Eyewitness testimonies

The memory and meta-memory effects of retellings and discussions with non-witnesses

Sarwar, Farhan

2011

[Link to publication]

Citation for published version (APA):
Sarwar, F. (2011). *Eyewitness testimonies: The memory and meta-memory effects of retellings and discussions with non-witnesses*. [Doctoral Thesis (compilation), Department of Psychology]. Department of Psychology, Lund University.

Total number of authors:
1

Eyewitness testimonies:
The memory and meta-memory effects of retellings and discussions with non-witnesses

Farhan Sarwar

Department of Psychology

Lund University, Sweden

Dissertation for the degree of
Doctor of Philosophy in Psychology

Academic dissertation which, by due permission of the Faculty of Social Sciences at
Lund University, will be publicly defended on January 28, 2011, at 10:15 in Palaestra,
hörsal övre, Paradisgatan 4, Lund, for the degree of Doctor of Philosophy in Psychology.

Farhan Sarwar

Eyewitness testimonies:
The memory and meta-memory effects of retellings
and discussions with non-witnesses

LUND
UNIVERSITY

In the memory of my mother

Abstract

This thesis investigated the effects of eyewitnesses retellings and discussions with non-witnesses on the eyewitness memory and meta-memory judgments. In *Study I*, the effect of eyewitness discussions with non-witnesses (persons who had not experienced the event) on eyewitness memory and meta-memory realism for the overall information about an event was investigated. The results suggest that discussions of an experienced event may reduce some of the beneficial memory and meta-memory effects caused by mere retellings, but may not have great negative effects compared to a control condition. Analysis of the type of questions asked suggests listeners ask more about the peripheral details as compared with the central details. In a follow-up study to study I conducted a year later participants in the Retell condition no longer showed evidence of the memory and meta-memory benefits evident at the original final test after about 24 days. However, participants in the Retell condition recalled a higher number of correct items than participants in the Control condition. In *Study II*, the effect of eyewitness discussions with non-witnesses on eyewitness memory and meta-memory realism for different types of information was investigated. The different types of information were Forensically central, Forensically peripheral, and Non-forensic information. These are types of information that the police may ask at the beginning of a crime investigation. The results from the two experiments showed that participants had better memory and meta-memory realism for Forensically central and Non-forensic information than for Forensically peripheral information. Moreover, participants in the four conditions were equally capable of distinguishing between correct and incorrect items. Further, in Experiment 1 participants in conditions involving retelling and discussing the event reported more total number and number of correct Forensically central items as compared to the Control condition. *Study III* investigated if retellings and discussions would cause more reminiscence and hypermnesia than mere retellings. The results showed that discussions indeed cause more reminiscence and hypermnesia over the five sessions as compared to mere retellings. The results also showed that the number of times a piece of information was repeated over the sessions was associated with a higher probability for that piece of information being retrieved at the final recall. Interestingly, if the information was retold or discussed in an earlier or later session did not predict if this information would be reported in the testing session or not. Last, the results showed that the forensically peripheral information, but not forensically central information was affected by the reiteration effect (i.e., the effect that confidence tends to increase when a person asserts the same statement many times). This may be due to the fact that the peripheral information was less integrated than the central information.

Svensk Sammanfattning

Ögonvittnens vittnesmål: Effekter på minne och meta-minne av återberättande och diskussioner med icke-vittnen

Farhan Sarwar

Institutionen för psykologi, Lunds Universitet, 2011

Det är vanligt att ögonvittnen återger och diskuterar en upplevd brottshändelse med sin familj och vänner. Syftet med dessa diskussioner är att uppdatera familj och vänner om vad som är nytt. Dessa diskussioner har konsekvenser för ögonvittnets senare minnesrapportering och meta-minnesbedömningar. Med minnesrapportering menas här vad ett ögonvittne kan återerinra sig om den bevittnade händelsen och med meta-minnesbedömningar menas ögonvittnets känslor av säkerhet på att hans/hennes minnen av händelsen är korrekta. Vittnens meta-minnesbedömningar kallas nedan *konfidensbedömningar.* Med uttrycket god realism i meta-minnesbedömningarna menas att nivån på vittnets konfidensbedömningar matchar nivån på korrektheten i vittnets minnesrapportering av den upplevda händelsen. Vittnets minnesrapportering och konfidensbedömningar av sina rapporterade minnen är viktig information i kriminaltekniska sammanhang. Minnet hjälper oss att förstå detaljerna i brottshändelsen och vittnets konfidensbedömningar ger hjälp att förstå korrektheten av dessa minnen. Denna avhandling har undersökt effekterna av att ögonvittnen återberättar och diskuterar (en filminspelning av) en upplevd brottshändelse med icke-vittnen (personer som inte upplevt händelsen) på ögonvittnens minnesrapportering och på deras konfidensbedömningar av de rapporterade minnena.

I Studie I fick deltagarna först se en kort film (c:a 4 minuter) och därefter *fem gånger* över en tre-veckorsperiod uppdelade i tre betingelser antingen enbart fick återberätta händelsen eller både återberätta och diskutera händelsen med icke-vittnen (antingen i en laboratoriemiljö eller med familjemedlemmar och vänner). Lyssnarna var nya personer i var och en av de fem sessionerna. Dessa tre försöksbetingelser jämfördes mot en kontrollbetingelse där dessa aktiviteter inte skedde. Alla deltagarna genomförde en avslutande sjätte *testsession* där vittnena fick instruktionen att återberätta allt vad de kan minnas av den upplevda händelsen (dvs *öppen fri framtagning*) och tre dagar senare ge konfidensbedömningar av de olika ingående elementära minnesutsagorna i de rapporterade minnena. Av intresse i studien var alltså effekten av upprepat återberättande och diskussioner av en upplevd händelse på ögonvittnens

minnesrapportering och på realismen i vittnenas konfidensbedömningar i den avslutande sjätte testsessionen. Resultaten tyder på att diskussioner om en upplevd händelse kan minska några av de positiva effekterna på kvalitén i minnesrapporteringen och på realismen i meta-minne som orsakas av enbart återberättande, men att diskussionerna inte får stora negativa effekter jämfört med en kontrollbetingelse där återberättande och diskussioner av händelsen inte skett. Analys av vilken typ av frågor lyssnarna ställde visade att man frågade mer om de perifera detaljerna i händelsen, jämfört med händelsens centrala detaljer (i första hand handlingar). En uppföljande studie till studie I genomfördes ett år senare. Denna studie visade inga kvarstående tecken på de fördelar på minnesrapportering och meta-minne som deltagarna i den betingelse som enbart återberättat händelsen fem gånger uppvisade i slutsessionen i Studie I efter c:a 24 dagar. Däremot hade deltagarna, i den betingelse som enbart återberättat händelsen, efter ett år ett högre antal korrekta minnesrapporterade utsagor jämfört med deltagarna i kontrollbetingelsen.

Studie II utgick delvis från samma data som i Studie I. Här undersöktes effekten av ögonvittnens diskussioner med icke-vittnen på ögonvittnens minne och meta-minne realism för olika typer av information. De olika typer av information som analyserades var Forensiskt central, Forensiskt perifer och Icke-forensiskt relevant information. De två Forensiskt relevanta informationstyperna är sådan information som det är troligt att polisen kan vilja ha i början av en brottsutredning. Resultaten från de två experimenten i *Studie II* visade att deltagarna hade bättre minne och meta-minne realism för Forensiskt central och för Icke-forensiskt relevant information än för Forensiskt perifer information. Dessutom var deltagarna i de fyra betingelserna i Experiment I (samma fyra betingelser som i *Studie I*) lika kapabla att skilja mellan korrekt och inkorrekt objekt med hjälp av nivån på sina konfidensbedömningar för alla tre informationstyperna. Experiment 1 i Studie II visade också att deltagarna i de betingelser där deltagarna återberättade och diskuterade händelsen rapporterade ett högre antal korrekta Forensiskt centrala minnesutsagor jämfört med kontrollbetingelsen.

Studie III gällde data från två av betingelserna i Studie I, närmare bestämt inspelade data från de fem återgivningsomgångarna i den betingelse där deltagarna *enbart* återberättat händelsen och inspelade data från den betingelse där deltagarna både återberättade och diskuterade händelsen i laboratoriet. Studien visade att deltagarna i den betingelse där deltagarna både återgav och diskuterade händelsen uppvisade mer reminiscens (fler minnesutsagor, både korrekta och inkorrekta över de fem sessionerna) och mer hypermnesi (tillskott av mer korrekta minnesutsagor över de fem sessionerna) än deltagarna i den betingelse där de bara återberättade händelsen fem gånger över tre veckor. Resultaten visade också att antalet gånger en minnesutsaga upprepades under sessionerna var förenat med en högre sannolikhet för att minnesutsagan skulle återges vid den slutliga återgivningen. Däremot hade det, intressant nog, ingen effekt om minnesutsagorna hade återberättats eller diskuterats i en tidigare eller senare session (av de 5 sessionerna) på om minnesutsagan skulle redovisas i den sjätte testsession eller inte. Slutligen visade resultatet i Studie III att de Forensiskt perifera minnesutsagorna, men inte de Forensiskt centrala utsagorna, uppvisade en så kallad *reitereringseffekt* vilken innebär att säkerhetskänslan (dvs upplevd konfidens att utsagan är korrekt) höjs som en effekt av att utsagan upprepas fler gånger. Detta kan bero på att den Forensiskt perifera informationen är mindre välintegrerad än den Forensiskt centrala informationen.

List of Papers

The doctoral thesis is based on the following original papers, which are referred to in the text by roman numerals.

I   Sarwar, F., Allwood, C. M., & Inner-Ker, Å. (2010). Effects of communication with a non-witness on eyewitnesses recall correctness and meta-cognitive realism. Applied Cognitive Psychology. Advance online publication. doi: 10.1002/acp.1749

II   Sarwar, F., Allwood, C. M., & Inner-Ker, Å. (2010). Effects of repeated recall and discussion on eyewitness accuracy and meta-memory realism for different types of forensic information. Unpublished Manuscripts, Lund University, Sweden.

III   Sarwar, F., Allwood, C. M., & Inner-Ker, Å. (2010). Content analysis of eyewitnesses repeated recalls and discussions. Unpublished Manuscripts, Lund University, Sweden.

Acknowledgements

The completion of this dissertation would not have been possible without other people who, in one way or another, inspired me and help me to keep going.

First, I owe my deepest gratitude to Prof. Carl Martin Allwood for his inspiring and candid supervision. I always enjoyed and learned a lot from our casual conversations at the conferences. Carl Martin with his knowledge and enthusiasm improved my understanding of meta-cognition particularly the meta-cognition in eyewitness context. I would also like to show my gratitude to my assistant supervisor, Åse Innes-Ker, for her valuable suggestions and comments. Her feedback always provided me with an opportunity to question my ideas and critically analyze others ideas.

I am indebted to my friends and colleagues Rachel Maddux, Simon Graner, Devin Blair Terhune, Johan Mårtensson, Karin Moesh, Anne-Cecile Treese, Farida Rasulzada, Magnus R. Larsson, and Ilkka Salo for their friendship and support. I also thank Eva Hensiksson, Louise Petersson, Martina Roskvist, Tomy Bauer and Lars Kindberg who have made available their support whenever I needed it.

I am grateful to the Department of Psychology, Lund University for giving me the opportunity to pursue my research studies and provided me with the resources to conduct my research. I am also thankful to my colleagues and friends, at John Jay College of Criminal Justice, New York, Steven D. Penrod, Maria Hartwig, Jennifer E. Dysart, Deryn Strange, Sarah Jordan, and Brian Wallace, Marina Sorochinski, and Rafaele Dumas. They all made my stay at John Jay pleasant and productive. I am also thankful to Prof. Ulf Stridbeck, University of Oslo for the time we shared together at John Jay College of Criminal Justice, New York. I would also like to show my gratitude for my friends, Kristina Myrvold, Ahmad Sher, Umair Tamim, Waqas Rana and Naveed R. Butt, who were always there to help me.

Finally, special thanks to my parents for their support, though distant, that was always with me. I am indebted to Noreen, Rayyan, and Feri for their unconditional support and who always supported me and comforted me with their tender love during the whole journey and never complain about anything.

Contents

Introduction

Eyewitnesses are an important source of information in criminal cases. Frequently, they are the only source of information for the police investigators, lawyers and courts. Although other pieces of evidence (e.g. blood and DNA samples) also provide valuable information about a crime, an eyewitness testimony has a significant role in determining the nature of crime and finding the culprit. Moreover, as Ebbesen and Rienick (1998) pointed out, the police are likely to use information from eyewitnesses because such information is readily available which makes it possible to start the crime investigation and search for the culprit quickly as compared to other time consuming procedures, e.g. collecting blood or other samples from the crime scene and having them analyzed. However, an important problem faced by the criminal justice is how to judge the accuracy of eyewitness statements. In many cases, there is only a single eyewitness to the crime, and thus no independent source of information available to compare his or her statements against (Castelli et al., 2006). One important way that professionals in the criminal justice system (e. g. judges, juries, lawyers, investigators) try to assess the credibility of eyewitness claims is by using the confidence expressed by an eyewitness about his or her claims (Brewer & Burke, 2002; Wells, Lindsay, & Ferguson, 1979). Confidence judgments made by eyewitnesses about their memory statements is a form of meta-memory judgment.

A growing body of research has demonstrated that eyewitness memory and corresponding meta-memory judgments are prone to distortions. A number of factors at the encoding, storage, and retrieval stages contribute to this. Wells (1978) classified the variables that influence the eyewitness memory at each of these stages into two groups, namely *estimator* variables and *system* variables. The essential property of estimator variables (e.g. characteristics of the witness, situational factors, etc.) is that they are not under the control of the criminal justice system. In contrast, system variables (e.g. how to interview the witness, how to construct a lineup, etc.) are, at least to some degree, under the control of the criminal justice system and the handling of these variables can be improved by using appropriate measures (Wells, 1978).

Some examples of system variables that can affect the eyewitness memory during the storage phase are the eyewitness discussions with a co-witness (Hollin & Clifford, 1983; Shaw III, Garven, & Wood, 1997; Yarmey, 1992), the eyewitness exposure to the media coverage of the witnessed event (Loftus & Hoffman, 1989), questions asked by investigators and lawyers (Kebbell & Johnson, 2000; Loftus, 1975; Mark & Shane, 2000; Roebers & Schneider, 2000) and eyewitness discussions with their friends and family. Although most of these factors have been well-studied, this latter factor (discussions with friends and family) has not been subjected to much empirical

scrutiny, outside the present work.

Some of the system variables that can distort the eyewitness confidence judgments during the retrieval stage are: feedback about the correctness or incorrectness of eyewitness statements (Wells & Bradfield, 1998), how many times an eyewitness has reasserted a statement (Hertwig, Gigerenzer, & Hoffrage, 1997), eyewitness personal understanding about how good he or she is in remembering things (Perfect, 2004), and eyewitness discussions with his/her family and friends.

This thesis investigated the impact of eyewitness retellings and discussions with non-witnesses (generally eyewitness family and friends) on their memory and meta-cognitive judgments of the correctness of these memories. Calibration measures were used to study the effects of eyewitness retellings and discussions with non-witnesses on eyewitnesses meta-cognitive realism. The reasons for this are discussed below.

The thesis begins with a brief review of the main empirical findings about the different factors that could affect the eyewitness memory and meta-memory judgments in the case of eyewitness multiple retellings and discussions with non-witnesses. The empirical findings about the impact of different types of forensic information on the memory and confidence judgments are also discussed. Then, the methods used in the studies in the thesis are described followed by brief summaries of the three studies and the short report. Finally, the results of the studies are discussed in relation to the relevant previous research and suggestions are made for future research.

Eyewitnesses' Communications with Non-witnesses

Eyewitnesses tend to discuss the experienced events with their family and friends. The main purpose of discussing the experienced events may be to update their family and friends about whats new (Skowronski & Walker, 2004). Eyewitnesses usually engage in such discussions multiple times before they testify in court (Paterson & Kemp, 2006). Other research has found that in general the frequency of discussions about a tragic incident is high immediately after the event and this frequency of discussions decreases with the passage of time (Pennebaker & Harber, 1993). This conclusion may be applicable also to the events experienced by eyewitnesses.

Discussions vs. Retellings

In discussions people repeat and discuss the details of the witnessed event. Since repetition is known to improve the memory of the repeated content, one could also assume that discussing the details of a witnessed event would also improve memory of the discussed details of an event. However, this need not be the case since retellings and discussions are two different phenomena.

Retellings

In retellings one simply tells something to other person/s multiple times, but the other person/s does not contribute anything either in the form of questions or comments. As elaborated below, multiple retellings can have multiple effects on the memory of an experienced event. For example, retellings are similar to test taking, where active repetitions of the learned material occur without accessing the original study

material. Such active repetition has usually been found to improve the accuracy of the repeated content in later recall (Roediger III & Karpicke, 2006a). This phenomenon is called the *testing effect* (Cull, 2000; Roediger III & Karpicke, 2006a, 2006b). It should be noted that retellings are different from rehearsals because in rehearsals the content is repeated with access to the original content (Roediger III & Karpicke, 2006a). Although rehearsals are known to improve the memory of the rehearsed content, research results show that retelling of material results in better performance on the final test as compared with rehearsing the same content the same number of times (Karpicke & Roediger III, 2007, 2008). Moreover, the testing effect has also been shown to improve the memory of content that is related to the tested content also when the participants were not tested for that content (Chan, McDermott, & Roediger III, 2006).

Successive retrieval attempts may result in the recall of new information from the original information that was not recalled during the earlier retrieval occasions. This phenomenon is referred to as *reminiscence* or *spontaneous recovery* in the research literature (e. g. La Rooy, Pipe, & Murray, 2005; Payne, 1987; Turtle & Yuille, 1994). By definition retrieval of both new correct and new incorrect information is considered reminiscence. A number of studies have been consistent in showing support for reminiscence (e. g. La Rooy, Pipe, & Murray, 2007; Scrivner & Safer, 1988; Turtle & Yuille, 1994).

The amount of *correct information* recalled may also increase with each retrieval attempt. This phenomenon is referred to as *hypermnesia* (Mulligan, 2001; Payne, 1987; Roediger III, Jacoby, & McDermott, 1996). Hypermnesia depends both on the recall of previously unrecalled information and the recall of previously recalled information (La Rooy et al., 2005; Payne, 1987). Empirical support for hypermnesia has been inconsistent across studies. A commonly used method for studying hypermnesia, developed by Erdelyi and Becker (1974) exposes participants to pictures and words in a learning phase. In a subsequent testing phase, participants are asked to recall a fixed number of pictures and words. Participants are allowed to guess. Experiments using this paradigm tend to find support for hypermnesia (e g. Henkel, 2004; Scrivner & Safer, 1988). Taking advantage of an actual event (the O.J. Simpson case), Bluck, Levine, and Laulhere (1999) found evidence for hypermnesia. Participants were interviewed three times with the first interview taking place roughly eight months after the televised verdict. The results showed that both the amount of information and the amount of correct information recalled increased from first interview to third interview. However, there are studies using similar methods to Bluck et al. (1999) that has not found support for hypermnesia (La Rooy et al., 2007; La Rooy et al., 2005). In addition, Turtle and Yuille (1994) in Experiment 1 used free recall and focused questions. The results showed no support for hypermnesia. Ebbesen and Rienick (1998) tested the participants for the descriptive details about the individuals in the event (e.g. height, weight) and found what they called the *freezing effect.* By this they meant that retrieval attempts stopped forgetting, but did not cause hypermnesia. Ebbesen and Rienick (1998) suggested that the freezing effect and hypermnesia might be related to each other because both act to protect the memory.

Increase in recall may also cause people to retrieve incorrect information together with the correct information (Ceci, Huffman, Smith, & Loftus, 1994; Henkel, 2004, 2007; Roediger III et al., 1996; Scrivner & Safer, 1988). Thus multiple retrieval

attempts may also cause people to include incorrect information in the memory of an event. This is most likely to occur when people engage in discussions. The impact of discussions on memory is discussed below.

Multiple retellings may also hinder the retrieval of information because of *retrieval induced forgetting* (Coman, Manier, & Hirst, 2009; MacLeod, 2002). According to the retrieval induced forgetting hypothesis, if a person fails to recall some information in the first recall attempt after learning the material the individual will most likely not be able to recall it in the subsequent recalls as well (MacLeod, 2002). The method used to study retrieval induced forgetting has three stages, namely learning, retrieval, and testing. In the learning stage participants study lists of categories (e.g. fruit: banana, apple). In the retrieval stage participants recall half of the items from each category. In the testing phase participants memory is tested for all items. The results show that the items that were retrieved during the retrieval stage were more often recalled at the final test than the items that were not retrieved during the retrieval stage (Anderson, Bjork, & Bjork, 1994). The basic reason assumed for retrieval induced forgetting is inhibition of non-retrieved items in favor of retrieved items. Many experimental findings support this explanation (e.g. Anderson et al., 1994; Erdelyi, 2010; Storm, Bjork, & Bjork, 2007).

Retrieval induced forgetting does not lead to a permanent loss of information, nor does it lead to a weakness in the storage strength (Anderson et al., 1994; Storm, Ligon Bjork, & Bjork, 2008). The expression *storage strength* refers to how well connected a piece of information is with other relevant information in the memory (Anderson et al., 1994; Storm et al., 2008). Instead retrieval induced forgetting is only a temporary unavailability of the information because of the weakness in retrieval strength as compared to the information that was retrieved (Anderson et al., 1994; Storm et al., 2008). The expression *retrieval strength* refers to how accessible a piece of information is in response to a given cue at a certain point of time.

Retrieval induced forgetting is dependent on how well the information is integrated with other information. If the encoded content is well-integrated then inhibition is less likely to occur and there may be less or no retrieval induced forgetting (Anderson, 2003; Anderson & McCulloch, 1999).

Anderson (2003) defines *well-integrated* by noting that items in a category are well-integrated (or at least better integrated) when they are associated with a cue other then the cue used to link them together in an experiment, compared to items not having such an association. For example, the items used in a category of fruit can be associated with other cues as well besides the keyword fruit used to represent them in the experiment. A participant can link the item apple to a famous proverb an apple a day keeps the doctor away. Lemon can be associated with lemonade and so on. Thus, for Anderson, for an item to be better integrated than another item appears to mean that the item in question has more associations.

In this thesis it is suggested that, in event memory, items that are part of a sequence of action details in an event can be seen as well-integrated. The reason is that information describing an action is likely to be associated with information describing other actions in the same action sequence and thus to have a high probability to be activated as an effect of earlier items in the sequence being activated. Thus well-integrated is here taken to mean to have a *strong* (reliable) association with other items.

4

Not all researchers agree with the inhibitory explanation for retrieval induced forgetting (e.g. Butler, Williams, Zacks, & Maki, 2001; Perfect et al., 2004) According to Perfect et al. (2004) retrieval induced forgetting is a context-specific phenomenon. Retrieval of some specific information at a given time depends on the context of retrieval. Alternatively, according to Butler et al. (2001) it is the cue attached to a category that causes retrieval induced forgetting for the non-retrieved items in that category. For example, when Fruit is a common cue attached to all the items in this category (e.g. Fruit: apple, orange, banana) retrieval induced forgetting occurs, but when each item is assigned a specific cue (different cue for each item) there is no retrieval induced forgetting.

Interestingly, the information that is forgotten because of the retrieval induced forgetting can be relearned even faster than the information that was retold at the retrieval stage (Storm et al., 2008).

Discussions

Discussion is an interactive process between two or more people. In discussions the listeners not only ask questions, but also contribute their opinion. In forensic situations an eyewitness is likely to engage in discussions about the experienced event with different people with a variety of interests. Characteristics of both the teller and the listener as well as the context of their discussion determine what to share and how to respond to the tellers event descriptions (Pasupathi, 2001). In response to the eyewitness description of the experienced event listeners may also share their similar personal experiences (Loftus, 2003) and communicate incomplete and misleading information (Loftus, 1979).

There is much research evidence that shows that information supplied by others can distort eyewitness memory (Garcia-Bajos, Migueles, & Anderson, 2009; Loftus, 1992; Nourkova, Bernstein, & Loftus, 2004; Wright, Self, & Justice, 2000). It appears that it often is the witnesses that initially introduce incomplete and misleading information when discussing the event with other individuals and thereby make the listener mention it (Alper, Buckhout, Chern, Harwood, & Slomovits, 1976; Gabbert, Memon, & Allan, 2003; Hollin & Clifford, 1983; Luus & Wells, 1994; Marsh, 2007; Tversky & Marsh, 2000). A reason is that when witnesses forget details of the witnessed event they may compensate the missing memories with the memories from their previous listeners recollections (Wright, Mathews, & Skagerberg, 2005). During discussions people sometimes also deliberate about the speculative contra-factual possibilities, and such discussions can later affect the eyewitness memories of the discussed event (Wells & Gavanski, 1989). The feedback people receive from their discussion partners can thus cause people to make incorrect judgments about the different details of the witnessed event. This phenomenon is referred to as the *ripple effect* (Pizarro, Laney, Morris, & Loftus, 2006).

The listeners, for their better understanding and clarity, ask questions about different aspects of the forensic event. Such questions can be leading, misleading or confusing and can cause distortions and deterioration in the eyewitness memory (Kebbell & Johnson, 2000; Loftus, 1975). These distortions can be particularly strong if the witness does not realize the difference between the memories of the experienced event and the contents of the questions. Moreover, discussing an event is likely to activate memory schemas that represent how similar events normally occur (Tversky &

Marsh, 2000). During later recall, lost information may then be replaced by information from the event schema that was activated at the time of the earlier discussion. The effect could be that the memory schema contributes information which was not in the original event. In addition, the memory schema may block out experienced information which is inconsistent with the contents of the schema (Marsh, 2007; Tversky & Marsh, 2000). Similar schema dependent effects may also happen in the context of mere retellings. However, this is likely to only happen to a lesser extent because no input from outside is involved.

Discussions with Non-witnesses: How does this differ from Co-witness collaboration and testifying in court?

The nature of an eyewitness discussions with non-witnesses (e.g. family and friends) is different from the nature of eyewitness discussions with the co-witnesses and with the police investigators, lawyers, and judges. The reason is that speakers keep the listeners interest in mind and make their stories relevant for the listeners (Russell & Schober, 1999).

When eyewitnesses share the witnessed event with their family and friends they may lower their certainty criterion for reporting and provide their more free reactions and conclusions (Koriat & Goldsmith, 1994, 1996; Roebers, Moga, & Schneider, 2001). As a result eyewitnesses may report more incorrect details about the witnessed event to their family and friends as compared with the people in the criminal justice system. Moreover, when eyewitnesses repeat incorrect details over multiple discussions it may become impossible for the eyewitnesses to distinguish between incorrect and correct details of the witnessed event (Loftus, 1983). Family and friends may also react by giving their subjective analyses of the eyewitness description of the forensic event and may also contribute their personal similar experience to the discussion which may then be incorporated in the witness narrative (Dritschel, 1991).

The situation when eyewitnesses share the witnessed event with co-witnesses is somewhat different. Here the communication often may be like a collaboration process where both witnesses influence each other in the process of comparing their information with each other and try to complete their collection of the information about the witnessed event (Gabbert et al., 2003; Wright et al., 2000). As a result, co-witnesses can affect each other immediately (Shaw III et al., 1997). Eyewitnesses have been shown to be influenced by the co-witnesses both if the co-witnesses communicate with each other directly (Shaw III et al., 1997) or if the co-witnesses' statements are presented to them through another person (Garven, Wood, & Malpass, 2000). The phenomenon that co-witnesses influence each others memory is referred to as social conformity in eyewitness literature (Wright et al., 2000). Roediger, Meade, and Bergman (2001) called this phenomenon *social contagion of memory*. Interestingly, longer intervals between witnessing a crime and discussing it with a co-witness make the witnesses more susceptible to the incorrect information supplied by the co-witness (Garcia-Bajos et al., 2009). Moreover, if eyewitnesses are acquainted they may incorporate their co-witness account into their own recall to a greater degree than co-witnesses that are not previously acquainted (Hope, Ost, Gabbert, Healey, & Lenton, 2008). Eyewitnesses are also more likely to incorporate information from a co-witness who is more confident and had a better exposure to the forensic event

(Wright et al., 2000).

According to Koriat and Goldsmith (1994) and Tetlock (1983a) when eyewitnesses share the witnessed events with the people in the criminal justice system (e.g. investigators, lawyers, or judges) they may tend to share only such information that they feel is absolutely true. The reason could be that they are quite aware of the possible impact of their testimony.

Different Types of Forensic Information

After a crime eyewitnesses will be expected to answer questions posed to them by the police. The police is first likely to ask the eyewitnesses to describe the witnessed event in as much detail as possible, and the police naturally expect the eyewitnesses to provide them with as accurate details as possible of what happened, how it happened, description of offenders (e.g. age, height, body type, special features, dress), description of objects used (e.g. weapon, vehicle), and the time and place of incidence. In this context, it is important to know how satisfactorily eyewitnesses can provide information relevant to such questions in their first free recall. This is also important because the police will ask further probing questions on the basis of information received in the first eyewitness report. Incorrect details provided by the eyewitnesses in the first report can lead the crime investigation in the wrong direction. In this thesis eyewitnesses free recalls are analyzed to investigate the eyewitness memory and meta-memory judgments for the different types of information that could provide answer to the police questions. According to the authors knowledge, so far eyewitnesses open free recalls have not been subject to empirical investigation for their potential to provide answers to the police questions and corresponding meta-memory judgments.

Classifying the Eyewitness' Statements

In eyewitness research the eyewitness information has commonly been categorized into central and peripheral details. However, there is a lack of consensus among researchers about what information should be categorized into central and peripheral categories. In spite of this almost all the researchers use the terms central and peripheral information when they divide the information into information that they consider important/essential/central or less important/not essential/peripheral. There is consensus in the eyewitness research literature that eyewitnesses remember the central information from the forensic event better as compared with the peripheral information (Christianson & Loftus, 1987, 1991; Heath & Erickson, 1998; Parker & Carranza, 1989; Roebers & Schneider, 2000; Wessel & Merckelbach, 1997). This has been found to be true for children as well (Hershkowitz & Terner, 2007; Memon & Vartoukian, 1996; Saywitz, Goodman, Nicholas, & Moan, 1991).

Although there is consensus that central information is better recalled than peripheral information, there is less of a consensus on what information constitutes central information, and what information should be considered peripheral. This has led to the somewhat paradoxical situation where information, such as the color of a suspects or victims hair or shirt, that is considered central within one classification scheme (for example, Brown, 2003; Christianson, 1992; Christianson & Loftus, 1987, 1991; Memon & Vartoukian, 1996; Parker & Carranza, 1989; Wessel & Merckelbach,

1997), is regarded as peripheral information by another group of researchers (for example, Candel, Merckelbach, Jelicic, Limpens, & Widdershoven, 2004; Hershkowitz & Terner, 2007; Heuer & Reisberg, 1990; Orbach et al., 2000; Roebers et al., 2001; Roebers & Schneider, 2000). It is surprising that both groups agree that central information is remembered best while sometimes categorizing the same information into the opposite groups. It is not the aim of the present work to reconcile these different views, but instead to review the different types of classification systems, and select the system that seems best fit to answer the polices questions.

The classification models used by the different researchers to divide the eyewitness reports into forensically central and forensically peripheral information can be broadly sorted into four types of models: the *Visual attention model*, the *Plot relevancy model*, the *Mixed model*, and the *Empirically based model*. Below, the four models are described, and each model is discussed in terms of its use to divide the eyewitness statements into categories that could be used to assess the eyewitness' ability to answer the police questions.

The Visual attention model

The idea behind the Visual attention model is Easterbrook's (1959) hypothesis claiming that arousal results in the narrowing of attention. As a result, people increase the processing of central information but at the cost of neglecting to process peripheral information. According to this model information that is at the focus of attention, or is the source of arousal is considered central, for example, the gist of the event and its central details (for example the color of the shirt or height of the suspect). In contrast, the information that is not at the focus of attention or is not the source of arousal (for example, a car parked on the other side of the street) is considered peripheral (Brown, 2003; Christianson, 1992; Christianson & Loftus, 1987, 1991; Easterbrook, 1959; Parker & Carranza, 1989; Parker, Haverfield, & Baker-Thomas, 1986; Vandermaas, Hess, & Baker-Ward, 1993; Wessel & Merckelbach, 1997).

It is not possible to understand the eyewitness capacity to answer the polices questions by using the Visual attention model because no distinction is made between action details and descriptive details. In the Visual attention model the distinction between central and peripheral information concerns if the information was at the focus of attention or not. Since the Visual attention model allows that both the action and descriptive details can be present in both the central and the peripheral category it is hard to know if arousal will facilitate the memory for action or descriptive details. In addition, the empirical support for the Visual attention model has been mixed. Christianson and Loftus (1987, 1991) in two separate studies found support for the Visual attention model. In their 1987 study, they tested the memory for traumatic and non-traumatic events by showing emotional and neutral slides to the participants. The results indicated that traumatic events were better remembered. Moreover, the information regarding the source of arousal in a traumatic event (central information) was even better remembered than the information that was not the source of arousal (peripheral information). In their 1991 study, they showed a thematic series of slides to the participants. The pictures in the series were identical, except for the critical slide, which was either emotional (a woman injured near a bicycle) or neutral (a woman riding a bicycle) depending on condition. The results showed that the participants remembered the central details better than the peripheral details if

the critical slide was emotional. In contrast, Wessel and Merckelbach, (1997) used a spider phobic group as compared to a control group. The idea was that the presence of a spider in the environment would make the spider phobic group focus their attention on the spider and consequently show better memory for the central details as compared to the control group who would not react to the spider. The results showed no improvement in the memory of central details for the spider phobic group as compared with the control group. However, the phobic participants provided fewer peripheral details as compared with the control group. In brief, the findings regarding the Visual attention model are inconsistent.

The Plot relevancy model

According to this model, information or facts related to the event that cannot be changed without changing the storyline in the event is regarded as central, for example, the suspect put a gun to the victims head. Information or facts that can be changed without changing the story, for example, the suspect was wearing a blue shirt, is considered peripheral (Heuer & Reisberg, 1990). Other researchers have also used this model (e. g. Candel et al., 2004; Hershkowitz & Terner, 2007; Orbach et al., 2000; Roebers et al., 2001; Roebers & Schneider, 2000).

The Plot relevancy model divides the information into action details and descriptive details. The argument for this assertion is that only change of action details causes alteration in the story while change in descriptive information does not cause alteration in the story. Division of the eyewitness statements by using the Plot relevancy model can be helpful to assess the eyewitness ability to answer the polices questions, since the action details determine what happened and how it happened while the descriptive details provide information about the suspect, objects used etc.

Studies using the Plot relevancy model have shown that people are better at remembering action details than descriptive details (Ibabe & Sporer, 2004; Roebers & Schneider, 2000; Yuille & Cutshall, 1986). Results from studies using the Plot relevancy model are also in line with the results of research showing that people are better at describing actions than descriptions (Migueles & Garcia-Bajos, 1999). The Plot relevancy model also has support from other research results that show that when there is a moving stimulus the descriptive details become background to the moving stimulus and the moving stimulus becomes the focus of attention (Ramachandran & Anstis, 1986; Rock & Palmer, 1990).

Mixed models

In mixed models researchers have proposed more comprehensive models to divide the information into different categories. In these models researchers use two or more criteria to sort out the information. For example, Burke, Heuer, and Reisberg (1992) divided the focused questions about their stimulus slides in two stages. In the first stage the information was divided into central and peripheral information by following the classification of the Plot relevancy model (Heuer & Reisberg, 1990). In the second stage the central information was divided into gist and basic level visual information about the slides. Gist was a basic level information about the persons and things in the slides, e. g. the father was a doctor, while the basic level visual information was about the specific actions shown in a slide, e.g. the father was talking to a policeman

9

(Burke et al., 1992). The peripheral information was further divided into questions regarding central details and background details by following the Visual attention model (Christianson & Loftus, 1991).

Ibabe and Sporer (2004) also divided the event information into sublevels using different criteria. At first all information was divided into actions and descriptive details and then these two types of information were subdivided into their respective central and peripheral categories according to the Visual attention model.

Interestingly, regardless of how the information was classified the general findings from the mixed models are in line with the common findings described above, that participants are more accurate on central details as compared to the peripheral details (Burke et al., 1992; Ibabe & Sporer, 2004).

Empirically based models

Some researchers have classified central and peripheral information in more empirically based ways. The basis for these attempts is what information people generally *consider* central (important) and peripheral (unimportant). Heath and Erickson (1998) asked the adults in their study to rate the importance of the actions and props in a story on a 6-point scale where 1 meant very peripheral and 6 meant very central. Memon and Vartoukian (1996) asked students to list as many details as they remembered from a witnessed event. The items mentioned by four or more people were considered central and the items mentioned by less than four participants were considered peripheral. Roberts and Higham (2002) used four police officers and one crown counsel to classify the information of the stimulus event used into correct relevant, correct peripheral, errors, and confabulations. Saywitz, Goodman, Nicholas, and Moan (1991) asked five judges to rate childrens reports on a 5-point scale ranging from 1 (very central) to 4 (very peripheral). Items with mean ratings below 3 were considered central and items with a mean rating above 2.9 were considered peripheral. It is not clear from this study what instructions were given to the judges or whether they were asked to use some specific criteria or not.

A common finding in studies using empirically based models is that people remember central information (forensically relevant information) better as compared with peripheral information (forensically unimportant information) (Heath & Erickson, 1998; Memon & Vartoukian, 1996; Roberts & Higham, 2002; Saywitz et al., 1991). One problem with this approach is that the various criteria used in the different versions of the empirical approach do not provide a stable ground for classification. Another problem is that only dividing the information into relevant and irrelevant categories may not be useful in an applied context. The knowledge of relevant and irrelevant may not be very helpful for researchers in trying to understand witnesses ability to answer the different types of questions asked by the police. An important reason for this is that no distinction is made between action and descriptive details. In brief, empirical approaches may not be helpful in an applied context and for answering the questions posed in this dissertation.

Classification model used in this dissertation

In this dissertation the plot relevancy model by Heurer and Reisberg (1990) is used but with modification. This model is useful since it divides the information into

action details and descriptive details. Since most of the questions that the police have belong to either the action details (what happened and how it happened) or descriptive details (e.g. color of the offenders cloths) this division will help to test the eyewitnesss capacity to answer questions from these two categories.

In order to separate out the irrelevant information, the participants statements are at the first stage divided into the forensic and non-forensic statements. Non-forensic information such as houses, roads, surroundings, etc., is either not needed to solve the crime or this information can be collected from the crime scene without the help of an eyewitness. At the second stage, the forensic statements are divided into forensically central and forensically peripheral categories by following the plot relevancy model presented by Heurer and Reisberg (1990).

As noted above it is surprising that the research findings about forensically central and peripheral findings have been consistent regardless of the fact of how these two types of information were defined. That is why in the following discussion I summarize the findings about central and peripheral information without going into the details of how the different types of information were defined.

Quantity and Quality of Central vs. Peripheral Information

Quantity of information refers to the total amount of correct and incorrect information reported by the eyewitnesses. The quality of information is referred to in the research literature as the *accuracy*, which is the proportion of correct information reported of all reported information. Peoples memory of central and peripheral information can be compared by looking at the total amount of information recalled in each category and the mean accuracy of correct information recalled in each category.

As noted above, research shows that people recall more central information as compared to the peripheral information (see for example, Roberts & Higham, 2002; Roebers et al., 2001; Wessel & Merckelbach, 1997). It is true for imagined events as well. For example Jelicic et al. (2006) asked the participants if they had seen the non-existent video footage of the murder of Dutch politician Pim Fortuyn, 63% of the participants said that they had seen the non-existing footage while only 23% could provide peripheral details of the event. Similar results were found by Riniolo et al. (2003) by studying peoples memory of the Titanics final plunge.

With respect to the *accuracy* of the forensically central and forensically peripheral information, as mentioned above, adults are more accurate about information that is forensically central to an event than the information that is peripheral to that event (See for example, Heath & Erickson, 1998; Ibabe & Sporer, 2004; Wessel & Merckelbach, 1997). This is also the case for children (Hershkowitz & Terner, 2007; Memon & Vartoukian, 1996; Saywitz et al., 1991). Moreover, younger children (4-5 yrs) perform poorer on peripheral items as compared to older children (7-8 yrs) (Vandermaas et al., 1993).

Misinformation in Central and Peripheral information

When people are provided with misleading information about an event it may cause a memory alteration of that event. This phenomenon is called the *misinformation effect* and it can distort the memory of an event (Allen & Lindsay, 1998; Loftus, 1979, 2003; Zaragoza & Lane, 1994). People can even include misinformation into the

details of the experienced event from details belonging to a misinformation event in spite of the fact that they are aware that the two events are different. For example, Allen and Lindsay (1998) found that participants included details from the irrelevant post-event narrative into the memory of the details of the stimulus event. This makes it even more relevant in forensic situations to know which kind of information is more vulnerable to misinformation.

Studies involving both adult participants (Heath & Erickson, 1998) and child participants (Candel et al., 2004; Roebers & Schneider, 2000) show a greater misinformation effect for peripheral information than for central information. Moreover, in the absence of any misinformation people tend to fill the memory gaps in central information with their own imagination (Erskine, Markham, & Howie, 2001; Greenberg, Westcott, & Baily, 1998).

People also have a tendency to distort the descriptive details of an event because of the ripple effect described above (Pizarro et al., 2006). People may also include misinformation into the memory for the descriptive details of an experienced event because of biased retellings (Tversky & Marsh, 2000). The reasons according to Tversky and Marsh (2000) are that people elaborate the event details in line with the retelling specific information and the reorganization of schema that was created during the retellings. Moreover, eyewitness divided attention during experiencing the criminal event may not allow proper encoding of the event details. This makes eyewitnesses vulnerable to suggestions (Lane, 2006).

Emotion and Central vs. Peripheral Information

Forensic situations are emotional in nature and it is important to understand how emotional arousal affects the memory of an event. However, the research literature shows conflicting results about the impact of emotions on the kind of information remembered better. The Easterbrook hypothesis (1959) claims that arousal results in the narrowing of attention. As a result during arousal people increase the processing of information that is the source of arousal but at the cost of neglecting the processing of peripheral information. An excellent example of attention narrowing is the weapon focus effect where victims focus most of their attention on the weapon and neglect the other information about the perpetrator. Many laboratory studies have successfully demonstrated the weapon focus effect (See for example, Loftus, 1979; Loftus, Loftus, & Messo, 1987). Christianson (1992) and Christianson and Loftus (1987, 1991) found support for the attention narrowing hypothesis. Their results showed that the participants had better memory of the central information in the emotional slide as compared to the peripheral information. Although the classification used by Burke et al. (1992) was complex, their results did support enhanced performance on central information (gist and basic level visual information). Further, participants in the arousal condition performed worse on peripheral information (details not associated with the events main theme) as compared to the controls.

In contrast, other researchers (for example, Heuer & Reisberg, 1990) have not found support for the attention narrowing hypothesis. The results reported by Heuer and Reisberg (1990) showed that arousal facilitates memory of both central and peripheral details. Wessel and Merckelbach, (1997) found partial support for the attention narrowing hypothesis. Their results show no improvement in the memory of

12

central details for a spider phobic group as compared to a control group. Moreover, phobic participants provided few peripheral details as compared to the controls, but they also discuss that it could be due to their definitions of central and peripheral information.

Confidence as a Regulator of Memory Report

Koriat and Goldsmith (1996) proposed and empirically tested a *model of free report monitoring and control* by combining traditional signal detection theory with meta-memory theory. This model describes the regulation of both the quantity and memory accuracy of memory report by attending to both retrieval and monitoring processes. The monitoring and control processes are based on three factors: 1) *Monitoring effectiveness*, 2) *Control sensitivity*, and 3) *Response criteria*. 1) *Monitoring effectiveness* shows the competency of the retrieval system in identifying the correct and incorrect answers. 2) *Control sensitivity* shows the subjective control on what answer to share and what answer not to share on the basis of effective monitoring of the retrieved information. 3) *Response criteria* refers to the stakes involved in sharing the correct and incorrect answers (Koriat & Goldsmith, 1996). The *Control sensitivity*, in other words strategic control, is a function of monitoring output and the stakes involved (Koriat & Goldsmith, 1996). Moreover, the strategic control is based on an implicit confidence judgment about the level of correctness of a requested piece of information and the demand criterion. The empirical findings from the two experiments indeed support their model by showing that the participants were able to improve their quantity and accuracy of the information shared by using strategic control according to the stakes involved. Moreover, participant confidence accuracy correlation was high for the free report condition in contrast to the forced report condition. The results also showed that the participants confidence can be a better indicator of correctness under a free report option than a forced report option (Koriat & Goldsmith, 1996). For discussion of a similar *model of free report monitoring and control* see Blank (2009).

Realism in Confidence Judgments

Whenever memories are shared they can be assumed to have been confidence judged as part of the process of reporting them (Kelley & Lindsay, 1993; Koriat & Goldsmith, 1996; Shaw & McClure, 1996). The degree of Realism in confidence judgments is a function of the relationship between accuracy and confidence level (Allwood, 2010; Yates, 1994).

Surveys show that police, prosecuting and defense attorneys and jury-eligible samples consider eyewitness confidence as an important indicator of eyewitness accuracy (Brewer, Potter, Fisher, Bond, & Luszcz, 1999). Moreover, people have at least some strategic control over what to report and when to report based on the level of accuracy needed (Brainerd, Wright, Reyna, & Payne, 2002; Koriat & Goldsmith, 1996).

As noted above, a number of factors at the encoding, storage, and retrieval stages can distort the eyewitness confidence judgments. Examples of such factors are positive or negative feedback about the accuracy of the eyewitness statements (Wells & Bradfield, 1998), the eyewitness personal understanding about how good he is at

remembering things (Perfect, 2004), and the effects of retellings and discussions on realism in confidence judgments.

Effect of Retellings and Discussions on Realism in Confidence Judgments

In the criminal justice system eyewitnesses repeat and discuss the witnessed event many times. For example, according to Christianson (1994) in the US witnesses describe the witnessed event, on average, 11 times to different people. Since the different listeners have different interests, the speakers will adjust what they say to make their stories relevant to their audience. Just like the retellings and discussions have consequences for the quality of the memory, they also have consequences for confidence. In this context two factors, the *reiteration effect* (Hertwig et al., 1997), and the *accountability effect* (Tetlock, 1983a), may play an important role in tempering the realism in confidence judgments.

The Reiteration effect

Repeating a witnessed event multiple times may lead to increased feelings of confidence without any improvement in the memory accuracy. This phenomenon is known as the *reiteration effect* (Hertwig et al., 1997), that is, the effect that confidence tends to increase when a person asserts a statement many times. In line with this, other research has shown that when a witness is questioned many times confidence tends to increase (Shaw, 1996; Shaw & McClure, 1996; but see Granhag, 1997).

The reason for the *reiteration effect* is likely to be changes in retrieval fluency (Shaw & McClure, 1996). According to the *retrieval fluency* hypothesis, when a piece of information is retrieved multiple times it makes that piece of information more readily available when needed (Anderson et al., 1994). Because the feeling of confidence is partially based on how easily a memory is accessed, multiple retrievals may cause the confidence judgments for that information to be inflated without any change in the corresponding accuracy. As just noted, the reason for this may be the increase in *retrieval fluency* experienced by the individual as an effect of multiple retrievals (Shaw, 1996; Shaw & McClure, 1996).

The findings regarding the reiteration effect have been mixed. Shaws studies (e.g. Shaw, 1996; Shaw & McClure, 1996; Shaw, McClure, & Dykstra, 2007) show that confidence tends to increase if information is retrieved multiple times. Note that they used focused questions. In contrast, studies by Granhag and colleagues (Granhag, 1997; Granhag, Stromwall, & Allwood, 2000) show no increase in confidence with repeated confidence assertions of the previous answers to memory questions presented in print. One possible explanation for these conflicting results is how the memory report was revisited (Shaw et al., 2007). The reason may be that information was repeated differently in the studies by Shaw et al. and Granhag et al. In the studies by Shaw et al. the participants were questioned about the same information multiple times. In the studies by Granhag et al. the participants answered the focused questions once but were later presented with their previous answers and were asked to give their confidence judgments. Thus *active* reassertion of the statement may be needed for the reiteration effect to occur.

Accountability effect

Research also shows that confidence may be tempered in social contexts due to the *accountability effect* (Tetlock, 1983b). Specifically, people may lower their confidence when they consider that they will be held accountable for the correctness of their statements by other persons, for example, when testifying in court. Jermias (2006) studied the accountability effect for managers and found that if people are made accountable for their decisions they tend to show underconfidence in their decisions and vice versa. Note that this is very much in line with Koriat and Goldsmiths model (1996). The implications of these findings suggest that making people realize the possible consequences of their testimony may help to control the reiteration effect or at least its consequences.

Measuring Realism in Confidence Judgments

In eyewitness research the confidence-accuracy (CA) relationship is traditionally measured by the point bi-serial correlation, especially in research on lineups. Early results from these studies showed a weak CA relationship (Bothwell, Deffenbacher, & Brigham, 1987; Luus & Wells, 1994; Sporer, Penrod, Read, & Cutler, 1995; Wells, 1993). However, other researchers have noted weaknesses in this method as an indicator of realism in confidence judgments. One reason is, as pointed out by Juslin, Olsson and Winman (1996), that the correlation size in a partly non-relevant way depends on the spread of the confidence judgments over the total confidence judgment scale. Further, as also noted by Juslin et al. (1996) the confidence-accuracy correlation measure primarily picks up witnesses ability to discriminate correct from incorrect reports by means of their confidence judgments in contrast to the witnesses tendency to be over- or underconfident (see, Brewer, 2006; Brewer & Burke, 2002; Brewer & Wells, 2006). These and other researchers (e.g., Weingardt, Leonesio, & Loftus, 1994; Wells, Olson, & Charman, 2002) instead recommend the use of calibration methodology which gives a more differentiated and informative understanding of realism in confidence judgments.

The Calibration approach separates the issue of realism in confidence judgments into various aspects of realism (Yates, 1994). Two such aspects are *bias*, and *separation* (Yates, 1994). *Bias* refers to the correspondence between confidence and accuracy and can be measured by measures such as calibration and over/-underconfidence. The calibration measure punishes deviation from perfect realism at each confidence level whereas the over-/underconfidence measure shows the average degree of deviation between confidence and accuracy over all confidence levels. *Separation* refers to the eyewitness' ability to separate correct and incorrect items by means of the level of their confidence judgments. *Separation ability* can be measured by the slope measure. These measures are explained in more detail in the methodology section.

Researchers have pointed out that a problem in using calibration measures is that a large amount of data is needed to get reliable values for the measures (e.g, Brewer & Wells, 2006). This can be a problem in a within-subject design where sample size is small. When the sample size is small each participant can sometimes be asked to give many confidence judgments.

Realism in Confidence Judgments: Central vs. Peripheral Information

There appear to be only a few studies on the realism in confidence judgment for forensically central and peripheral information (e.g. Ibabe & Sporer, 2004; Migueles & Garcia-Bajos, 1999; Roberts & Higham, 2002). In the reported studies the mean confidence and accuracy levels are compared with the mean accuracy scores, that is, over-/underconfidence, is reported. Focused questions were used when probing the participants memory. The results from these studies show that participants assigned higher confidence judgments to the forensically central information as compared with the forensically peripheral information (Ibabe & Sporer, 2004; Migueles & Garcia-Bajos, 1999; Roberts & Higham, 2002). It is hard to draw any conclusions from these results because these studies used different criteria to divide the information into central and peripheral categories. For example, Ibabe and Sporer (2004), and Migueles and Garcia-Bajos, used the plot relevancy model to divide the information into central and peripheral categories. In contrast, Roberts and Higham (2002), used the empirically based approach to distinguish between central and peripheral information (for detailed description of the Plot relevancy model and the Empirical based model see above). Since the different criteria allocate different types of information into central and peripheral categories it is difficult to know the participants precise confidence levels for different types of information. Furthermore, the studies by Ibabe and Sporer (2004), and Robert and Higham (2002) and Migueles and Garcia-Bajos (1999) only used different versions of focused questions, not reports under open free recall instructions, but this is also investigated in the present thesis.

Methodological Background

All the three papers and the brief report included in this dissertation are either completely or to a large extent based on a main experiment. Therefore, a method description for the main experiment is relevant for all the three studies in this dissertation and will be included first. Following the method description comes a summary of each of the three individual manuscripts followed by the short report. The methodological issues specific to each manuscript will be addressed in the relevant summary.

Main Experiment

Participants

The participants consisted of eighty-nine undergraduate students from Lund University. There were 62 women and the mean age of the participants was 25 years (18 to 47 years). Each participant received a movie ticket worth 90 SEK (approximately US$ 12).

Initially, 23 participants were recruited for each of the four conditions in the experiment. There were 4, 7, 6, and 1 dropouts from the Retell condition, Lab-discussion condition, Family discussion condition and the Control condition respectively. New recruits replaced the dropped out participants from the study. After that, there were two more dropouts from the Lab-discussion condition and one more dropout from the Control condition that were not replaced. A chi-square test was performed on the total number of participants recruited for each condition (30, 29, 27 and 24) to check if there was any significant difference between the numbers of dropouts in the four conditions. The chi-square test did not reach significance ($[3, N = 110] = .074$, $p = ns$, $Phi = .08$). Consequently, we can still assume that the assignment of the participants into the different conditions was random.

Design

The experiment had a between-subjects design with four conditions. The four conditions were: Retell (n= 23), Lab-discussion (n = 21), Family discussion (n = 23), and Control (n= 22). In the Lab-discussion condition one participant only attended four sessions out of five, but the analysis after removing that participant did not change the results so this participant was included in the final analysis.

1) *Retell condition*, the participants retold the witnessed event five times over a three week period in the laboratory to the experimenter. The participants were instructed to tell whatever they remembered about the witnessed event in detail. The

experimenter did not pose any questions. 2) *Lab-discussion condition*, the participants first retold the witnessed event five times over a three week period in the laboratory to a confederate (each time new) who then posed questions about the event, which the witness answered. 3) *Family discussion condition*, the participants first retold the witnessed event five times over a three week period to their own family and friends (each time to a new person) who posed questions about the event, which the witness answered. 4) *Control condition*, no retelling or discussion before the final recall took place. The critical condition was the *Lab-discussion*. The *Family discussion* was a more ecologically valid, but methodologically looser, version of the Lab-discussion condition. Both of these conditions were intended to investigate how formal discussions and more informal discussion may impact recall and confidence. The Retell condition can be considered a control condition where participants retell the information, but do not receive any potentially distorting input.

Material

Videotape

A color film about the kidnapping of a woman at a bus stop by two men was shown. The film was 3 min and 50 s long and was shown on a 28-inch color television. This film has been used in previous research (Allwood, Ask, & Granhag, 2005; Allwood, Jonsson, & Granhag, 2005; Granhag, 1997).

The film is shot from an eyewitness perspective. It shows a woman coming to the bus stop. She checks the bus timetable and sits on the bench to wait for the bus. A few cars pass in front of the stop and three women walk by the scene. One of the passing-by women also checks the bus timetable. When she is leaving the bus stop the first woman asks her Excuse me, what is the time?. Quarter to one the second woman answers and leaves. The first woman then stands up and waits for the bus. A car stops by the bus stop and two men appear from the car. One man presents an identity card to the woman. The other man catches the woman from behind. She resists but is over-powered by the two men. The womans handbag falls on the ground and some items from the handbag fall out. One man goes to get the handbag. When the man is collecting items from the pavement the witness (the camera perspective) attempts to have a closer look. The man pulls out a revolver and threatens the witness and the witness retreats instantly. The man then collects the items and returns to his partner. They then force the woman into the car and drive away.

Questions about the film

Forty-four focused questions about the short film were used. Each question had two alternatives where one was always correct. The participants were instructed to choose one of the answer alternatives. If they did not remember the correct answer, the participants were instructed to make a guess and choose one. The questions were about different details like the persons appearances, clothes, ages, and the surrounding environment with letterboxes, cars, busses, and the offenders car.

18

Confidence judgment scales

Two confidence judgment scales were used in the main experiment. First: to rate the confidence for the detailed parts of the free recalls an 11-point scale was used. This scale went from 0% (Completely sure that I remember wrong) and then in steps of 10%, 20%, 30%, to 100% (Completely sure that I remember correctly). The other confidence judgment scale was used by the participants to confidence judge the correctness of their answers to the 44 focused questions where, as noted above, each question had two alternatives. Here the probability to choose the right answer was 50%. Therefore, this confidence scale went from 50% (Guessing), 60%, 70%, to 100% (Completely sure). Consequently, this was a 6-point scale.

Procedure

The participants were received in the lab and they were informed that the research was about human perception in different forensic situations. First, they watched a short film after which they got further instructions. The film was shown to groups of between four and eight individuals. After the end of the film, the participants were randomly assigned to one of the four conditions: (a) Retell, (b) Lab-discussion, (c) Family discussion and (d) Control condition. The participants in the three experimental conditions were given time schedules for the five meetings over a twenty day period. They were instructed that they would receive further instructions at the next meeting.

In the Retell condition, participants returned to the laboratory a total of five times over a twenty day period. In each of the five sessions they simply told the story in the short film to the experimenter. They were instructed to tell whatever they remembered about the film. They were asked no questions. All retellings were recorded on an MP3 recorder.

In the Lab-discussion condition, participants also returned to the laboratory a total of five times over a twenty day period. Each time, the participants first retold the witnessed event and then discussed it with a confederate who was an unknown person. One hundred and five individuals were recruited solely to work as a discussion partner to the participants in the Lab-condition. Each discussion partner took part only in one discussion. The discussion partners instructions were to listen when the participant was telling the events of the film and later to ask unprepared questions about the film. The discussion partners were also instructed that their questions were to be constructed in such a way that the discussion partner could understand the complete course of events in the film. All discussions were recorded on an MP3 recorder.

In the Family discussion condition, participants discussed the contents of the film five times over a twenty-day period with either a family member or a friend. Participants were instructed first to give an account of the film to their discussion partner and then discuss the contents of the film with him/her. Participants were also instructed to discuss the short film every time with a new family member or friend. The discussion partner could ask questions or share relevant experiences if he or she had any. In the first meeting the participants were given a time schedule for the days when they would discuss the film and were also instructed to carry out their discussions at the same time of the day on the scheduled dates. On each day the participants had

to confirm that they had completed their discussions by 7:00 pm by sending an SMS (Short Message Service: Cell phone text messaging function) to the experimenter. If the experimenter did not receive an SMS by 7:00 pm the experimenter called the participants and reminded them of their task. Participants in the Control condition were simply instructed not to talk about the contents of the short film with anybody and did not return to the lab until day 21.

All participants returned to the lab on day 21 for the memory tests. First, an open free recall test was conducted where all participants typed in Microsoft Word whatever they remembered about the events of the short film. They were asked to type as many details about the film as they could remember. Next, participants answered 44 focused questions about the short film. Participants were instructed to choose one answer from the two alternatives provided (one alternative was always correct). They were further instructed that if they could not remember the correct answer, they should make a guess.

The participants made their last visit to the lab on the twenty-fourth or twenty-fifth day to give their confidence judgments. First, the participants were asked to give their confidence judgments on their free recall statements that had been prepared as explained below. Participants gave their confidence judgments on the 11-point scale (described above). Next, participants gave their confidence judgments for their answers to the 44 forced-choice questions. Participants were then debriefed, thanked and dismissed.

Preparation of material for participants confidence judgments

In order to prepare for the participants confidence judgments their free recalls were broken down into single pieces of information. This was done by applying the criteria used by Allwood et al. (2005). The procedure was as follows: (1) The statements about actors and actions carried out were treated as one unit, for example, a woman passed by was treated as a single unit. (2) An object with one attached attribute was treated as one unit, for example, a blond woman was treated as one unit. (3) An object described by more than one attribute was treated as two units, where the additional attributes were treated as separate units, for example, the tall blond woman was treated as two units. (4) If actors and actions were described by many attributes, the actor and act was used as one unit while the attributes were rendered individually, for example, a blond woman with a long coat walked by was rendered as three units: a woman walked by, blond and long coat. A single experimenter did the coding. To help the participants to remember the context in which they wrote some statement, we added one or two sentences related to that specific statement. The items to be confidence judged were underlined while the reference items were put in the brackets. Finally, an 11-point confidence scale was inserted directly below each piece of information. Similarly a six-point confidence scale from 50% (Guessing) to 100% (Completely sure) was inserted below each of the 44 questions.

Measurements

We calculated three calibration measures to measure the realism in the participants confidence judgments: calibration and over/underconfidence, and slope. In addition to these calibration measures we also calculated Number of correct items and Num-

ber of incorrect items. *Calibration and over-/underconfidence* relate to the relation between the levels of confidence and accuracy. The formula used for computing the calibration measure is

$$\text{Calibration} = \frac{1}{n} \sum_{t=1}^{T} (r_t - c_t)^2$$

Here n is the total number of questions answered and T is the number of confidence levels used. We used eleven (free recall; 0, 10, 20, , 90, 100) or six (focused questions; 50, 60, ,90, 100) confidence levels. c_t is the proportion of correct answers for all items at confidence level r_t, and n_t is the number of times the confidence level r_t was used.

Over/underconfidence is computed the same way as calibration with the only difference being that the differences between the mean confidence and the proportion of correct units at each confidence level are not squared.

Slope measures another aspect of the realism in confidence judgments, namely to how well the participant can use his/her confidence judgments to separate correct and incorrect answers. It is computed by subtracting the mean level of confidence for a participants incorrect items from the mean level of the confidence for his/her correct items.

Summary of the Empirical Studies

Study I.

Effects of communication with a non-witness on eyewitnesses recall correctness and meta-cognitive realism Sarwar, Allwood, Innes-Ker, 2010, Applied Cognitive Psychology)

Aim of Study

This study aimed to investigate the effects of eyewitness discussions with non-witnesses (persons who have not experienced the event) on eyewitness memory and meta-memory realism for overall information about an event. In brief, previous research findings regarding the difference between the impact of eyewitness discussions and mere retellings show that mere retellings improve the memory of the learned material because of the *Testing effect* (Roediger III Karpicke, 2006a). In contrast, while sharing the witnessed event with family and friends eyewitnesses can be expected to also share such information that they are not sure about (Koriat Goldsmith, 1994, 1996; Roebers, Moga, Schneider, 2001). In response to an eyewitness story their family and friends may also share their personal similar experiences (Dritschel, 1991). Consequently, the interaction between an eyewitness and his/her family and friends may distort the eyewitness memory. Moreover, repeating the witnessed event multiple time may also distort the eyewitness confidence because of the *reiteration effect* (Hertwig, Gigerenzer, Hoffrage, 1997).

Method

All data from the main study except the recordings of sessions 1-5 in the Lab- and Retell condition, and the focused questions was used in this study.

Results

Analysis of the recall information from the participants free recall showed that the participants in the Lab-discussion and the Retell condition conditions reported a higher number of correct details than the Control condition. There was no difference between the Family discussion condition and the other three conditions for the number of correct details reported. Participants in the Lab-discussion condition also

reported more incorrect details than those in the Retell condition, while there was no difference between the other conditions for the number of incorrect details reported. Accuracy was higher in the Retell condition as compared with the Lab, Family, and Control conditions. There was no difference between the Family and Control conditions for accuracy. Participants in the Retell condition were more confident and better calibrated then the participants in the Control condition. There were no other differences between the conditions for confidence and calibration. There were also no differences between the four conditions for over-/underconfidence and slope. The analysis of the transcription from the Lab-discussion condition and Retell condition showed that the participants in the Lab-discussion reported more confabulations then the Retell condition. Moreover, the discussion partners asked more questions about peripheral information than about central questions.

Discussion

The results in study I suggest that discussion of an experienced event with non-witnesses may reduce some of the beneficial memory and meta-memory effects caused by mere retelling, but may have no great negative effects compared to a control condition. The results showed no support to the reiteration effect (Hertwig et al., 1997). This study also successfully induced memory distortions using a novel method of Lab-discussion. Implications of these findings for forensic situations are that discussing an experience event should be avoided since these discussions are likely to affect the eyewitness memory and meta-memory judgments. In contrast, mere retellings can be helpful for the eyewitnesses memory and meta-memory realism for the experienced event.

Study II.

Effects of repeated recall and discussion on eyewitness accuracy and meta-memory realism for different types of forensic information (Sarwar, Allwood, & Innes-Ker, unpublished manuscript)

Aim of Study

This study aimed to investigate the effects of multiple retellings and discussions on eyewitness accuracy and meta-memory realism for the different types of information the police may ask for at the beginning of a crime investigation. This information consists of Forensically central information (e.g. what happened, how it happened) and Forensically peripheral information (e.g. description of offenders, objects used, and the time and place of incidence). Moreover, non-forensic information was analyzed. Participants memory and meta-memory for Forensically central and peripheral information, so far, have not been investigated in the context of open free recall. Moreover, we used the different calibration measures (calibration, over-/underconfidence, and slope) to get a comprehensive understanding of meta-memory process. In addition, previous similar studies use different methods to classify the forensically central and peripheral information (e.g. Roberts & Higham, 2002 using cognitive interview and an empirical model to classify information into different categories).

Experiment 1

Aim and Predictions

Experiment 1 aimed at investigating the participants performance for different types of information in the free recalls and focused questions. In this experiment all the four conditions from the main experiment were included in the analysis. It was predicted that: 1) The participants would have better memory and meta-memory realism for the Forensically central information than the Forensically peripheral information. The reason was that the forensically central information is well-integrated with other forensically central information and that is likely to make it possible for the participants to take advantage of the *testing effect* (Roediger III Karpicke, 2006a, 2006b) and *monitoring effectiveness* (Koriat Goldsmith, 1996) as compared to the forensically peripheral information, which consists of individual facts about the description of persons involved and objects used. 2) The participants in the Lab-discussion condition and Family discussion condition were expected to have lower accuracy and meta-memory realism for forensically central information as compared to the Retell condition. The reason was that the discussions were expected to introduce information that might decrease the accuracy of these participants (Nourkova, Bernstein, Loftus, 2004; Wright, Self, Justice, 2000). In contrast, confidence of the participants in the same conditions would be inflated due to increase in retrieval fluency caused by multiple retrievals (Shaw, 1996; Shaw McClure, 1996). 3) Participants in the three experimental conditions were expected to show poorer calibration than the Control condition for Forensically peripheral information. The reason is that since the participants memory for Forensically peripheral information has been consistently found to be weak in previous research (e.g. Migueles Garcia-Bajos, 1999; Yuille Cutshall, 1986) the three experimental conditions are not likely to differ in accuracy because of a floor effect. Moreover, and more importantly the post-event retellings would inflate the confidence in the three experimental conditions due to an increase in retrieval fluency (Shaw, 1996; Shaw McClure, 1996). Consequently, with poor accuracy and inflated confidence, the participants in the three experimental conditions would show poorer calibration than the Control condition.

Classification of information

Forensically relevant information was separated from the forensically irrelevant information. Thus, the participants statements were first divided into forensic and non-forensic statements. Next, the *Forensic information* was further subdivided into the Forensically central information and Forensically peripheral information by using the plot relevancy model by (Heuer and Reisberg (1990). The plot relevancy model was used because the forensically central information, which addresses the questions of what happened and how it happened can often be considered central to the police interests in a crime investigation. The forensically peripheral information addresses the questions of the descriptions of offenders, objects used, and the time and place of incidence.

25

Method

The participants results for the free recall and focused questions including the confidence judgments from the main experiment were used.

Results

The results showed that the conditions did not differ in the *proportion* of different types of information reported. When comparing the amount of information reported the results showed that participants recalled more Forensically central items than Forensically peripheral and Non-forensic items. The participants in the three experimental conditions reported more forensically central items than the control condition. The results for the correct items were that participants recalled more Forensically central items than Forensically peripheral and Non-forensic items. Participants in the Lab-discussion condition and the Retell condition reported more correct forensically central and Non-forensic items than the Control condition. The difference between the number of Forensically central and Forensically peripheral items recalled was high in the Retell condition as compared to the other three conditions.

The results also showed that the participants had higher accuracy and confidence, better calibration, and less over-/underconfidence for the Forensically central information and Non-forensic information as compared to the Forensically peripheral information.

The *Focused questions* were all forensically peripheral, therefore differences for the different types of information could not be analyzed for forensic questions. For forensically peripheral information the results of the Focused questions showed that there was no difference between the conditions for any of the measures.

Experiment 2

Aim and Predictions

Experiment 1 was somewhat limited in its comparison of the participants performance for *Forensically central* and *Forensically peripheral* information and there were two reasons for this. 1) The participants in each condition reported a low number of *forensically peripheral* items as compared to the *forensically central* items. 2) The 44 focused questions only asked for Forensically peripheral information. This did not allow a full comparison of the participants performance for the forensically central and peripheral information. To further explore the participants performance for Forensically central and peripheral information using focused questions we conducted Experiment 2. In light of the findings of Experiment 1 it was predicted that the participants would have better memory and meta-memory realism for the Forensically central questions as compared to the Forensically peripheral questions.

Method

A within-subject design was used where the within-subject factor was the two types of focused questions, Forensically central and Forensically peripheral. The participants watched the same video as in Experiment 1 and after a filler task answered the questionnaire. The questionnaire consisted of 63 questions about the details of the

events shown in the film. Eighteen questions were about the forensically central details and 45 questions were about the forensically peripheral details.

Results

The results showed that participants had significantly higher accuracy and confidence, better calibration and slope for the Forensically central information as compared with the Forensically peripheral information. However, participants showed less over-/underconfidence for the Forensically peripheral information as compared to the Forensically central information.

Discussion

To summarize, it was found in Experiment 1 that participants in general had better memory and meta-memory realism for the Forensically central information and Non-forensic information as compared to the Forensically peripheral information. However, it was also found that the discussions and mere retellings make people recall more total and correct Forensically central items as compared to no discussion or retelling, i.e., the Control condition. Results also showed that the Forensically peripheral information was more difficult to remember than Forensically central information. A possible contributing explanation to this is that a lower degree of integration of forensically peripheral information is likely to make this information more vulnerable to retrieval induced forgetting as compared to forensically central information

Further, the results of Experiment 2 supported the findings of Experiment 1 and showed that the participants performed better on all the measures except over-/underconfidence for Forensically central information as compared to the Forensically peripheral information. These results have implications for professionals in the criminal justice system. The results suggest that they may put greater trust in the eyewitness description of the Forensically central information (e.g. who did what) than in the participants Forensically peripheral information (e.g. description of the suspect). Moreover, participants confidence for Forensically central information is likely to be realistic and it may be more trusted as a signature of correctness than participants confidence for Forensically peripheral information.

Study III.

Content analysis of eyewitnesses repeated recalls and discussions (Sarwar, Allwood, & Innes-Ker, Unpublished manuscript)

Aim of Study

This study aimed to investigate the quantitative and qualitative change in eyewitness memory and the realism in confidence over successive retellings and discussions of an experienced event. The first hypothesis was that the amount of reminiscence (retrieval of previously unrecalled information both correct and incorrect) and hypermnesia (improvement in the retrieval of correct information over successive retrieval attempts) would increase more over the successive discussions in the Lab-discussion condition

as compared to the successive retellings in the Retell condition. The reasons for this prediction are: First, the questions asked by the discussion partner would make the eyewitness recall even more new information. Second, information contributed by the discussion partners could be remembered (Gabbert, Memon, Allan, 2003; Marsh, 2007; Tversky Marsh, 2000). In contrast, in the Retell condition the reason for reminiscence and hypermnesia may only be the repeated recall attempts (La Rooy, Pipe, Murray, 2005; Payne, 1987; Turtle Yuille, 1994). The second hypothesis was that information that was more often retold or discussed in the five experimental sessions was more likely to be reported in the final test. One possible reason would be the testing effect (Roediger III Karpicke, 2006b). Last, the final hypothesis that the multiple retellings or discussions of the Forensically peripheral information would inflate the participants confidence judgments about the Forensically peripheral information without improving the accuracy (Hertwig, et al., 1997) may be because of increased retrieval fluency.

Method

Only two of the four conditions in the main experiment were used in this study. These were the Retell (n= 23) and Lab-discussion (n = 21) conditions. The data used in this study was collected in the main experiment but was not analyzed in Study I and Study II. The retellings and Lab-discussions that took place in the five sessions in the lab were recorded and later transcribed. This transcribed data was used in this study as were the data collected in the final sixth recall and confidence judgment session (session 6). The participants statements for each session were coded into single units of information by following the procedure described in the methodology section.

Results

The results showed that the participants reported significantly more correct and incorrect items in session 6 as compared to each of the five experimental sessions. The five experimental sessions did not differ in the number of correct and incorrect items reported. The participants reported more correct, incorrect, and new Forensically central items than Forensically peripheral items in each of the five experimental sessions. Results also showed that the participants in the Lab-discussion condition reported more correct, incorrect, and new items than the participants in the Retell condition. Further, the participants were more accurate for Forensically central information as compared to Forensically peripheral information in each of the five sessions. The results also showed that the participants in the Retell condition showed better accuracy in session 6 as compared to session 3 and session 4. In contrast, the participants in the Lab-discussion condition showed lower accuracy in session 6 than session 3 and session 4.

The results for the Forensically central and the Forensically peripheral information showed that the information that was retold or discussed four to five times was reported more at the testing session (session 6) as compared to the information that was retold one to two times. In addition, whether the information (all the types) was retold or discussed in the earlier sessions or in the later sessions did not affect the reporting of that information at the testing session.

Finally, the results for the Forensically peripheral information showed that the information that was retold or discussed more than three times was assigned higher confidence judgments as compared to the information that was retold or discussed less than three times. However, accuracy was not affected by whether the information was retold or discussed more or less than three times. Together this suggests the presence of a reiteration effect over the sessions, i.e. that confidence increases with more recalls in spite of the fact that accuracy is stable. No reiteration effect was found for Forensically central information.

Discussion

It was found that the Lab-discussion condition indeed caused more reminiscence and hypermnesia as compared to the Retell condition and hence supported the first hypothesis. This result is in line with the *testing effect* (Roediger III Karpicke, 2006a). The results also supported the second hypothesis and showed that the more sessions that a piece of information is repeated facilitates retrieval for that same information at the final recall. One reason why the information that was repeated less in the earlier sessions was not reported in the testing session may have been retrieval induced forgetting. Interestingly, if the information was retold or discussed in earlier or later sessions did not predict if this information would be reported in the testing session or not. Last, the analysis explored the presence of a reiteration effect for the Forensically central and peripheral information. The results showed that the Forensically peripheral information was vulnerable to the reiteration effect. One possible explanation for this result is that the forensically peripheral information is difficult to remember correctly due to it being poorly integrated but the increased retrieval fluency caused by the multiple retrievals may have affected the participants judgments about the accuracy of the forensically peripheral information. That there was no clear sign of the reiteration effect for Forensically central information may thus be due to that it was better integrated than the Forensically peripheral information.

Short report.

One year follow-up of the effects of communication with non-witnesses on eyewitnesses memory and meta-memory realism (Sarwar, F. unpublished short report)

Aim of Study

This short study was a one year follow-up of the participants in Study I in this thesis Sarwar, Allwood, and Innes-Ker (2010). This study investigated the effects of discussing an experienced event with others on both accuracy and the realism in confidence in recall (i.e. meta-memory) after one year.

Method

The 89 participants of the Sarwar et al. (2010) study were contacted after a year. Seventy-six participants (54 women) agreed to participate in this follow-up study, Lab-discussion (n = 19), Family discussion (n = 19), Retell (n= 20), and Control

condition (n= 18). This time all the participants completed the same memory tasks as in the final sessions of the main experiment. However, rather than performing the task in the lab, the participants were e-mailed the information, and asked to complete all the tasks at home.

Results

The results (Table 1) showed that there was a significant difference between the conditions for number of correct items recalled. Further analysis showed that the participants in the Lab-discussion condition recalled significantly more correct items than the participants in the Control condition. There were no significant differences between the conditions for the incorrect items, accuracy, confidence, calibration, over-/underconfidence and slope.

Discussion

The results suggest that over a one-year period, the memory and meta-memory benefits for the Retell condition as compared with the Lab-discussion condition and the Family discussion conditions disappeared. It is difficult to interpret the significant difference between the Lab-discussion condition and the Control condition for the number of correct items. The reason is that these two conditions did not differ on any other measure. An implication of these results for forensic situations is that the witnesses may show poor memory when there is a long interval between experiencing the crime and reporting it and that long intervals are likely to reduce the effects of discussions and retellings as well.

Table 1.

Free Recall: Means (and SDs) for Correct items, Incorrect items, Accuracy, Confidence, Calibration, Over-/underconfidence and Slope. and F-values for the Corresponding ANOVAs

	Lab-discussion	Family discussion	Retell	Control	F
Correct items	24.00 (9.15)	20.21 (7.79)	22.10 (9.10)	15.83 (8.05)	3.08*
Incorrect items	7.63 (5.27)	7.63 (5.23)	6.35 (4.03)	6.33 (5.36)	0.42
Accuracy	.77 (.12)	.74 (.12)	.78 (.12)	.73 (.14)	0.71
Confidence	81.15 (8.81)	84.19 (8.64)	86.49 (7.63)	79.66 (17.62)	1.41
Calibration	.05 (.03)	.07 (.05)	.07 (.05)	.08 (.08)	1.03
O/U confidence	.04 (.11)	.10 (.14)	.08 (.14)	.06 (.19)	0.68
Slope	12.85 (10.87)	10.81 (7.88)	10.42 (16.33)	9.09 (10.65)	0.3

*$p < 0.05$

Note. O/U confidence = Over-/underconfidence.

General Discussion

Eyewitnesses often share the witnessed event with their family and friends before they describe it to the investigators or formally testify in court (Paterson & Kemp, 2006). These discussions have consequences for the eyewitness memory and meta-memory realism, but so far these consequences on later memory performance, especially on meta-memory performance, have not been well investigated.

The present thesis aimed to study the effects of eyewitness retellings and discussions with non-witnesses on eyewitness memory and meta-memory realism. To recapitulate, Study 1 investigated the effects of eyewitness retellings and discussions on memory and meta-memory realism for the witnessed event details. *Study II* investigated the effects of eyewitness retellings and discussions on memory and meta-memory judgments for Forensically central information and Forensically peripheral information about the experienced forensic episode. *Study III* investigated if eyewitness successive retrieval attempts would increase reminiscence and hypermnesia over five recalls. It also investigated if the number of times a piece of information was repeated would predict the reporting of that information at the testing session. Finally, it analyzed if the multiple retrieval attempts would increase the level of the eyewitness confidence judgments (the reiteration effect). A short report presented after the summaries of the three studies explored if the effects of eyewitness retellings and discussions was sustained over a one year long time interval.

In the following sections, first the main results are briefly summarized followed by the discussion of the factors that may have caused errors in eyewitness memory and increased the lack of realism in the participants confidence judgments.

Summary of the Main Findings

The results suggested that discussing an experienced forensic event may reduce some of the beneficial memory and meta-memory effects caused by merely retelling it. Moreover, discussing an event may have no great negative effects compared to not retelling or discussing it. Interestingly, the participants also introduced misinformation in their recollections even when no misinformation was deliberately suggested to them.

The results also showed better memory and meta-memory realism for the Forensically central and Non-forensic information as compared to the Forensically peripheral information. Moreover, the mere retellings and discussions make people recall more total and correct Forensically central items as compared to not involving any kind of communication. The findings that there was no difference between the four conditions for the memory and meta-memory realism for the Forensically peripheral information

may have been due to that in general the forensically peripheral information was difficult to remember.

Finally, the results showed that discussing an experienced event caused more reminiscence and hypermnesia as compared to merely retelling it. The number of times a piece of information was retold or discussed during the five sessions facilitated the retrieval of such information at the testing session. Finally support for the reiteration effect was found only for the Forensically peripheral information.

Effects of Eyewitness Discussions with Non-witnesses on Memory Accuracy

It was hypothesized that the participants in the Retell condition would show higher accuracy for the overall information (*Study I*) and for the Forensically central information (*Study II*) than the participants in the discussion conditions and the control condition. The reason was that active repetition (i.e., mere retellings) is known to improve accuracy because of the *testing effect* (Roediger III & Karpicke, 2006a, 2006b). In contrast, the input from the discussion partners in the two discussion conditions was predicted to have negative consequences for accuracy (Nourkova et al., 2004; Wright et al., 2000). It was also predicted that the participants in the Lab-discussion condition would show higher reminiscence and hypermnesia than the participants in the Retell condition (*Study III*). Further, it was also predicted that the more a piece of information was retold or discussed during the five experimental sessions the more likely it would be that that piece of information would also be reported in the testing session.

The results from *Study I* supported our prediction that the Retell condition would show higher accuracy for the overall information as compared to the other three conditions. The reason for the lower accuracy scores in the Lab-discussion may be attributed to the higher number of confabulations reported in this condition compared to the Retell condition. One source of these confabulations may be the questions asked (on average about 14 questions per session) by the discussion partners in the Lab-discussion condition. This result supports the findings of Kebbell and Johnson (2000) and Loftus (1975) that the questions posed to an eyewitness can potentially distort eyewitness memory. The one year follow-up study revealed that the memory benefits for the Retell condition as compared to the other three conditions (particularly as compared to the two discussion conditions) had disappeared after twelve months.

In this context it is also interesting to note that the results of *Study II* showed that there was no significant difference for accuracy of the Forensically central information among the four conditions. Some possible reasons why such a difference was not detected might be that the mixed ANOVA conducted in *Study II* is a fairly stringent test and also that it excluded those participants from the analysis who did not report either type of information. Consequently, the ANOVA results presented were based on 71 participants (total N = 89). This might have lowered the power of the analysis. However, inspection of the means showed that the results were in the expected direction. Participants in the Retell condition were more accurate than participants in the two discussion conditions. One possible explanation for these tendencies may be that the contributions of the participants discussion partners might have distorted the participants memory for forensically central information (see Russell & Schober,

34

1999).

Further, in *Study I* and *Study II* there was no difference between the two discussion conditions and the Control condition for the accuracy of overall information and Forensically central and peripheral information. It is interesting that the two discussion conditions, despite having the greatest number of correct items in absolute numbers for overall information and Forensically central and peripheral information, did not differ in accuracy from the Control condition. The reason was of course that the participants in the discussion conditions also reported more incorrect details. This result supports previous research findings showing that discussing an event does increase the total number of correct items, but it also causes people to recall more incorrect details (Loftus, 2003; Luus & Wells, 1994; Marsh, 2007; Tversky & Marsh, 2000).

There was no difference between the four conditions in the memory for the Forensically peripheral information and the same was the case for the 44 focused questions in Experiment 1 (*Study II*), which all consisted of Forensically peripheral information. However, a limitation of this result was that the number of peripheral information items reported in each condition was quite low (Mean = 10). This result suggests that forensically peripheral information is quite difficult to remember. This low number of forensically peripheral items is in line with the previous research findings that people recall less of the Forensically peripheral information as compared to the Forensically central information (Roberts & Higham, 2002; Roebers et al., 2001; Wessel & Merckelbach, 1997).

The results of *Study III* showed that both the Retell condition and the Lab-discussion condition showed reminiscence and hypermnesia over the six analyzed recall sessions. This result supports the previous findings of retrieval of new information with successive retrievals (La Rooy et al., 2005; Mulligan, 2006; Payne, 1987) and that successive recalls improve memory in terms of amount of correct information recalled (Henkel, 2004; Mulligan, 2006; Payne, 1987). As predicted, the results also showed that the amount of reminiscence and hypermnesia was higher in the Lab-discussion condition than in the Retell condition. The apparent reason, as suggested in the introduction, is likely to be the questions asked by the confederates to the participants during the discussions because these questions may have caused the participants to recall more information.

Effects of Discussions and Retellings on Forensically Central and Peripheral Information

Keeping the applied perspective in mind, *Study II* investigated what kind of police questions eyewitnesses could answer better. The results showed that the amount of recalled Forensically central, peripheral and Non-forensic information reported was 58%, 20.5%, and 21.5 respectively. This shows that the major portion of recalled information was forensic in nature, 78.5% (Forensically central plus Forensically peripheral). This result is compatible with the idea that the participants exerted control over what to share on the basis of their expectation of what was required from them (Grice, 1975; Russell & Schober, 1999). Moreover, the major portion of information consisted of Forensically central information. This result supports previous research findings that people in their free recall report less Forensically peripheral information

35

as compared to the Forensically central information (Hershkowitz & Terner, 2007; Roebers et al., 2001). In addition, the proportion of forensically peripheral information may have decreased even further as an effect of multiple retrieval attempts since previous research has shown that multiple retrieval attempts decrease the amount of peripheral information (Hershkowitz & Terner, 2007). Interestingly, though the participants discussions and retellings did affect the quantity and quality of Forensically central, peripheral, and Non-forensic information as compared between the conditions, the composition of proportions of these three types of information did not differ between the conditions.

As predicted, the results also showed that the participants accuracy was better for the Forensically central information as compared to the Forensically peripheral information. This result also supports earlier empirical findings where focused questions were used to investigate the participants memory for the Forensicaly central and peripheral information (e.g. Burke et al., 1992; Heuer & Reisberg, 1990; Ibabe & Sporer, 2004). The findings of this thesis extend the previous findings to eyewitness open free recall.

The results also showed that participants had better memory and meta-memory realism for Non-forensic information as compared to the Forensically peripheral information. Non-forensic information, such as houses, roads, surroundings, etc., is either not needed to solve the crime or can be collected from the crime scene without the help of an eyewitness. However, this result is important from an applied perspective since a witness better accuracy for the non-forensic information can affect the investigators credibility judgments of the witness recall of forensically relevant information because of a possible Halo effect (Dennis, 2007). An example of a halo effect in another context is that teachers evaluation of a students performance in one subject may be influenced by how the student is performing in another subject (Dompnier, Pansu, & Bressoux, 2006). However, if the halo effect would include Forensically peripheral information it could cause error since the results showed that participants had low accuracy for the Forensically peripheral information.

Should Confidence be Trusted?

In *Study I* the results showed that for the overall information participants in the Retell condition were more confident than participants in the Control condition. Further, there was no difference between the confidence level of participants in the two discussion conditions and the Control condition. In *Study II* the results of Experiment 1 (using open free recall) and Experiment 2 (using focused questions) showed that the participants had higher confidence levels for the Forensically central information than Forensically peripheral information. The result of Experiment 2 is in line with the previous finding using focused questions that participants show higher confidence for the Forensically central information than the Forensically peripheral information (Ibabe & Sporer, 2004; Migueles & Garcia-Bajos, 1999; Roberts & Higham, 2002). The results of Experiment 1 in Study 2 extended these findings to open free recall. The realism in confidence judgments was analyzed in this thesis.

It was predicted that the participants in the Retell condition would show better realism in their confidence judgments as compared to the Lab-discussion and the Family discussion conditions for the Overall information (*Study I*) and the Forensically

central information (*Study II*). This was expected because the empirical findings regarding the improvement in memory accuracy as a result of active repetition has been found to be more consistent in previous research (Roediger III & Karpicke, 2006a) than the increase in confidence as a result of repetition with accuracy being constant, which is the reiteration effect (Hertwig et al., 1997).

Both the aspects of realism, bias and separation, were analyzed. To analyze bias calibration and over-/underconfidence were calculated. The results showed, as predicted, that for the overall information the participants in the Retell condition were better calibrated than the participants in the Control condition (*Study I*). For the Forensically central and peripheral information there was no difference between the participants in the four conditions (*Study II*). Further the results of Experiment 1 (*Study II*) showed that, for open free recall, participants were better calibrated, and showed less over-/underconfidence for Forensically central information than Forensically peripheral information. Experiment 2 (*Study II*) showed that, for the focused questions, participants were better calibrated for Forensically central information than for Forensically peripheral information. In contrast, participants showed less over-/underconfidence for Forensically peripheral information than Forensically central information. However, this result was an effect of the balancing of underconfidence at the lower end of the confidence scale with overconfidence at the higher end of the confidence scale.

It is not meaningful to compare the present findings with the earlier relevant research work in this context since, as discussed above, either the researchers have been using different classification methods to categorize the Forensically central and peripheral information (Roberts & Higham, 2002) or they did not use calibration measures to study the confidence accuracy relationship (e.g. Migueles & Garcia-Bajos, 1999).

To measure the *separation* aspect of the realism, slope was calculated. The results for the overall information (*Study I*) and Forensically central, peripheral, and Non-forensic information (*Study II*) showed that the participants in the four conditions were equally capable to discriminate between correct and incorrect items by means of their confidence. Further, the results of Experiment 1 (*Study II*) showed that there was no difference between the participants ability to discriminate between correct and incorrect Forensically central and Forensically peripheral information. In contrast Experiment 2 (*Study II*) showed that, for the focused questions, participants ability to discriminate between correct and incorrect was better for Forensically central information than Forensically peripheral information. These results showed that though the discussions affected the memory and confidence of the participants, it did not have a strong impact on the participants ability to monitor their accuracy levels and to separate correct and incorrect items. In general the results give support to the idea that confidence can be used as a predictor of accuracy. Sorting the items into items that were assigned high confidence judgments and items that were assigned low confidence judgments may help to separate the correct items from the incorrect items. The sorting should be done separately for each type of information (Forensic, central, peripheral, etc) because the results also showed that participants assigned different levels of confidence judgments to different types of information. For example, Forensically central items had higher levels of confidence as compared to the Forensically peripheral information. In brief, eyewitness confidence is an important piece of in-

formation and if used carefully it can help to determine the level of correctness of eyewitness statements.

Effects of Eyewitness Retellings and Discussions on the Presence of the Reiteration Effect

It was predicted that confidence would be higher in the three experimental conditions (Lab, Family, and Retell) than in the Control condition for the overall information and the Forensically central information. This prediction was made on the assumption of the presence of a reiteration effect (Hertwig et al., 1997). The result of *Study I* showed insufficient support for the hypothesis by showing that only the Retell condition had a higher confidence than the Control condition and moreover *Study II* showed that this difference held only for the Overall information. There was no difference between the three experimental conditions and the Control condition on confidence for the Forensically central and peripheral information. In brief, these results did not support the presence of a reiteration effect (Hertwig et al., 1997) since the higher confidence in the Retell condition could simply be because of the fact that accuracy was higher in the Retell condition than in the other conditions.

There could be two possible reasons that worked against the detection of a reiteration effect when analyzed as just described. One reason could be that the participants in the discussion conditions may have felt a pressure for social accountability that acted to attenuate the increase in confidence in the discussion conditions (Tetlock, 1983b). This pressure may have been most clearly felt in the Lab-discussion condition, which, although not strictly formal, was of a more formal character than the Family discussion condition. How confidence is affected by the reiteration effect in different forms of social situations should be further investigated in future research.

A second possible reason could be reminiscence (retrieval of new information). *Study III* showed that the participants in the Lab-discussion condition had higher reminiscence than the Retell condition. However, it also showed that the participants in the Retell condition were repeating almost the same information during the five sessions. In contrast the participants in the Lab-discussion condition actually did retrieve a lot of new information in each session and in the final testing session that was not discussed in the previous sessions. Moreover, the information that was reported only at the testing session did receive a lower level of confidence judgments than the other information reported at the testing session (this result was not reported in the result section of *Study III*) and may be due to a lack of influence of the reiteration effect for this new information. Therefore, this information caused the mean confidence level of the participants in the Lab-discussion condition to decrease and dilute the signs of a reiteration effect at the final testing session.

Although we did not find evidence for the reiteration effect in *Study I* further exploration of the reiteration effect in *study III* showed that participants showed the reiteration effect for the Forensically central information but not for the Forensically peripheral information. A possible reason is that, as discussed in the introduction, the forensically peripheral information may not be well integrated with other information. Additional recall of the Forensically peripheral information may increase its association to other relevant information (i.e. make it more integrated) and hence cause an increase in the retrieval fluency as compared to the Forensically central information

(Shaw & McClure, 1996). The Forensically central information may be considered as well-integrated and additional recalls might cause no, or only little, effect in terms of retrieval fluency and hence no reiteration effect would be expected to occur. The lack of reiteration effect for the Forensically central information may be a contributing explanation of why confidence seems to be a better predictor of the accuracy of the Forensically central information compared to the Forensically peripheral information.

In *Study II* the results of Experiment 1 (using open free recall) and Experiment 2 (using focused questions) also showed that the participants had higher confidence levels for the Forensically central information than Forensically peripheral information. The result of Experiment 2 is in line with the previous findings using focused questions (Ibabe & Sporer, 2004; Migueles & Garcia-Bajos, 1999; Roberts & Higham, 2002). Further, the results of Experiment 1 extend the findings to open free recall.

Limitations

Like in all studies there are some limitations to this empirical work as well. Three main limitations are discussed here. These are: 1) the choices of the conditions used, 2) the time lapse between the final recall and the subsequent confidence judgments, and 3) the low frequency of the Forensically peripheral items reported by the participants.

Choice of conditions

As discussed above. The Lab-discussion condition was seen as the central condition in the main data collection of this thesis. The purpose of the Family discussion condition was to achieve ecological validity so that the effect of spontaneous informal discussions with family and friends could be explored. The participants were simply asked to discuss the experienced events five times with family and friends. After each discussion they confirmed by SMS to the experimenter that the task was done. An attempt to have more control over this condition through other means (e.g. recording protocols) could have affected the participants performance and would not have allowed us to achieve the ecological validity aspired for. Moreover, the Family discussion condition would have become quite similar to the Lab-discussion condition.

We were aware of the problems in having such a loosely controlled condition, where, other than trusting the participants, we had no possibility to make sure if the participants actually performed the assigned tasks. Moreover, we also did not have information about the variance of the duration and contents of the discussions. The performance of the participants in the Family discussion condition at the testing session (session 6) was found to be similar to the performance of the participants in the Lab-discussion condition. This suggests that the participants in the Family discussion condition did comply with the instructions to a certain extent. A credible reason for the differences in results between the Family discussion condition and the Lab-discussion condition was that the participants in the Family discussion condition spoke with people they knew. In contrast, in the Lab-discussion condition the participants communicated with strangers. According to Hope et al. (2008), communicating with strangers and communicating with people you know affects memory and confidence differently. Differences between the effects of situations similar to those in the Lab- and Family discussion conditions should be further investigated in future work.

The time lapse between the final recall and the subsequent confidence judgments

Since the participants did not record their confidence judgments immediately after the memory tests, one could argue that these confidence judgments might have been contaminated by other non-relevant factors. This argument seems valid for lineup situations where the witness confidence judgment should be recorded instantly after making the identification (Brewer & Burke, 2002; Brewer & Wells, 2006). However, it seems that this argument might not be equally relevant to the present work because of the following two reasons: 1) The participants made their confidence judgments of their statements that were present in front of them. This situation is similar to a natural forensic situation where eyewitnesses are asked how sure they are about what they have said earlier in the court. For example, You have earlier said that.: are you sure about it? 2) Even if the participants recorded their confidence judgments a few days after the memory recall it is common in a forensic context that eyewitnesses discuss the witnessed event with their family and friends many times before they formally describe it to the police or testify in the court (Paterson & Kemp, 2006). Whenever an eyewitness retrieves information in a given context it is, according to the model of Koriat and Goldsmith (1996) followed by a spontaneous confidence judgment to evaluate if he should share the information with the current audience or not (Koriat & Goldsmith, 1996). In the present context where a person has already made confidence judgments (though implicitly) many times when retelling or discussing the witnessed event with other people, it seems that some days delay between the recall test and giving the confidence judgments may not have much effect.

The low frequency of the Forensically peripheral items

A third limitation of the results was that the participants did not recall as many Forensically peripheral items as could have been ideal from a methodological perspective. For this reason any differential effects may have been buried in a floor effect. Likewise, the focused questions in *Study II* Experiment 1 only used Forensically peripheral information. For this reason, Experiment 2 *Study II* included focused questions asking for both Forensically central and peripheral information. Since the police would commonly ask probing questions on the basis of an eyewitness first free recall it is important to know how accurately eyewitnesses can report the Forensically central and peripheral information in their first report. This issue needs to be explored further in future research using a different research design where the participants would be able to report a great amount of both forensically central and peripheral information in their free recall.

Applied Implications

The results of the three studies that comprise this dissertation have some implications for the criminal justice system. First, discussions may cause eyewitnesses to recall more details about the forensic event. Discussions may also make eyewitnesses recall more incorrect details. The fact that eyewitnesses are likely to discuss the witnessed event with the people they know (Paterson & Kemp, 2006) may have as a consequence that they are able to recall more details. Because the police investigators and

the courts place more confidence on the eyewitnesses who recall more details than the eyewitnesses who cannot recall that many details (Bell & Loftus, 1988; Heath, Grannemann, Sawa, & Hodge, 1997), they should be aware that the additional details can be both correct and incorrect. Second, the results in the present thesis, as in other research, show that when the time duration after the event is fairly short (three weeks) eyewitnesses in their open free recall can very well describe what happened and how it happened (i.e., action details). In contrast, eyewitnesses may not be able to provide an accurate description of the culprit/s and the object/s used. Thus, the police need to be more careful in using the eyewitnesses description of the culprit. Third, sorting eyewitness statements about different types of information separately into statements with low and high confidence judgments can be helpful in predicting the accuracy of these statements. The statements with high confidence in each information category are likely to be more accurate as compared to the statements with low confidence in the same information category. Hopefully, the studies in this thesis have contributed to further our understanding of the effects of eyewitnesses retellings and discussions with Non-witnesses on the witnesses memory and meta-memory.

References

Allen, B. P., & Lindsay, D. S. (1998). Amalgamations of memories: intrusion of information from one event into reports of another. *Applied Cognitive Psychology*, 12(3), 277-285.

Allwood, C. M. (2010). Eyewitness confidence. In P. A. Granhag (Ed.), *Forensic Psychology in context* (pp. 281-303). Uffculme, Devon, UK: Willan Publishing.

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.

Allwood, C. M., Jonsson, A.-C., & Granhag, P. A. (2005). The effects of source and type of feedback on child witnesses' metamemory accuracy. *Applied Cognitive Psychology*, 19(3), 331-344.

Alper, A., Buckhout, R., Chern, S., Harwood, R., & Slomovits, M. (1976). Eyewitness identification: Accuracy of individual vs. composite recollections of crime. Bulletin of the Psychonomic *Society*, 8, 147-149.

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, 49(4), 415-445.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 20(5), 1063-1087.

Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 25(3), 608-629.

Bell, B. E., & Loftus, E. F. (1988). Degree of detail of eyewitness testimony and mock juror judgments. *Journal of Applied Social Psychology*, 18(14), 1171-1192.

Blank, H. (2009). Remembering: A Theoretical Interface Between Memory and Social Psychology. *Social Psychology*, 40(3), 164-175.

Bluck, S., Levine, L. J., & Laulhere, T. M. (1999). Autobiographical remem bering and hypermnesia: A comparison of older and younger adults. *Psychology and Aging*, 14(4), 671-682.

References

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology*, 72(4), 691-695.

Brainerd, C. J., Wright, R., Reyna, V. F., & Payne, D. G. (2002). Dual-Retrieval Processes in Free and Associative Recall. *Journal of Memory and Language*, 46(1), 120-152.

Brewer, N. (2006). Uses and abuses of eyewitness identification confidence. *Legal and Criminological Psychology*, 11, 3-23.

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior,* 26(3), 353-364.

Brewer, N., Potter, R., Fisher, R. P., Bond, N., & Luszcz, M. A. (1999). Beliefs and data on the relationship between consistency and accuracy of eyewitness testimony. *Applied Cognitive Psychology*, 13(4), 297-313.

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil Similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, 12(1), 11-30.

Brown, J. M. (2003). Eyewitness memory for arousing events: putting things into context. *Applied Cognitive Psychology*, 17(1), 93-106.

Burke, A., Heuer, F., & Reisberg, D. (1992). Remembering emotional events. *Memory & Cognition*, 20(3), 277-290.

Butler, K. M., Williams, C. C., Zacks, R. T., & Maki, R. H. (2001). A limit on retrieval-induced forgetting. *Journal of experimental psychology. Learning, memory, and cognition*, 27(5), 1314-1314.

Candel, I., Merckelbach, H., Jelicic, M., Limpens, M., & Widdershoven, K. (2004). Childrens suggestibility for peripheral and central details *The Journal of Credibility Assessment and Witness Psychology* 5(1), 9-18.

Castelli, P., Goodman, G. S., Edelstein, R. S., Mitchell, E. B., Alonso, P. M. P., Lyons, K. E., & Newton, J. W. (2006). Evaluating Eyewitness Testimony in Adults and Children *The handbook of forensic psychology*, 3rd ed. (pp. 243-304): Hoboken, NJ, US: John Wiley & Sons Inc.

Ceci, S. J., Huffman, M. L. C., Smith, E., & Loftus, E. F. (1994). Repeatedly Thinking about a Non-event: Source Misattributions among Preschoolers. *Consciousness and Cognition*, 3(3-4), 388-407.

Chan, J. C. K., McDermott, K. B., & Roediger III, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, 135(4), 553-571.

Christianson, S.-Å. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin*, 112(2), 284-309.

Christianson, S.-Å. (1994). Traumatiska minnen. Stockholm: Natur och Kultur.

Christianson, S.-Å., & Loftus, E. F. (1987). Memory for traumatic events. *Applied Cognitive Psychology*, 1(4), 225-239.

Christianson, S.-Å., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cognition & Emotion*, 5(2), 81 - 108.

Coman, A., Manier, D., & Hirst, W. (2009). Forgetting the unforgettable through conversation: Socially shared retrieval-induced forgetting of September 11 memories. *Psychological Science*, 20(5), 627-633.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, 14(3), 215-235.

Dennis, I. (2007). Halo effects in grading student projects. *Journal of Applied Psychology*, 92(4), 1169-1176.

Dompnier, B., Pansu, P., & Bressoux, P. (2006). An integrative model of scholastic judgments: Pupils' characteristics, class context, halo effect and internal attributions. *European Journal of Psychology of Education*, 21(2), 119-133.

Dritschel, B. H. (1991). Autobiographical memory in natural discourse: A methodological note. *Applied Cognitive Psychology*, 5, 319-330.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. Psychological Review, 66(3), 183-201.

Ebbesen, E. B., & Rienick, C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal of Applied Psychology*, 83(5), 745-762.

Erdelyi, M. H. (2010). The ups and downs of memory. *American psychologist*, 65(7), 623-633.

Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures - Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology*, 6(1), 159-171.

Erskine, A., Markham, R., & Howie, P. (2001). Children's script based inferences. Cognitive Development, 16(4), 871-887.

Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology*, 17(5), 533-543.

Garcia-Bajos, E., Migueles, M., & Anderson, M. C. (2009). Script knowledge modulates retrieval-induced forgetting for eyewitness events. *Memory*, 17(1), 92-103.

Garven, S., Wood, J. M., & Malpass, R. S. (2000). Allegations of wrongdoing: The effects of reinforcement on children's mundane and fantastic claims. *Journal of Applied Psychology*, 85(1), 38-49.

Granhag, P. A. (1997). Realism in eyewitness confidence as a function of type of event witnessed and repeated recall. *Journal of Applied Psychology*, 82(4), 599-613.

Granhag, P. A., Stromwall, L. A., & Allwood, C. M. (2000). Effects of reiteration, hindsight bias, and memory on realism in eyewitness confidence. *Applied Cognitive Psychology*, 14(5), 397-420.

Greenberg, M. S., Westcott, D. R., & Baily, S. E. (1998). When Believing Is Seeing: The Effect of Scripts on Eyewitness Memory. *Law and Human Behavior*, 22(6), 685-694.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics* (Vol. 3, pp. 41-58): Speech Acts, New York, Academic Press.

Heath, W. P., & Erickson, J. R. (1998). Memory for central and peripheral actions and props after varied post-event presentation. *Legal and Criminological Psychology*, 3, 321-346.

Heath, W. P., Grannemann, B. D., Sawa, S. E., & Hodge, K. M. (1997). Effects of detail in eyewitness testimony on decisions by mock jurors. *Journal of Offender Rehabilitation*, 25(3), 51-71.

Henkel, L. A. (2004). Erroneous memories arising from repeated attempts to remember. *Journal of Memory and Language*, 50(1), 26-46.

Henkel, L. A. (2007). The benefits and costs of repeated memory tests for young and older adults. *Psychology and Aging*, 22(3), 580-595.

Hershkowitz, I., & Terner, A. (2007). The effects of repeated interviewing on children's forensic statements of sexual abuse. *Applied Cognitive Psychology*, 21(9), 1131-1143.

Hertwig, R., Gigerenzer, G., & Hoffrage, U. (1997). The reiteration effect in hindsight bias. *Psychological Review*, 104(1), 194-202.

Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition*, 18(5), 496-506.

Hollin, C. R., & Clifford, B. R. (1983). Eyewitness Testimony: The Effects of Discussion on Recall Accuracy and Agreement. *Journal of Applied Social Psychology*, 13(3), 234-244.

Hope, L., Ost, J., Gabbert, F., Healey, S., & Lenton, E. (2008). "With a little help from my friends...": The role of co-witness relationship in susceptibility to misinformation. *Acta Psychologica*, 127(2), 476-484.

Ibabe, I., & Sporer, S. L. (2004). How you ask is what you get: On the influence of question form on accuracy and confidence. *Applied Cognitive Psychology*,

18(6), 711-726.

Jelicic, M., Smeets, T., Peters, M. J. V., Candel, I., Horselenberg, R., & Merckelbach, H. (2006). Assassination of a controversial politician: remembering details from another non-existent film. *Applied Cognitive Psychology*, 20(5), 591-597.

Jermias, J. (2006). The influence of accountability on overconfidence and resistance to change: A research framework and experimental evidence. *Management Accounting Research*, 17(4), 370-388.

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1304-1316.

Karpicke, J. D., & Roediger III, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57(2), 151-162.

Karpicke, J. D., & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966-968.

Kebbell, M. R., & Johnson, S. D. (2000). Lawyers' questioning: The effect of confusing questions on witness confidence and accuracy. *Law and Human Behavior*, 24(6), 629-641.

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.

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.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103(3), 490-517.

La Rooy, D., Pipe, M.-E., & Murray, J. E. (2007). Enhancing children's event recall after long delays. *Applied Cognitive Psychology*, 21(1), 1-17.

La Rooy, D., Pipe, M. E., & Murray, J. E. (2005). Reminiscence and hypermnesia in children's eyewitness memory. *Journal of Experimental Child Psychology*, 90(3), 235-254.

Lane, S. M. (2006). Dividing attention during a witnessed event increases eyewitness suggestibility. *Applied Cognitive Psychology*, 20(2), 199-213.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, 7(4), 560-572.

Loftus, E. F. (1979). Eyewitness testimony. Cambridge, MA: Harvard University Press.

Loftus, E. F. (1983). Misfortunes of Memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 302(1110), 413-421.

Loftus, E. F. (1992). When A Lie Becomes Memory's Truth: Memory Distortion After Exposure to Misinformation. *Current Directions in Psychological Science*, 1(4), 121-123.

Loftus, E. F. (2003). Our changeable memories: legal and practical implications. *Nature Reviews Neuroscience*, 4(3), 231-234.

Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of experimental psychology. General*, 118(1), 100-104.

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "Weapon Focus". *Law and Human Behavior*, 11(1), 55-62.

Luus, C. A. E., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perserverance effects. *Journal of Applied Psychology*, 79(5), 714-724.

MacLeod, M. (2002). Retrieval-induced forgetting in eyewitness memory: forgetting as a consequence of remembering. *Applied Cognitive Psychology*, 16(2), 135-149.

Mark, R. K., & Shane, D. J. (2000). Lawyers' Questioning: The Effect of Confusing Questions on Witness Confidence and Accuracy. *Law and Human Behavior*, 24(6), 629-641.

Marsh, E. J. (2007). Retelling is not the same as recalling: Implications for memory. *Current Directions in Psychological Science*, 16(1), 16-20.

Memon, A., & Vartoukian, R. (1996). The effects of repeated questioning on young children's eyewitness testimony. *British Journal of Psychology*, 87(3), 403-416.

Migueles, M., & Garcia-Bajos, E. (1999). Recall, recognition, and confidence patterns in eyewitness testimony. *Applied Cognitive Psychology*, 13(3), 257-268.

Mulligan, N. W. (2001). Generation and hypermnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(2), 436-450.

Mulligan, N. W. (2006). Hypermnesia and total retrieval time. *Memory*, 14(4), 502-518.

Nourkova, V., Bernstein, D., & Loftus, E. (2004). Altering traumatic memory. *Cognition and Emotion*, 18(4), 575-585.

Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child Abuse & Neglect*, 24(6), 733-752.

Parker, J. F., & Carranza, L. E. (1989). Eyewitness testimony of children in

target-present and target-absent lineups. *Law and Human Behavior*, 13(2), 133-149.

Parker, J. F., Haverfield, E., & Baker-Thomas, S. (1986). Eyewitness testimony of children. *Journal of Applied Social Psychology*, 16(4), 287-302.

Pasupathi, M. (2001). The social construction of the personal past and its implications for adult development. *Psychological Bulletin*, 127(5), 651-672.

Paterson, H. M., & Kemp, R. I. (2006). Co-witnesses talk: A survey of eyewitness discussion. *Psychology, Crime & Law*, 12(2), 181-192.

Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin*, 101(1), 5-27.

Pennebaker, J. W., & Harber, K. D. (1993). A social stage model of collective coping: The Loma Prieta earthquake and the Persian Gulf War. *Journal of Social Issues*, 49, 125-145.

Perfect, T. J. (2004). The role of self-rated ability in the accuracy of confidence judgements in eyewitness memory and general knowledge. *Applied Cognitive Psychology*, 18(2), 157-168.

Perfect, T. J., Stark, L. J., Tree, J. J., Moulin, C. J. A., Ahmed, L., & Hutter, R. (2004). Transfer appropriate forgetting: The cue-dependent nature of retrieval-induced forgetting. *Journal of Memory and Language*, 51(3), 399-417.

Pizarro, D. A., Laney, C., Morris, E. K., & Loftus, E. F. (2006). Ripple effects in memory: Judgments of moral blame can distort memory for events. *Memory & Cognition*, 35(3), 550-555.

Ramachandran, V. S., & Anstis, S. M. (1986). The perception of apparent motion. *Scientific American*, 254(6), 102.

Riniolo, T. C., Koledin, M., Drakulic, G. M., & Payne, R. A. (2003). An archival study of eyewitness memory of the Titanic's final plunge. *Journal of General Psychology*, 130(1), 89-95.

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.

Rock, I., & Palmer, S. (1990). The legacy of gestalt psychology. *Scientific American*, 263(6), 84.

Roebers, C. M., Moga, N., & Schneider, W. (2001). The role of accuracy motivation on children's and adults' event recall. *Journal of Experimental Child Psychology*, 78(4), 313-329.

Roebers, C. M., & Schneider, W. (2000). The impact of misleading questions on eyewitness memory in children and adults. *Applied Cognitive Psychology* 14(6), 509-526.

Roediger, I. H. L., Meade, M. L., & Bergman, E. T. (2001). Social contagion of memory. *Psychonomic Bulletin and Review*, 8, 365-371.

Roediger III, H. L., Jacoby, J. D., & McDermott, K. B. (1996). Misinformation Effects in Recall: Creating False Memories through Repeated Retrieval. *Journal of Memory and Language*, 35(2), 300-318.

Roediger III, H. L., & Karpicke, J. D. (2006a). The power of testing memory. *Perspectives on Psychological Science*, 1(3), 181-210.

Roediger III, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning. Psychological Science, 17(3), 249-256.

Russell, A. W., & Schober, M. F. (1999). How beliefs about a partners goals affect referring in goal-discrepant conversation. *Discourse Processes*, 27, 530-541.

Sarwar, F., Allwood, C. M., & Innes-Ker, Å. (2010). Effects of communication with a non-witness on eyewitnesses recall correctness and meta-cognitive realism. *Applied Cognitive Psychology.* Advance online publication. doi: 10.1002/acp.1749

Saywitz, K. J., Goodman, G. S., Nicholas, E., & Moan, S. F. (1991). Children's memories of a physical examination involving genital touch: Implications for reports of child sexual abuse. *Journal of Consulting and Clinical Psychology*, 59(5), 682-691.

Scrivner, E., & Safer, M. A. (1988). Eyewitnesses show hypermnesia for details about a violent event. *Journal of Applied Psychology*, 73(3), 371-377.

Shaw III, J. S., Garven, S., & Wood, J. M. (1997). Co-Witness Information Can Have Immediate Effects on Eyewitness Memory Reports. *Law and Human Behavior*, 21(5), 503-523.

Shaw, J. S. (1996). Increases in eyewitness confidence resulting from postevent questioning. *Journal of experimental psychology. Applied*, 2(2), 126-146.

Shaw, J. S., & McClure, K. A. (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior*, 20(6), 629-653.

Shaw, J. S., McClure, K. A., & Dykstra, J. A. (2007). Eyewitness confidence from the witnessed event through trial. In M. P. Toglia, J. D. Read, D. F. Ross & R. C. Lindsay (Eds.), *Handbook of Eyewitness Psychology, Vol. 1.* (pp. 371-397). Mahwah, NJ: Erlbaum.

Skowronski, J. J., & Walker, W. R. (2004). How describing autobiographical events can affect autobiographical memories. *Social Cognition*, 22, 555-590.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, 118(3), 315-327.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2007). When intended remembering leads to unintended forgetting. *Quarterly Journal of Experimental Psychology*, 60(7), 909-916.

Storm, B. C., Ligon Bjork, E., & Bjork, R. A. (2008). Accelerated Relearning After Retrieval-Induced Forgetting: The Benefit of Being Forgotten. *Journal of Experimental Psychology / Learning, Memory & Cognition*, 34(1), 230-237.

Tetlock, P. E. (1983a). Accountability and complexity of thought. *Journal of Personality and Social Psychology*, 45(1), 74-83.

Tetlock, P. E. (1983b). Accountability and the perseverance of first impressions. *Social Psychology Quarterly*, 46(4), 285-292.

Turtle, J. W., & Yuille, J. C. (1994). Lost but not forgotten details: Repeated eyewitness recall leads to reminiscence but not hypermnesia. *Journal of Applied Psychology*, 79(2), 260-271.

Tversky, B., & Marsh, E., J. (2000). Biased retellings of events yield biased memories. *Cognitive Psychology*, 40(1), 1-38.

Vandermaas, M. O., Hess, T. M., & Baker-Ward, L. (1993). Does anxiety affect children's reports of memory for a stressful event? *Applied Cognitive Psychology*, 7, 109-127.

Weingardt, K. R., Leonesio, R. J., & Loftus, E. F. (1994). Viewing eyewitness research from a metacognitive perspective. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 155184). Cambridge, MA: MIT Press.

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology*, 36(12), 1546-1557.

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, 48(5), 553-571.

Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology*, 83(3), 360-376.

Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology*, 56(2), 161-169.

Wells, G. L., Lindsay, R. C. L., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology,* 64(4), 440-448.

Wells, G. L., Olson, E. A., & Charman, S. D. (2002). The confidence of eyewitnesses in their identifications from lineups. *Current Directions in Psychological Science*, 11(5), 151-154.

Wessel, I., & Merckelbach, H. (1997). The impact of anxiety on memory for details in spider phobics. *Applied Cognitive Psychology*, 11(3), 223-231.

Wright, D. B., Mathews, S. A., & Skagerberg, E. M. (2005). Social recognition memory: The effect of other people's responses for previously seen and unseen items. *Journal of Experimental Psychology. Applied*, 11(3), 200-209.

Wright, D. B., Self, G., & Justice, C. (2000). Memory conformity: Exploring misinformation effects when presented by another person. *British Journal of Psychology,* 91(2), 189-203.

Yarmey, A. D. (1992). The Effects of Dyadic Discussion on Earwitness Recall. *Basic and Applied Social Psychology*, 13(2), 251 - 263.

Yates, J. F. (1994). Subjective probability accuracy analysis. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 381-410). New York: John Wiley & Sons.

Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology,* 71(2), 291-301.

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 934-945.

Effects of communication with a non-witness on eye-witnesses recall correctness and meta-cognitive realism

Paper I

Effects of repeated recall and discussion on eyewitness accuracy and meta-memory realism for different types of forensic information

Paper II

Content analysis of eyewitnesses repeated recalls and discussions

Paper III