline police investigators. *Applied Cognitive Psychology, 23*, 138–147. doi: 10.1002/acp.1451

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

Davis, M. R., McMahon, M., & Greenwood, K. M. (2005). The efficacy of mnemonic components of the Cognitive Interview: Towards a shortened variant for time-critical investigations. *Applied Cognitive Psychology, 19*(1), 75–93. doi: 10.1002/acp.1048

Dietze, P. M., Powell, M. B., & Thomson, D. M. (2012). Examination of the effect of mental reinstatement of context across developmental level, retention interval and type of mnemonic instruction. *Psychiatry, Psychology and Law, 19*(1), 89–103. doi: 10.1080/13218719.2010.543410

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*(4), 290–301. doi: 10.1037/lhb0000034

Emmett, D., Clifford, B., & Gwyer, P. (2003). An investigation of the interaction between cognitive style and context reinstatement on the memory performance of eyewitnesses. *Personality and Individual Differences, 34*(8), 1495–1508. doi: 10.1016/S0191-8869(02)00131-9

Emmett, D., Clifford, B., Young, K., Kandova, M., & Potton, A. (2006). The use of a computer presented virtual licence plate and mental context reinstatement to improve recall accuracy of both licence plate and related event details. *Applied Cognitive Psychology, 20*(9), 1195–1207. doi: 10.1002/acp.1259

Erev, I., & Cohen, B. L. (1990). Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational Behavior and Human Decision Processes, 45*(1), 1–18. doi: 10.1016/0749-5978(90)90002-Q

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

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532–538. doi: 10.1037/a0015808

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). New York, NY: Cambridge University Press.

Fisher, R. P. (1996). Implications of output-bound measures for laboratory and field research in memory. *Behavioral and Brain Sciences, 19*, 197–197. doi: 10.1017/S0140525X00042229

Fisher, R. P. (2010). Interviewing cooperative witnesses. *Legal and Criminological Psychology, 15*(1), 25–38. doi: 10.1348/135532509X441891

Fisher, R. P., Compo, N. S., Rivard, J., & Hirn, D. (2013). Interviewing witnesses. In T. J. Perfect & D. S. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 559–578). London, England: SAGE Publications.

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

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

Fisher, R. P., Geiselman, R., Raymond, D. S., Jurkevich, L. M., & Warhaftig, M. L. (1987). Enhancing enhanced eyewitness memory: Refining the Cognitive Interview. *Journal of Police Science and Administration, 15*(4), 291–297.

Fisher, R. P., Geiselman, R. E., & Amador, M. (1989). Field test of the Cognitive Interview: Enhancing the recollection of actual victims and witnesses of crime. *Journal of Applied Psychology, 74*(5), 722–727. doi: 10.1037/0021-9010.74.5.722

Flexser, A. J., & Tulving, E. (1978). Retrieval independence in recognition and recall. *Psychological Review, 85*(3), 153–171. doi: 10.1037/0033-295X.85.3.153

Gabbert, F., Hope, L., Carter, E., Boon, R., & Fisher, R. (2016). 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* (pp. 107–131). Chichester, UK: 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*(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

Galvin, S., Podd, J., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review, 10*(4), 843–876. doi: 10.3758/BF03196546

Gastwirth, J. L., Gel, Y. R., Hui, W. L. W., Lyubchich, V., Miao, W., & Noguchi, K. (2014). An R package for biostatistics, public policy, and law (Version 2.4.1) [R computer software package]. Retrieved from http://CRAN.R-project.org/package=lawstat

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

Geiselman, R. E., & Callot, R. (1990). Reverse versus forward recall of script-based texts. *Applied Cognitive Psychology, 4*(2), 141–144. doi: 10.1002/acp.2350040206

Geiselman, R. E., Fisher, R. P., Cohen, G., Holland, H., & Surtes, L. (1986). Eyewitness responses to leading and misleading questions under the Cognitive Interview. *Journal of Police Science and Administration, 14*(1), 31–39.

Geiselman, R. E., Fisher, R. P., Firstenberg, I., Hutton, L. A., Sullivan, S. J., Avetissian, I. V., & Prosk, A. L. (1984). Enhancement of eyewitness memory: An empirical evaluation of the Cognitive Interview. *Journal of Police Science and Administration, 12*(1), 74–80.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus

hypnosis. *Journal of Applied Psychology, 70*(2), 401–412. doi: 10.1037/0021-9010.70.2.401

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1986). Enhancement of eyewitness memory with the Cognitive Interview. *The American Journal of Psychology, 99*(3), 385–401. doi: 10.2307/1422492

Goldsmith, M., & Koriat, A. (1999). The strategic regulation of memory reporting: Mechanisms and performance consequences. In A. E. Barnes, T. O. Nelson, J. Dunlosky, G. Mazzoni, & L. Narens (Eds.), *Attention and performance XVII* (pp. 372–400). Cambridge, MA: MIT Press.

Goldsmith, M., & Koriat, A. (2008). The strategic regulation of memory accuracy and informativeness. *Psychology of Learning and Motivation, 48*, 1–60. doi: 10.1016/S0079-7421(07)48001-X

Goldsmith, M., Koriat, A., & Pansky, A. (2005). Strategic regulation of grain size in memory reporting over time. *Journal of Memory and Language, 52*(4), 505–525. doi: 10.1016/j.jml.2005.01.010

Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size in memory reporting. *Journal of Experimental Psychology: General, 131*(1), 73–95. doi: 10.1037/0096-3445.131.1.73

Goldsmith, M., Pansky, A., & Koriat, A. (2013). Metacognitive control of memory reporting. In T. J. Perfect & D. S. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 481–500). London, England: SAGE Publications.

Granhag, P. A., Jonsson, A.-C., & Allwood, C. M. (2004). The Cognitive Interview and its effect on witnesses' confidence. *Psychology, Crime & Law, 10*(1), 37–52. doi: 10.1080/1068316021000030577

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Greenberg, M. S., Westcott, D. R., & Bailey, S. E. (1998). When believing is seeing: The effect of scripts on eyewitness memory. *Law and Human Behavior, 22*(6), 685–694. doi: 10.1023/A:1025758807624

Grice, J. B. (1975). Logic and conversation. *Syntax and Semantics, 3*, 41–58.

Gwyer, P., & Clifford, B. R. (1997). The effects of the Cognitive Interview on recall, identification, confidence and the confidence/accuracy relationship. *Applied Cognitive Psychology, 11*(2), 121–145. doi: 10.1002/(SICI)1099-0720(199704)11:2<121::AID-ACP443>3.0.CO;2-L

Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition, 18*(5), 496–506. doi: 10.3758/BF03198482

Higham, P. A. (2007). No special K! A signal detection framework for the strategic regulation of memory accuracy. *Journal of Experimental Psychology: General, 136*(1), 1–22. doi: 10.1037/0096-3445.136.1.1

Hill, J. A., & Moston, S. (2011). Police perceptions of investigative interviewing: Training needs and operational practices in Australia. *The British Journal of Forensic Practice, 13*(2), 72–83. doi: 10.1108/14636641111134314

Holliday, R. E., Humphries, J. E., Milne, R., Memon, A., Houlder, L., Lyons, A., & Bull, R. (2012). Reducing misinformation effects in older adults with Cognitive Interview mnemonics. *Psychology and Aging, 27*(4), 1191–1203. doi: 10.1037/a0022031

Hope, L., Gabbert, F., & Brewer, N. (2011, August). *Regulating memory: Increasing output quantity and diagnosing accuracy with a grain size approach.* Paper presented at the 5th International Conference on Memory, York, England. Retrieved from https://www.york.ac.uk/conferences/icom5/Abstract%20Booklet_ICOM5220711.pdf

Hope, L., Gabbert, F., Brewer, N., Tull, M., & Nagesh, A. (2010, June). *Interviewing eyewitnesses: Enhancing output quality and diagnosing accuracy using the Grain Size*

*technique.* Paper presented at the 20th Conference of the European Association of
Psychology and Law, Gothenburg, Sweden.

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. doi: 10.1111/j.2044-8333.2011.02015.x

Hope, L., Mullis, R., & Gabbert, F. (2013). Who? What? When? Using a timeline technique
to facilitate recall of a complex event. *Journal of Applied Research in Memory and Cognition, 2*(1), 20–24. doi: 10.1016/j.jarmac.2013.01.002

Howell, D. C. (2009). *Permutation tests for factorial ANOVA designs.* Retrieved from
http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Permutation%20Anova/PermTestsAnova.html

Jack, F., Leov, J., & Zajac, R. (2014). Age-related differences in the free-recall accounts of
child, adolescent, and adult witnesses. *Applied Cognitive Psychology, 28*(1), 30–38. doi: 10.1002/acp.2951

Kahneman, D., & Tversky, A. (1982). Intuitive prediction: Biases and corrective procedures.
In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 414–421). New York, NY: Cambridge University Press.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings.*
New York, NY: Oxford University Press.

Kebbell, M. R., & Milne, R. (1998). Police officers' perceptions of eyewitness performance in
forensic investigations. *The Journal of Social Psychology, 138*(3), 323–330. doi: 10.1080/00224549809600384

Kebbell, M. R., Milne, R., & Wagstaff, G. F. (1999). The Cognitive Interview: A survey of its
forensic effectiveness. *Psychology, Crime and Law, 5*(1–2), 101–115. doi: 10.1080/10683169908414996

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*(1), 1–24. doi: 10.1006/jmla.1993.1001

Köhnken, G., Milne, R., Memon, A., & Bull, R. (1999). The Cognitive Interview: A meta-analysis. *Psychology, Crime & Law, 5*(1–2), 3–27. doi: 10.1080/10683169908414991

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123*(3), 297–315. doi: 10.1037/0096-3445.123.3.297

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory and accuracy. *Psychological Review, 103*(3), 490–517. doi: 10.1037//0033-295X.103.3.490

Lamb, M. E., Orbach, Y., Hershkowitz, I., Esplin, P. W., & Horowitz, D. (2007). A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: A review of research using the NICHD Investigative Interview Protocol. *Child Abuse and Neglect, 31*(11–12), 1201–1231. doi: 10.1016/j.chiabu.2007.03.021

Liberman, V., & Tversky, A. (1993). On the evaluation of probability judgments: Calibration, resolution, and monotonicity. *Psychological Bulletin, 114*(1), 162–173. doi: 10.1037/0033-2909.114.1.162

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). New York, NY: Cambridge University Press.

Loftus, E. F. (1975). Reconstructing memories: The incredible eyewitness. *Jurimetrics, 15*(3), 188–193.

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*(2), 148–158. doi: 10.1037/a0023276

Luna, K., & Martin-Luengo, B. (2012). Improving the accuracy of eyewitnesses in the presence of misinformation with the plurality option. *Applied Cognitive Psychology, 26*(5), 687–693. doi: 10.1002/acp.2845

Luna, K., Martin-Luengo, B., & Brewer, N. (2015). Are regulatory strategies necessary in the regulation of accuracy? The effect of direct-access answers. *Memory & Cognition, 43*(8), 1180–1192. doi: 10.3758/s13421-015-0534-2.

Luna, K., & Migueles, M. (2009). Acceptance and confidence of central and peripheral misinformation. *Spanish Journal of Psychology, 12*(2), 405–413. doi: 10.1017/S1138741600001797

Mantwill, M., Köhnken, G., & Aschermann, E. (1995). Effects of the Cognitive Interview on the recall of familiar and unfamiliar events. *Journal of Applied Psychology, 80*(1), 68–78. doi: 10.1037/0021-9010.80.1.68

McCallum, N. A., Brewer, N., & Weber, N. (2015, June). *Why eyewitnesses of crime withhold coarse-grain information from memory reports.* Paper presented at the SARMAC XI, Victoria, BC, Canada. Retrieved from http://static1.squarespace.com/static/504170d6e4b0b97fe5a59760/t/55887e52e4b0eb9 9d4b5b905/1435008594892/Program15June.pdf

McCallum, N. A., 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*(3), 375–386. doi: 10.1002/acp.3212

McLean, M. (1995). Quality investigation? Police interviewing of witnesses. *Medicine, Science and Law, 35*(2), 116–122. doi: 10.1177/002580249503500205

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision, 6*(1), 8–17. doi: 10.1167/6.1.2

Mello, E. W., & Fisher, R. P. (1996). Enhancing older adult eyewitness memory with the Cognitive Interview. *Applied Cognitive Psychology, 10*(5), 403–417. doi: 10.1002/(SICI)1099-0720(199610)10:5<403::AID-ACP395>3.0.CO;2-X

Memon, A., & Higham, P. A. (1999). A review of the Cognitive Interview. *Psychology, Crime & Law, 5*(1–2), 177–196. doi: 10.1080/10683169908415000

Memon, A., Holley, A., Milne, R., Köhnken, G., & Bull, R. (1994). Towards understanding the effects of interviewer training in evaluating the Cognitive Interview. *Applied Cognitive Psychology, 8*(7), 641–659. doi: 10.1002/acp.2350080704

Memon, A., Holley, A., Wark, L., Bull, R., & Köhnken, G. (1996). Reducing suggestibility in child witness interviews. *10*(6), 503–518. doi: 10.1002/(SICI)1099-0720(199612)10:6<503::AID-ACP416>3.0.CO;2-R

Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law, 16*(4), 340–372. doi: 10.1037/a0020518

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. doi: 10.1111/j.2044-8295.1997.tb02629.x

Memon, A., Wark, L., Holley, A., Bull, R., & Köhnken, G. (1997). Eyewitness performance in cognitive and structured interviews. *Memory, 5*(5), 639–656. doi: 10.1080/741941481

Migueles, M., & García-Bajos, E. (1999). Recall, recognition, and confidence patterns in eyewitness testimony. *Applied Cognitive Psychology, 13*(3), 257–268. doi: 10.1002/(SICI)1099-0720(199906)13:3<257::AID-ACP566>3.0.CO;2-7

Milne, R. (2004). *The Enhanced Cognitive Interview: A step-by-step guide*. Unpublished manuscript. Retrieved from https://www.google.com.au/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=The+Enhanced+Cognitive+Interview%3A+A+step-by-step+guide

Milne, R., & Bull, R. (1999). *Investigative Interviewing: Psychology and Practice*. West Sussex, England: John Wiley & Sons.

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., Clare, I. C. H., & Bull, R. (1999). Using the Cognitive Interview with adults with mild learning disabilities. *Psychology, Crime & Law, 5*(1–2), 81–99. doi: 10.1080/10683169908414995

Milne, R., Sharman, S. J., Powell, M. B., & Mead, S. (2013). Assessing the effectiveness of the Cognitive Interview for children with severe intellectual disabilities. *International Journal of Disability, Development and Education, 60*(1), 18–29. doi: 10.1080/1034912X.2013.757137

Murphy, A. H. (1973). A new vector partition of the probability score. *Journal of Applied Meteorology, 12*(4), 595–600. doi: 10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2

Navarro, D. J. (2015). Learning statistics with R: A tutorial for psychology students and other beginners (Version 0.5) [R computer software package and lecture notes]. Retrieved from http://health.adelaide.edu.au/psychology/ccs/teaching/lsr/

Nelson, D. L., & Goodmon, L. B. (2003). Disrupting attention: The need for retrieval cues in working memory theories. *Memory & Cognition, 31*(1), 65–76. doi: 10.3758/BF03196083

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109–133. doi: 10.1037/0033-2909.95.1.109

Nelson, T. O. (1986). ROC curves and measures of discrimination accuracy: A reply to Swets. *Psychological Bulletin, 100*(1), 128–132. doi: 10.1037/0033-2909.100.1.128

Nelson, T. O. (1996). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item. *Applied Cognitive Psychology, 10*(3), 257–260. doi: 10.1002/(SICI)1099-0720(199606)10:3<257::AID-ACP400>3.0.CO;2-9

Neuhäuser, M. (2010). A nonparametric two-sample comparison for skewed data with unequal variances. *J Clin Epidemiol, 63*(6), 691–693. doi: 10.1016/j.jclinepi.2009.08.026

Newson, R. (2002). Parameters behind "nonparametric" statistics: Kendall's tau, Somers D and median difference. *The Stata Journal, 2*(1), 45–64.

Ord, B., Shaw, G., & Green, T. (2008). *Investigative interviewing explained* (2nd ed.). Chatswood, Australia: LexisNexis Butterworths.

Pansky, A., Koriat, A., & Goldsmith, M. (2005). Eyewitness recall and testimony. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 93–150). New York, NY: Guilford Press.

Paulo, R. M., Albuquerque, P. B., & Bull, R. (2015). Improving the Enhanced Cognitive Interview: Expressions of uncertainty, motivation and its relation to report accuracy. *Psychology, Crime & Law*, 366–381. doi: 10.1080/1068316X.2015.1109089

Paz-Alonso, P. M., Goodman, G. S., & Ibabe, I. (2013). Adult eyewitness memory and compliance: Effects of post-event misinformation on memory for a negative event. *Behavioral Sciences and the Law, 31*(5), 541–558. doi: 10.1002/bsl.2081

Perfect, T. J., & Weber, N. (2012). How should witnesses regulate the accuracy of their identification decisions: One step forward, two steps back? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(6), 1810–1818. doi: 10.1037/a0028461

Powell, M. B. (2002). Specialist training in investigative and evidential interviewing: Is it having any effect on the behaviour of professionals in the field? *Psychiatry, Psychology and Law, 9*(1), 44–55. doi: 10.1375/pplt.2002.9.1.44

Powell, M. B., Cavezza, C., Hughes-Scholes, C., & Stoove, M. (2010). Examination of the consistency of interviewer performance across three distinct interview contexts. *Psychology, Crime and Law, 16*(7), 585–600. doi: 10.1080/10683160902971063

Py, J., Ginet, M., Desperies, C., & Cathey, C. (1997). Cognitive encoding and cognitive interviewing in eyewitness testimony. *Swiss Journal of Psychology, 56*(1), 33–41.

R Core Team. (2014). R: A language and environment for statistical computing (Version 3.2.2) [Computer software]. Retrieved from http://www.R-project.org/

Reid, N., Brewer, N., & Weber, N. (2013a, June). *Why do eyewitnesses withhold information: The role of informativeness in grain size choice.* Paper presented at the SARMAC X, Rotterdam, The Netherlands. Retrieved from http://static1.squarespace.com/static/504170d6e4b0b97fe5a59760/t/52457a9be4b0aeb 534dd2e3a/1380285083110/SARMAC_X_PaperJune28.pdf

Reid, N., Brewer, N., & Weber, N. (2013b, July). *Why do eyewitnesses withhold information?* Paper presented at the 33rd International Congress on Law and Mental Health, Amsterdam, The Netherlands. Retrieved from http://ialmh.org/wp-content/uploads/2016/03/Amsterdam-2013-Abstracts.pdf

Revelle, W. (2014). Psych: Procedures for psychological, psychometric, and personality research (Version 1.4.8) [R computer program software]. Retrieved from http://CRAN.R-project.org/package=psych

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

Sauer, J. D., & Hope, L. (in press). The effects of divided attention at study and reporting procedure on regulation and monitoring for episodic recall. *Acta Psychologica*.

Schollum, M. (2005). *Investigative interviewing: The literature*. Wellington, New Zealand: Office of the Commissioner of Police.

Schraw, G. (1995). Measures of feeling-of-knowing accuracy: A new look at an old problem. *Applied Cognitive Psychology, 9*(4), 321–332. doi: 10.1002/acp.2350090405

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4*(1), 33–45. doi: 10.1007/s11409-008-9031-3

Schuurmans, E., & Vandierendonck, A. (1985). Recall as communication: Effects of frame anticipation. *Psychological Research, 47*, 119–124. doi: 10.1007/BF00309126

Schwarz, N. (2015). Metacognition. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. Bargh (Eds.), *APA handbook of personality and social psychology, Volume 1: Attitudes and social cognition* (pp. 203–229). Washington, DC: American Psychological Association.

Shapiro, L. R., Blackford, C., & Chen, C. F. (2005). Eyewitness memory for a simulated misdemeanor crime: The role of age and temperament in suggestibility. *Applied Cognitive Psychology, 19*(3), 267–289. doi: 10.1002/acp.1089

Shaw, J. S., Appio, L. M., Zerr, T. K., & Pontoski, K. E. (2007). Public eyewitness confidence can be influenced by the presence of other witnesses. *Law and Human Behavior, 31*(6), 629-652. doi: 10.1007/s10979-006-9080-6

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy, 85*(3), 257–268.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language, 32*, 25–38. doi: 10.1006/jmla.1993.1002

Snook, B., & Keating, K. (2011). A field study of adult witness interviewing practices in a Canadian police organization. *Legal and Criminological Psychology, 16*, 160–172. doi: 10.1348/135532510X497258

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

Sutherland, R., & Hayne, H. (2001). The effect of postevent information on adults' eyewitness reports. *Applied Cognitive Psychology, 15*(3), 249–263. doi: 10.1002/acp.700

Thompson, V. A., Ackerman, R., Sidi, Y., Ball, L. J., Pennycook, G., & Turner, J. A. P. (2013). The role of answer fluency and perceptual fluency in the monitoring and control of reasoning: Reply to. *Cognition, 128*(2), 256–258. doi: 10.1016/j.cognition.2013.03.003

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128*(2), 237–251. doi: 10.1016/j.cognition.2012.09.012

Thoresen, C., Lønnum, K., Melinder, A., & Magnussen, S. (2009). Forensic interviews with children in CSA cases: a large-sample study of Norwegian police interviews. *Applied Cognitive Psychology, 23*(7), 999–1011. doi: 10.1002/acp.1534

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*(2), 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*(5), 352–373. doi: 10.1037/h0020071

Vandierendonck, A., & Damme, R. (1988). Schema anticipation in recall: Memory process or report strategy? *Psychological Research, 50*(2), 116–122. doi: 10.1007/BF00309211

Verkampt, F., & Ginet, M. (2010). Variations of the Cognitive Interview: Which one is the most effective in enhancing children's testimonies? *Applied Cognitive Psychology, 24*(9), 1279–1296. doi: 10.1002/acp.1631

Vrij, A., Hope, L., & Fisher, R. P. (2014). Eliciting reliable information in investigative interviews. *Policy Insights from the Behavioral and Brain Sciences, 1*(1), 129–136. doi: 10.1177/2372732214548592

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General, 115*(4), 348–365. doi: 10.1037/0096-3445.115.4.348

Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of confidence. *Journal of Experimental Psychology: Applied, 14*(1), 50–60. doi: 10.1037/1076-898X.14.1.50

Wong, C. K., & Read, J. D. (2011). Positive and negative effects of physical context reinstatement on eyewitness recall and identification. *Applied Cognitive Psychology, 25*(1), 2–11. doi: 10.1002/acp.1605

Wright, A. M., & Holliday, R. E. (2007). Interviewing cognitively impaired older adults: How useful is a Cognitive Interview? *Memory, 15*(1), 17–33. doi: 10.1080/09658210601047351

Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General, 124*(4), 424–432. doi: 10.1037/0096-3445.124.4.424

Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *Journal of Behavioural Decision Making, 10*(1), 21–32. doi: 10.1002/(SICI)1099-0771(199703)10:1<21::AID-BDM243>3.0.CO;2-G

Yaniv, I., Yates, J. F., & Smith, J. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin, 110*(3), 611–617. doi: 10.1037/0033-2909.110.3.611

Zeelenberg, R. (2005). Encoding specificity manipulations do affect retrieval from memory. *Acta Psychologica, 119*(1), 107–121. doi: 10.1016/j.actpsy.2004.12.001