Eyewitness testimonies

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

Sarwar, Farhan

2011

Farhan Sarwar

Eyewitness testimonies 2011

Eyewitness testimonies

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

Farhan Sarwar

ISBN: 978-91-978718-9-1

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

Faculty Opponent: Jan Andersson, Linkoping University, Sweden

Thesis Advisors: Carl Martin Allwood, Åse Innes-Ker

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

Printed in Sweden 2011 at Wallin & Dalholm, Lund.
ISBN 978-91-978718-9-1

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 minneutsagorna 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.

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

(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

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.

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. ct is the proportion of correct answers for all items at confidence level rt, and nt is the number of times the confidence level rt 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.

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,

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

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.

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

Applied Cognitive Psychology, *Appl. Cognit. Psychol.* (2010)
Published online (wileyonlinelibrary.com) **DOI**: 10.1002/acp.1749

Effects of Communication with Non-witnesses on Eyewitnesses' Recall Correctness and Meta-cognitive Realism

FARHAN SARWAR[1]*, CARL MARTIN ALLWOOD[2] and ÅSE INNES-KER[1]

[1]*Department of Psychology, Lund University, Sweden*
[2]*Department of Psychology, University of Gothenburg, Sweden*

Summary: In forensic contexts it is common that witnesses retell and discuss the experienced event many times. It is of forensic importance to understand how this influences memory and meta-memory. Eighty-nine participants viewed a short film and were assigned to one of four conditions: (1) Laboratory discussion (five discussions of the event with a confederate), (2) Family discussion (five discussions of the event with a family member), (3) Retell (five retellings of the event) and (4) Control. Three weeks later participants gave an open free recall, and then 3 days later confidence judged the recalled information. The results showed significant differences between the four conditions on number of correct items, incorrect items, accuracy, confidence and calibration. The results suggest that discussion of an experienced event 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. Copyright © 2010 John Wiley & Sons, Ltd.

Eyewitnesses often talk about a witnessed event in different contexts as well as with people with varying requirements and interests before they testify in court (Paterson & Kemp, 2006). The communication may take place at different intervals after the incident, and could occur, for example, in discussions with their family and friends and, in some instances, with health care professionals (Marsh, Tversky, & Hutson, 2005). In addition, witnesses often discuss the event with co-witnesses, crime investigators from the police and public prosecutors, other legal experts and sometimes the media (Gabbert, Memon, & Allan, 2003; Lane, Mather, Villa, & Morita, 2001). As an example, research indicates that children in USA are heard on average 11 times in the forensic process alone before they testify in the court (Christianson, 1994).

Because the listeners often have different interests, the witness may discuss the event from many different perspectives, focusing on different aspects of the event at different retellings (Hyman, 1994). In a detailed review, Skowronski and Walker (2004) argued that the main reason behind discussions is social, and the purpose of these discussions is to update their family and friends about what they called 'what's new'. Thus, the tellers try to make their stories relevant and understandable to the audiences, as well as leave room for feedback and, finally, try to make a point out of the whole narrative.

As a consequence, the witness' subsequent memory-performance will likely be impacted. For example, the way an event is retold has demonstrable consequences for the content of future memory recall (Hyman, 1994; Marsh, 2007; Skowronski & Walker, 2004). In addition, the level of confidence in one's assertions may increase with each reassertion of a claim, without necessarily improving accuracy (Hertwig, Gigerenzer, & Hoffrage, 1997).

The purpose of the present research was to investigate how witness' discussions with non-witnesses (persons who have not experienced the event) influence the witness' memory and meta-cognitive realism (henceforth: meta-memory realism) for overall information about an event.

Below, we first review relevant research on how communicating an event to others influences the accuracy of subsequent recall. Next, we discuss how the realism in one's confidence judgments is affected by multiple retellings of the event (i.e. how well does a participant's confidence that he or she is correct relate to actual accuracy). The final section summarizes the hypotheses for the present research.

Effects of retelling on memory accuracy

Memory of an event is not only affected by the event itself and the circumstances when it is reported but also by what takes place during the time between when the event is stored in memory and when it is reported. For example, memory is affected by how, and how often, the event is retold as well as by the questions a witness has to answer when the event is retold and discussed (Lane et al., 2001). These circumstances can both improve and distort memories. Loftus (1992, 2002) used the 'Trojan horse' metaphor in this context to express how information introduced in discussions, or introduced *via* media coverage, sometimes can dramatically affect the memory of the event.

There are important differences between discussing an event and merely retelling it. Retelling an event can be considered the same thing as *actively* repeating learned materials, that is, self-rehearsal without access to the original script. Active repetition tends to improve accuracy in later recall (Roediger III & Karpicke, 2006b). Roediger and Karpicke called the effect that recall is improved after taking a test *the testing effect* (Cull, 2000; Roediger III & Karpicke, 2006a, 2006b). Number of tests taken as compared to the same amount of rehearsals also results in superior performance on the final test (Karpicke & Roediger III, 2007, 2008). Moreover, the testing effect can also improve memory of material that was not tested but was related to the tested material (Chan, McDermott, & Roediger III, 2006).

Multiple retellings can also hamper the retrieval of information because of *retrieval induced forgetting* (Coman,

*Correspondence to: Farhan Sarwar, Department of Psychology, Lund University, Box 213, 22100 Lund, Sweden.
E-mail: farhan.sarwar@psychology.lu.se

Manier, & Hirst, 2009; MacLeod, 2002) but if the encoded content is well integrated there may be less or no retrieval induced forgetting (Anderson, 2003; Anderson & McCulloch, 1999). In the case of eyewitness' memory where the events of a crime scene are very well connected, retrieval induced forgetting might not have much effect. In their review of the testing effect, Roediger and Karpice (2006b) also discussed research on the effects of retrieval induced forgetting. They concluded that although repeated recall may create some retrieval induced forgetting, previous research suggests that the over-all result of repeated recall in situations that are similar to recall in forensically relevant contexts is likely to be an increase in the number of correct, recalled items.

In contrast, *discussing an event with others* introduces many opportunities for the memory to become changed and distorted. For example, when discussing an event, a memory schema that represents how similar events usually happen may be activated (Tversky & Marsh, 2000). During later recall, lost information may then be replaced by information from the event schema that was not in the original event but that was activated at the time of the earlier discussion (Marsh, 2007; Tversky & Marsh, 2000). Although similar schema dependent effects might happen in the context of mere retellings this appear likely to happen only to a lesser extent. Pasupathi (2001) noted that both the qualities of the teller and the listener and their prior discussions define the context of their conversation and guide people as to what to share and how to react to the individual event memories. Also, when other people react to the witness' story they may communicate incomplete and misleading information (Loftus, 1979). Thus, such information may be introduced when the witness discusses the event with another witness (Alper, Buckhout, Chern, Harwood, & Slomovits, 1976; Gabbert et al., 2003; Hollin & Clifford, 1983; Luus & Wells, 1994; Marsh, 2007; Tversky & Marsh, 2000). In addition, co-witnesses that are acquainted with one another tend to incorporate their co-witness account into their own recall to a greater degree than co-witnesses that are not previously acquainted, resulting in lower accuracy in their recall (Hope, Ost, Gabbert, Healey, & Lenton, 2008).

Research shows that information supplied by others can have a negative effect on the witness' memory (Loftus, 1992; Nourkova, Bernstein, & Loftus, 2004; Wright, Self, & Justice, 2000). For example, listeners' recollections may affect witness' later recollections (Loftus, 2003). When witnesses forget details of the event they can compensate with memories from their previous listeners' recollections (Wright, Mathews, & Skagerberg, 2005). People sometime discuss speculative contra-factual possibilities, and such discussions can later affect witness' judgments of the discussed event (Wells & Gavanski, 1989). Leading, misleading or confusing questions from discussion partners can lead to deteriorations in witness' memory (Kebbell & Johnson, 2000; Loftus, 1975). This effect 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 a witnessed event with family and friends may have stronger negative effects on the correctness of subsequent recall than the more official discussions that take place within the legal system. The reason is that speakers understand the goals of the listener and try to meet their expectations (Russell & Schober, 1999). When witnesses tell an event to the forensic professionals such as investigators and lawyers, or testify in court, they may feel under pressure to stick to information which they are sure is true. However, when they share the experienced event with their friends, they may provide their own more free reactions and conclusions. This means that witnesses may use a lower certainty level when they decide which memories to share with their friends and family (Koriat & Goldsmith, 1994, 1996; Roebers, Moga, & Schneider, 2001). A consequence may be that witnesses report more incorrect memory details in that context as compared with the legal process. Moreover, family and friends may react to the witness' account more subjectively and contribute their own similar experiences during the discussion (Dritschel, 1991).

To sum up, intermediate repeated recalls of an event can affect the memory of the experienced event both as an effect of it merely being retold and also as an effect of the event being discussed. Similarly, as discussed next, the realism of the confidence with which the witness assert their memory may be affected by multiple retellings and discussions of the witnessed event.

Realism in confidence judgments

The confidence expressed by the witnesses about the accuracy of their reported memories is another essential type of evidence in the forensic process. Such judgments are called *confidence judgments* and are important when the court considers the correctness of the witness' statements (Allwood, Granhag, & Johansson, 2003; Wells & Bradfield, 1999). In the present study we analysed the *realism* of the witness' confidence judgments, i.e. the match between the witness' confidence and the accuracy of their memory statements (Yates, 1994).

The retelling of memories can affect witnesses' judgment of how sure they are about the correctness of their memory. As noted above, previous research shows that multiple retellings of an event can lead to increased feelings of confidence while memory correctness does not increase. This is called 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). In studies where confidence has been found to increase over many sessions this may depend on the reiteration effect (see, Granhag, Stromwall, & Allwood, 2000). However, research also shows that confidence may be tempered in social contexts due to the *accountability effect* (Tetlock, 1983), that is, that people may lower their confidence when they consider that they will be held accountable for correctness of their statements by other persons, for example in a discussion of the event, or when testifying in court.

In research on the realism in confidence judgments the relationship between correctness and confidence has in

forensic contexts traditionally been measured by the point bi-serial correlation (especially in research on lineups). A common result from these studies is that the level of confidence has a relatively weak relation to the correctness of witness' identifications (Bothwell, Deffenbacher, & Brigham, 1987; Luus & Wells, 1994; Sporer, Penrod, Read, & Cutler, 1995; Wells, 1993). However, other researchers have noted weaknesses in this method as a *general* indicator of realism in confidence judgments. Juslin, Olsson, and Winman (1996) pointed out that the correlation size in a partly non-relevant way depends on the spread of the confidence judgments over the total confidence judgment scale and that this measure primarily picks up witnesses' ability to discriminate correct from incorrect reports in contrast to the witnesses' tendency to be over- or under-confident (see also e.g. 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 that gives a more differentiated and informative understanding of realism in confidence judgments.

Overview of the present study

The present research investigated how discussing an experienced event with others may influence both accuracy and confidence in subsequent recall. Participants first viewed a short video of a kidnapping. They were then separated into four conditions (see details below). Three weeks after the initial viewing, participants were asked to recall the events freely. Finally, about 3–4 days after their recall, they confidence judged their responses.

The experiment consisted of four conditions: (1) retell the witnessed event five times in the laboratory to an unknown (each time new) person who posed questions about the event which the witness answered (*Lab-discussion condition*), (2) retell the witnessed event five times to one's own family and friends (each time to a new person) who posed questions about the event which the witness answered (*Family discussion condition*), (3) retell the witnessed event five times in the laboratory to a person who did not pose any questions (*Retell condition*) and (4) a *Control condition* where no retelling or discussion before the final recall took place.

The purpose of the Retell condition was to set the level for the effects on the outcome measures of merely retelling the event. In this way, the additional effects of the discussion aspect of the Lab-discussion condition could be better understood. The critical condition was thus the *Lab-discussion* condition that intended to investigate how discussions may impact recall and confidence. This condition had a relatively high level of control since it took place in the same setting each time, and that the discussions were recorded. At the same time, the discussions were less controlled than can be expected in a forensic interview in that the confederates had rather loose instructions on how to perform the discussion, and all questions the confederate asked were unprepared. Thus, it resembles the type of informal discussions one can assume witnesses perform with friends and family during the time between a witnessed event and more formal testimonies. Still, bringing participants to a laboratory to discuss a film with an unknown person under measured conditions may alter the experience and memory enough so that the results would not fully address the question whether informal discussions with family and friends may impact recall and meta-memory realism. Thus, we included the *Family discussion* condition, which was designed to minimize measurement-intrusion and enhance ecological validity.

Hypotheses

Hypothesis 1

The overall accuracy (proportion correct of all items recalled) for the participants in the Retell condition was predicted to be higher than for the participants in the other three conditions. Although participants in all conditions except the Control condition retell the information the same amount of times, the presence of discussion partners in both discussion conditions could introduce materials, such as questions or erroneous assertions about the event that will influence accuracy negatively, as discussed above. Moreover, the participants in the Retell condition, in contrast to those in the Control condition, are likely to benefit from retelling the material (Roediger III & Karpicke, 2006b).

Hypothesis 2

It was hypothesized that the confidence level would be especially high in the Retell condition since according to Hypothesis 1 the accuracy level was expected to be higher in this condition. Above we mentioned that confidence and accuracy are not always highly correlated. However, most eyewitness studies still show some positive association between confidence and accuracy. Moreover, as a consequence of the reiteration effect (Hertwig et al., 1997), the confidence level in all the three conditions where retelling occurred was expected to be higher than in the Control condition.

Hypothesis 3

It was also expected that the participants in the Retell condition would show the best realism in their confidence judgments. Given that the effect of active rehearsal (i.e. increased accuracy) has been found to be more reliable in previous research than the reiteration effect (increased confidence), we expected the Retell condition would also show better realism than especially the Control condition.

METHOD

Participants

Eighty-nine undergraduate students (62 women) from Lund University participated in the study. The mean age of the participants was 25, ranging from 18 to 47 years. Each participant received a movie ticket worth 90 SEK (approximately US$12).

Initially, 23 participants were recruited for each condition. Participants who left the study prior to completion were replaced. The dropout rates were somewhat uneven with

seven, six, four and one participants leaving the Lab-discussion, Family discussion, Retell and Control condition, respectively. These were all replaced. Subsequently, there were two more dropouts from the Lab-discussion condition and one more from the Control condition that were not replaced. A χ^2-test performed on the total number of participants recruited for each condition (30, 29, 27 and 24) did not reach significance ($X^2[3, N = 110] = .074$, $p = \text{ns}$, $\varphi = .08$), suggesting that we still could assume random assignment of participants to the different conditions.

Design

The experiment consisted of four between-subjects conditions: Lab-discussion ($n = 21$; one participant only attended four sessions out of five), Family discussion ($n = 23$), Retell ($n = 23$) and Control condition ($n = 22$). These were described above.

Material

Videotape

A 3 minutes and 50 seconds long colour film showing a kidnapping of a woman by two men at the bus stop was used. The film was shown on a 28 inch colour television and has been used in previous research (e.g. Allwood, Ask, & Granhag, 2005; Granhag, 1997).

The film is shot from an eyewitness perspective and first shows a woman at a bus stop. A woman passing by checks the bus timetable and leaves. A car stops by the bus stop and two men appear from the car. They proceed to kidnap the woman. She resists but finally she is overpowered. The woman's handbag falls on the ground. One man picks it up. When the man is picking up the handbag from the sidewalk, the witness (the camera perspective) attempts to have a closer look. The man pulls out a revolver and threatens the witness who retreats instantly. The man then picks up the bag and returns to his partner. They put the woman in the car and drive away.

Confidence judgment scales

The confidence scale was used by the participants to rate their confidence for the detailed parts of the free recalls (low-level statements). It used 11 levels beginning at 0% ('Completely sure that I remember wrong') and then 10, 20, 30, ... 100% ('Completely sure that I remember correct').

Procedure

The participants were informed that the research was about human perception in different forensic situations. Groups of between four and eight participants first watched the video film. Next, the participants were randomly assigned to one of the four conditions. In three of the conditions the participants were given a time schedule for the five occasions when they would return to the lab.

Participants in the Lab-discussion condition returned to the lab five times over a 20-day period. In each session, the participant discussed the content of the movie with a confederate. For this purpose 105 individuals (i.e. the confederates) were recruited who would act as discussion partners for the participants. The discussion partner's assignment was to first listen when the participant was telling the events of the film and then to ask unprepared questions about the film. The questions asked were to be constructed in such a way that the discussion partner could understand the complete course of events in the film. Each discussion partner performed only one discussion. All discussions were recorded on a MP3 recorder.

In the Family discussion condition the participants discussed the contents of the film five times with either a family member or a friend. Participants were instructed to first give an account of the film and then to discuss the contents of the film with the discussion partner. The participants should discuss the film with a new family member or friend each time. Discussion partners could ask questions, or share relevant experiences, if he or she had any. The participants in this condition were asked 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 had not received an SMS by 7:00 pm he called the participants and reminded him/her of the task. As mentioned we were interested in the spontaneous and natural discussion between witnesses and non-witnesses and its impact on recall and realism. Thus, we aimed for minimal interference in this process (for example only instruction of when to do the discussion and to alert the experimenter when done). We felt that the use of a recording device could make the participants and their partners conscious about their discussion (and also artificially prolong the time spent at the assigned task) and consequently render this condition not very much different from the Lab-discussion condition.

Participants in the Retell condition returned to the laboratory a total of five times over a 20-day period. Each time, the participant simply told the story about the film to the experimenter. No questions were asked. Participants were instructed to tell whatever they remembered about the film they had seen in the first visit. All rehearsals were recorded on a MP3 recorder. In the Control condition the participants were simply told they were not to talk about the contents of the film with anybody.

On the 21st day all participants returned to the lab for their free recall test, where they typed in Microsoft Word whatever they remembered about the incident that they had witnessed in the short film. They were instructed to type as many details as they could remember.

Finally, the participants came to the laboratory on the 24th or 25th day to give their confidence judgments. First the participants were asked to give their current confidence judgments on their free recall statements, which had been prepared as explained below. Participants were then debriefed, thanked and dismissed.

The sessions for the individual participants in the Retell and the Lab-discussion (and the Control condition) often took place in different rooms over the five times and for the recall and confidence ratings sessions and also compared

with the room where the film was seen. The rooms used depended on the rooms available.

Preparation of material for participants' confidence judgments

In order to prepare for the confidence judgments, the free recall of each participant was broken down into single units of information. We followed the principles described by Allwood et al. (2005) when dividing the free recall into single units (pieces of information). The following principles were used: (1) those statements that were about actors and actions carried out were rendered as one unit. For example 'a car passed by' was used as a single unit. (2) An object with one attached attribute was used as one unit. For example 'the car was blue' was used as one unit. (3) If an object was described by more than one attribute, the additional attributes were rendered as separate units. For example 'the tall blond woman' was rendered as two units. For the purpose of reminding the participants about the context in which they mentioned each unit, one or two short sentences related to that item were attached to that specific item. The items to be confidence judged were underlined while the reference items were put in brackets. Finally, an 11-point confidence scale was inserted directly below each single item.

Measurements

Three measures were computed to analyse the realism in the participants' confidence judgments: calibration, over-/underconfidence and slope. *Calibration* and *over-/underconfidence* measure aspects of the relation between the participant's *level* of confidence and accuracy. Calibration indicates the individual's squared deviance from perfect calibration averaged and weighted over all confidence levels (e.g. . . . 50, 60, 70, ...) divided by the number of items. Over-/underconfidence is calculated by taking the difference between the participant's average confidence and his/her percentage correct items. For both calibration and over-/underconfidence the value 0 shows perfect realism and for the over-/underconfidence measure positive values signal overconfidence and negative values underconfidence.

Slope measures the participant's ability to separate their correct answers from their incorrect answers by means of the level of their confidence judgments and is computed by subtracting the mean level of confidence for a participant's incorrect items from the mean confidence level for their correct items. Thus, greater separation is indicated by a larger slope value. The slope measure and its' outcome have an advantage in that they are intuitively easy to understand.

RESULTS

Calibration curves

Figure 1 shows the calibration curves for the four conditions. The *x*-axis shows the 11 different confidence levels (from 0 to 100%) and the *y*-axis shows the percent of correct answers. The numbers inside the graph give the percent of answers for each confidence level in each condition. The diagonal shows perfect calibration. As can be seen in Figure 1, the calibration curves for the four conditions show very low percentages for the confidence scores between 0 and 50. Consequently, these levels will not be further discussed. For confidence score 50 and above the scores are clustered close to the diagonal, indicating that confidence is fairly realistic. Importantly, about half of the confidence judgments are at the 100% level. All groups showed some overconfidence at the 100% confidence level. The Retell and the Family condition tended to show underconfidence at confidence levels below 100%, while the other two conditions showed overconfidence.

ANOVAs

Means and standard deviations for the four experimental conditions for correct items, incorrect items, accuracy (proportion correct statements), confidence, calibration, over-/underconfidence and slope can be found in Table 1. Each of the dependent measures, correct items, incorrect

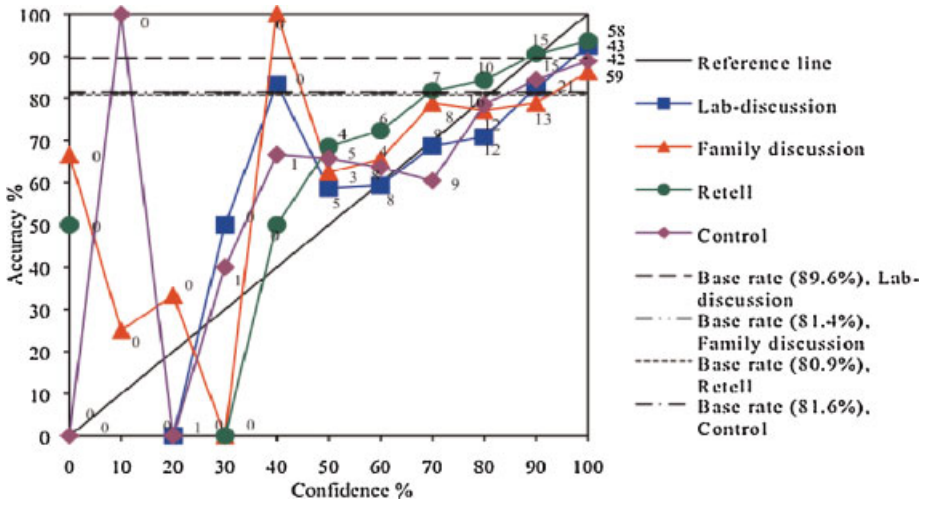

Figure 1. Calibration curves for the Lab-discussion, Family discussion, Retell and the Control condition. The numbers next to each point indicate % responses that used this confidence level at this particular accuracy level; 0 means that less than 1% of the total responses used this confidence level at this accuracy level. The total number of responses (*N*) for Lab-discussion condition, Family discussion condition, Retell condition, and Control condition are 1220, 1099, 1062 and 785, respectively.

Table 1. Free recall: Means (and SDs) for correct items, incorrect items, accuracy, confidence, calibration, over/underconfidence and slope. *F*-values and η^2 for the corresponding ANOVAs

	Lab-discussion	Family discussion	Retell	Control	*F*	η^2
Correct items	46.95 (15.83)	38.74 (15.81)	41.04 (12.68)	28.23 (11.01)	6.8^{***}	.19
Incorrect items	11.14 (7.23)	9.04 (5.37)	5.13 (3.43)	7.45 (5.84)	4.55^{**}	.14
Accuracy	.80 (.10)	.81 (.08)	.89 (.05)	.81 (.09)	5.69^{***}	.17
Confidence	85.77 (6.05)	88.23 (4.67)	89.61 (5.41)	84.12 (7.05)	3.95^{**}	.12
Calibration	.03 (.03)	.04 (.02)	.02 (.01)	.04 (.02)	3.44^{*}	.10
O/U confidence	.04 (.10)	.06 (.09)	.00 (.07)	.03 (.10)	2.17	.07
Slope	12.99 (8.8)	11.65 (13.99)	22.22 (28.49)	14.01 (20.14)	1.37	.05

Note: O/U confidence = Over-/underconfidence.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

items, accuracy, confidence, calibration, over-/underconfidence and slope was submitted to separate between-subjects one-way ANOVAs. Significant effects were found for correct items, $F(3, 85) = 6.8$, $p < .001$, $\eta^2 = .19$, incorrect items, $F(3, 85) = 4.55$, $p < .005$, $\eta^2 = .14$, accuracy, $F(3, 85) = 5.69$, $p < .001$, $\eta^2 = .17$, confidence, $F(3, 85) = 3.95$, $p < .01$, $\eta^2 = .12$, and calibration, $F(3, 85) = 3.44$, $p < .02$, $\eta^2 = .10$. There were no significant differences between the groups for the over-/underconfidence and slope measures. To further probe the significant differences, planned pairwise contrasts were performed. The results, with Bonferroni corrected significance levels, are presented below.

Number of correct items
Participants in the Lab-discussion condition and Retell conditions recalled significantly more correct items than those in the Control condition (Lab-discussion *vs*. Control: $t(85) = 4.32$, $p < .02$, Cohen's $d = 1.42$; Retell *vs*. Control: $t(85) = 3.08$, $p < .01$, Cohen's $d = 1.10$). The difference in number of correct items between the Family discussion and Control discussion did not quite reach significance ($p = .081$). The number correctly recalled items did not differ between the three experimental conditions.

Number of incorrect items
Participants in the Lab-discussion reported significantly more errors than those in the Retell condition, $t(85) = 3.56$, $p = .001$, Cohen's $d = 1.10$. No other differences reached significance.

Accuracy
As predicted in hypothesis 1, accuracy was higher in the Retell condition than in the Lab-discussion condition ($t(85) = 3.41$, $p < .002$, Cohen's $d = 1.18$) and higher than in the Family discussion condition ($t(85) = 3.29$, $p = .009$, Cohen's $d = 1.23$). Accuracy was also higher for the Retell condition than for the Control condition ($t(85) = 3.37$, $p = .007$, Cohen's $d = 1.13$). The Lab- and Family discussion conditions did not differ significantly from the Control condition, which is also evident when inspecting the means.

Table 2. Means (and SDs), *t*-values, and corresponding Cohen's *d* for the number of confabulations reported during the five sessions in the lab-discussion and the retell condition and in the final free recall

	Lab-discussion	Retell	*t*	Cohen's *d*
Session 1	1.33(1.53)	.13(.34)	-3.68^{***}	1.13
Session 2	1.71(1.95)	.22(.67)	-3.46^{***}	1.07
Session 3	1.43(1.57)	.35(.65)	-3.04^{**}	.94
Session 4	1.29(1.45)	.17(.49)	-3.46^{***}	1.08
Session 5	1.78(2.10)	.14(.47)	-3.57^{***}	1.13
Final free recall	1.14(1.39)	.35(.78)	-2.37^{*}	.73

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Confidence
Participants in the Retell condition were more confident than those in the Control condition ($t(85) = 3.15$, $p = .025$, Cohen's $d = .89$). No other differences reached significance.

Calibration
Participants in the Retell condition were better calibrated than those in the Control condition ($t(85) = 2.79$, $p < .05$, Cohen's $d = -1.30$). No other differences were significant, although the difference in calibration between the Retell condition and the Family discussion condition approached significance ($p = .081$).

Confabulations reported in the Retell and Lab-discussion conditions

Information reported by the participants that was not present in the film were considered confabulations (for the results see Table 2). For example, it was considered a confabulation if the participants reported that the offenders were wearing masks although they were not wearing masks. The participants in the Lab-discussion condition reported significantly more confabulations than the participants in the Retell condition in session 1 ($t(42) = -3.68$, $p < .001$, Cohen's $d = 1.13$), session 2 ($t(42) = -3.46$, $p < .001$, Cohen's $d = 1.07$), session 3 ($t(42) = -3.04$, $p < .004$, Cohen's $d = .94$), session 4 ($t(42) = -3.46$, $p < .001$, Cohen's $d = 1.08$), session 5 ($t(42) = -3.57$, $p < .001$, Cohen's $d = 1.13$) and in the final free recall session ($t(42) = -2.37$, $p < .022$, Cohen's $d = .73$).

Content analysis of the confederates' questions in the Lab-discussion condition

Although data only existed for one condition, we found it illuminating to look at the number of and type of questions the confederates asked in the Lab-discussion condition. The mean number of questions asked in each session was 14.8 ($SD = 3.8$). Most of these, 13.9 (3.7), were judged to be

relevant questions. Of the relevant questions 5.4 (2.3) were classified as central questions (questions about the facts related to the story that cannot be changed without changing the story) and 8.5 (2.9) as peripheral questions (questions about the facts related to the story that can be changed without changing the story. The division between central and peripheral questions was based on the 'plot relevancy' model by Heuer and Reisberg (1990).

DISCUSSION

The purpose of this study was to investigate how eyewitnesses' retelling and informal discussions with non-witnesses (persons who were not present at the event) can affect both the witnesses' memory and the realism in his or her meta-memory. Participants viewed a video and then, over a 3 week period, either (a) discussed the events with a discussion partner in a laboratory setting several times, (b) discussed the events with a family member or friend several times, (c) retold the events to a listener in a laboratory setting several times or (d) did not retell or discuss the events. Memory and meta-memory performance were then assessed.

The first hypothesis stated that accuracy in the Retell condition would be higher than in the three other conditions; the reasons being that active repetition (i.e. retelling in one's own words) is known to improve accuracy (Roediger III & Karpicke, 2006b) and that the input from the discussion partners in the two discussion conditions could have negative effects on the recall accuracy. The results did indeed show higher accuracy in the Retell condition than in the other conditions, thus supporting the first hypothesis. Our analysis showed that the participants in the Lab-discussion condition reported more confabulations as compared with the Retell condition. These confabulations obviously decreased accuracy in the Lab-discussion condition. The confabulations may to a large extent have been the result of the questions posed to the participants in the Lab-discussion condition (on average about 14 per session). These questions provided a possibility for implicit but erroneous assumptions on behalf of the questioner to affect the participant's memory. However, it is beyond the scope of the present paper to satisfactorily evaluate the extent to which the discussions in the Lab-condition in different ways contributed to the participants' accuracy in recall.

More interesting from a forensic standpoint is that the accuracy levels for both discussion conditions did not differ from the Control condition. For the Lab-discussion condition this is particularly interesting. Despite having the greatest number of correct items in absolute numbers, this condition did not differ in accuracy (the proportion correct items) from the Control condition. The reason was that the Lab-discussion condition also had a higher number of incorrect items than the control condition. This result supports that discussing an event with other people can increase the total number of correct items but can also affect the witness' recall by introducing incorrect information (Loftus, 2003; Luus & Wells, 1994; Marsh, 2007; Tversky & Marsh, 2000).

The second hypothesis predicted that confidence would be higher in the three discussion/Retell conditions than in the Control condition. Empirically this prediction was partly supported by the results since we found a significant difference between the Retell and the Control condition. The confidence displayed by the participants in the Lab-discussion group and the Family discussion condition did not significantly differ from those in the Control condition, however. It should be noted that these results do not clearly support the presence of a reiteration effect in the Retell condition (Hertwig et al., 1997) because a more likely conclusion is that the confidence was high in the Retell condition because the accuracy was high in the same condition. The effect sizes were also rather small.

If anything, there is some indication of a reiteration effect in the Lab- and Family discussion conditions because the accuracy level in these conditions was not higher than in the Control condition whereas the level of confidence showed a trend towards higher levels than in the Control condition and particularly in the Family discussion condition although not reaching conventional levels of confidence. Consider, however, that the Bonferroni correction may be overly cautious. The uncorrected levels indicated that the differences were significant. Speculatively, the effects of making an assertion may be promoted by making the assertion in a full social context (i.e. not just retelling the event to the experimenter as in the Retell condition) but at the same time 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, 1983). 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.

The third hypothesis predicted that the participants in the Retell condition would show better realism in their confidence judgments compared with the two discussion conditions. This was expected since the effect of active repetition on recall-accuracy has been found to be more stable in previous research (Roediger III & Karpicke, 2006b) than the reiteration effect (Hertwig et al., 1997). The hypothesis was supported for the calibration measure since the results showed that the participants in the Retell condition were better calibrated than the participants in the Control condition. The effect size was moderate. The results for the over-/underconfidence and the slope measures were in the same direction as for calibration, although not significant at the omnibus level.

No significant differences in results were found between the two discussion conditions (the Lab- the Family discussion conditions). Thus, although the Family discussion condition, compared with the Lab-discussion condition, can be assumed to have been more informal and that the participants in this condition for this reason may not have felt subjected to the same scrutiny as the participants in the Lab-discussion condition, this did not give rise to any substantial differences in the results. One possibility is that the many retellings included in our study acted to 'wash out' potential initial differences between the conditions. It remains for future research to evaluate this possibility.

At a general level, it can be noted that calibration for the free recall in all conditions indicated good calibration (none of the conditions showed a level of overconfidence that differed significantly from perfect realism, that is, zero). This suggests that it can sometimes be reasonable to use confidence as a criterion for correctness of free recall. These results for free recall are in line with the results from previous research (e.g. Allwood et al., 2005; Allwood, Innes-Ker, Homgren, & Fredin, 2008). Here it can be noted that the lack of significant results for the over-/underconfidence measure is obviously not a floor effect since the scale ranges from over- to underconfidence. When individuals freely chose what to report (i.e. the free recall instructions in this study) they, at least in the situations studied, appear to be able to monitor their recall so that their performance shows good accuracy and confidence judgments well in line with the level of accuracy, that is, an important aspect of meta-memory realism.

Limitations

As in all studies there are various limitations to the present results. Two such limitations will be considered here. The first one concerns the choices of conditions and the manner in which some of them were measured. The second concerns the delay between the time of final recall and the subsequent confidence judgments.

The main purpose of the research was to further understand how discussing a witnessed event with multiple others across a relatively large time-span may influence both recall and meta-memory realism. The selection of the Lab-discussion condition and the no-discussion Control condition is relatively straightforward. However the designs of the Retell condition and the Family discussion conditions warrant some discussion.

The purpose of the Retell condition was to control for the effect that simple re-telling of an event (without any discussion input) might have on memory and meta-memory. Thus, it seemed that simply asking participants to return to the lab to retell the events to a single individual, and one somewhat in an authority position (the experimenter) would be appropriate. But, as shown in previous research reviewed in the Introduction, humans are sensitive to who they talk to, and dynamically adjust their narratives to this. It is quite possible that recall might have been different if they had to re-iterate the event to a tape-recorder; a different listener each time; individuals of different status; or individuals of differing familiarity. However, at the time of the design, it was of some importance to control the variance in what in essence is a control condition, and to ensure that the participants did do the retell task. For that purpose a single listener seemed the optimal solution. Further research is needed to investigate the potential differences in subsequent memory performance as an effect of retelling to different types of listeners.

We next discuss the Family discussion condition, where participants were simply asked to discuss the events with familiar others at five set times, and to then alert us when this was done. The purpose was ecological validity, since one of the aspects we are interested in is how this type of spontaneous informal discussion with family and friends about a witnessed event may impact testimony both when it comes to accuracy and confidence. As such, we deliberately kept the condition largely uncontrolled, with the exception of the schedule for discussing the events, and the SMS verification. We felt that introducing more control, in the form of a recording or protocols would likely alter the performance to such an extent that it no longer could address that particular research question. As mentioned, additional controls would make this condition more similar to the Lab-discussion condition, and thus it would not serve any particular purpose. We are highly aware of the uncontrolled nature of this condition, that we have no idea of the variance both in duration and contents of the discussions or even whether the participants actually complied and did not simply respond that they had done the discussion to placate the experimenter. However, their responses can be seen as some kind of signature of their performance. The fact that the participants in the Family discussion condition appears closer in their performance to those in the other discussion/retell conditions than to those in the no-discussion Control condition suggests that they did at least partially comply. Of course, the Family discussion condition also differed in a possibly important way from the Lab-discussion and the Retell conditions in that they talked to people with whom they were highly acquainted. This feature is of interest since it mimics a common situation in the forensic context and, as we know from the co-witness literature, this may impact both accuracy and confidence in a different manner than discussing the event with unfamiliar others (Hope et al., 2008). Further research is needed to more specifically sort out the influences of the various differences between the Lab-discussion and the Family discussion conditions.

For the second limitation, it could be argued that since the confidence judgments were not given immediately after the memory recall task was carried out their result is influenced by various factors not relevant to the confidence judgment and are not representative for common forensic situations. However, presumably this argument is more valid for lineup situations where both in the US and, for example Sweden it is stipulated that the witness confidence in their identification response should be recorded immediately after the response. This recommendation is based on conclusions from research on lineups (Brewer & Burke, 2002; Brewer & Wells, 2006). However, we would argue that this argument may be less urgent in the case of event memory and for the present study, given the following observations. First, the confidence judgments were given with respect to assertions that were 'in front of' the participants at the time they made their confidence judgments (the participants' open free recall statements). This is likely to be a common situation, for example in court ('you have earlier said that . . .') or after the police have received new information, that they will return to a witness to ask how sure they are about their memory of what they reported. Second, even if the confidence judgments are given a few days after the participants' memory recall, the most commonly occurring situation in forensic contexts may be that witnesses have discussed the experienced event with other people many times after the event (Paterson & Kemp, 2006). Each time the event is retold

the witnesses is likely to have made a confidence judgment about whether he or she is sufficiently sure about the memory to be willing to retell it to the current audience. This assumption is at the fundament of the memory model presented by Koriat and Goldsmith (1996) which deals with the role of confidence judgments in memory retrieval of the kind investigated in the present study. According to the model, each time before a memory is reported the individual spontaneously carries out a confidence judgment. The function of this confidence judgment is to test that the individual remembering is sufficiently sure about the correctness of the memory to want to report it in the current social context.

From this reasoning it seems to follow that at the time of later recalls, that is in court and police interviews after the witness has retold the event to family and friends or discussed it with other witnesses, one or more confidence judgments of the same memory have already been made. In this situation it would not seem likely that some days delay between the final recall and the confidence judgment as in our study would have much influence on the level and realism of the confidence judgments compared with what occurs in many or most forensic situations.

Conclusions

To sum up, talking about a witnessed event is likely to increase the amount of information volunteered at a later time. This effect was particularly strong for the Lab-discussion condition. Although correct information obviously is to be preferred, it may, in general, be forensically beneficial, although risky, to get more information compared to less information from a witness even if it is not all correct, because it may provide cues that can be used to further probe and check the witness' statements. However, when looking at accuracy, the Lab-discussion condition did not outperform the conditions that never discussed the information. Instead it was the Retell condition that emerged, showing superior accuracy. Accordingly, and in brief, our results show that talking to other people about an experienced event does not necessarily *lower* the accuracy and the meta-memory aspects of recall, although it may obviously lead to that specific, potentially important details are altered in recall.

Our results also showed that merely retelling information (the Retell condition) tended to increase one's confidence in the information. (How repeated *written* description of an experienced event affects recall is a topic for future research.) However, the present results also showed that the same participants were better able to adjust their confidence level to the level of their accuracy, that is, they showed better performance for calibration. In contrast, the participants in the two discussion conditions are also likely to have been affected by the discussions they participated in. These may have acted to somewhat decrease their level of confidence, compared with the Retell condition (maybe as an effect of lower accuracy and social accountability). At the same time, an important finding was that discussing the event with others did not affect the realism in the confidence judgments adversely, as compared with the Control condition.

REFERENCES

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

Allwood, C. M., Granhag, P. A., & Johansson, M. (2003). Increased realism in eyewitness confidence judgements: The effect of dyadic collaboration. *Applied Cognitive Psychology*, *17*, 545–561.

Allwood, C. M., Innes-Ker, Å., Homgren, J., & Fredin, G. (2008). Accuracy and realism in confidence in children's and adults' event recall in response to free recall and focused questions. *Psychology, Crime & Law*, *14*, 529–547.

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*, 415–445.

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*, 608–608.

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

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*, 353–364.

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

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

Christianson, S.-Å. (1994). *Traumatiska minnen [Traumatic memories]*. Stockholm: Natur och Kultur.

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

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

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

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

Granhag, P. A. (1997). Realism in eyewitness confidence as a function of type of event witnessed and repeated recall. *Journal of Applied Psychology*, *82*, 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*, 397–420.

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

Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition*, *18*, 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*, 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*, 476–484.

Hyman, I. E., Jr. (1994). Conversational remembering: Story recall with a peer versus for an experimenter. *Applied Cognitive Psychology*, *8*, 49–66.

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*, 1304–1316.

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

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science*, *319*, 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*, 629–641.

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*, 297–315.

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

Lane, S. M., Mather, M., Villa, D., & Morita, S. K. (2001). How events are reviewed matters: Effects of varied focus on eyewitness suggestibility. *Memory & Cognition*, *29*, 940–947.

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

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

Loftus, E. F. (1992). When a lie becomes memory's truth: Memory distortion after exposure to misinformation. *Current Directions in Psychological Science*, *1*, 121–123.

Loftus, E. F. (2002). Memory faults and fixes. Retrieved 20 October 2005, from Publication of the National Academy of Sciences. Available at: http://www.issues.org

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

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

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

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

Marsh, E. J., Tversky, B., & Hutson, M. (2005). How eyewitnesses talk about events: Implications for memory. *Applied Cognitive Psychology*, *19*, 531–544.

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

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

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

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*, 313–329.

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

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

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

Shaw, J. S., III. (1996). Increases in eyewitness confidence resulting from postevent questioning. *Journal of Experimental Psychology: Applied*, *2*, 126–146.

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

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*, 315–327.

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

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

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. 155–184). Cambridge, MA: MIT Press.

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

Wells, G. L., & Bradfield, A. L. (1999). Distortions in eyewitnesses' recollections: Can the postidentification-feedback effect be moderated? *Psychological Science*, *10*, 138–144.

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

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*, 151–154.

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*, 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*, 189–203.

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

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

Paper II

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

Farhan Sarwar[1], Carl Martin Allwood[2], and Åse Innes−Ker[1]
[1]Department of Psychology, Lund University, Sweden
[2]Department of Psychology, University of Gothenburg, Sweden

Correspondence concerning this article should be addressed to Farhan Sarwar, Department of Psychology, Lund University, P.O. Box 213, SE-22100 Lund, Sweden. Tel.: +46 46-222 05 72; fax: +46 46-222 42 09. E-mail address: farhan.sarwar@psychology.lu.se

Abstract

Two experiments investigated the recall and the meta-memory performance for Forensically central information (e.g., what happened and how it happened at the crime scene), Forensically peripheral information (e.g., description of the suspect) and Non-forensic information for two types of questions. Experiment 1 studied the influence of eyewitness discussions with non-witnesses for open free recall and for focused questions. Eighty-nine participants watched a short film and participated in one of 4 conditions: 1) Laboratory discussion, 2) Family discussion, 3) Retell and 4) Control. Three weeks later participants did the recall tests, then 3 days later confidence judged their memory reports. The participants showed better memory and meta-memory realism for Forensically central and Non-forensic information than for Forensically peripheral information. Discussions and retellings caused participants to recall a higher total number of items and number of correct items for Forensically central information as compared to no discussion and retelling. In Experiment 2, 77 participants first watched a short film and later answered 63 focused questions on Forensically central details and on Forensically peripheral details of the viewed events. Here the participants showed better memory for Forensically central information but about equal meta-memory realism for the two types of information. In brief, the two experiments show the vulnerability of Forensically central information to witness discussions and of peripheral information to the context provided by the other questions asked.

Keywords: Eyewitness; Confidence; Meta-memory; Free recall; Focused questions; Retellings; Forensic information; Central information; Peripheral information.

Introduction

In the beginning of a crime investigation, the police are likely to ask for as accurate description as possible of the crime event: what happened, how it happened, description of offenders, description of objects used, and the time and place of incidence. Questions regarding what happened and how it happened can be answered by describing the central details (or action details) of the crime scene (e.g., the suspect pointed a gun at the victim, the victim tried to run). Questions regarding offenders, objects, time and place can be answered by describing the peripheral details (or descriptive details) of the crime scene (e.g. what kind of weapon was used, what kind of vehicle was used, what were the offenders wearing, what were their ages and heights).

Unfortunately, eyewitnesses are often not good in providing the required information, and different variables influence witnesses performance. The present paper investigates whether Forensically central or peripheral information is associated with better memory and meta-memory performance and whether this is influenced by discussing the witnessed event repeatedly or by question type. Two types of interview questions are investigated in this context: Open free recall questions (Experiment 1) and focused questions (Experiment 1 and 2). Focused questions are questions that ask for specific information; for example, What was the color of the girls hair?. To our knowledge, no empirical work has investigated the accuracy of forensically central and forensically peripheral information an eyewitness can provide in his or her early free recall reports. This issue is of forensic relevance because eyewitnesses first reports are likely to have an important influence on the subsequent course of the forensic investigation.

There is a lack of consensus among eyewitness researchers about how *central* and *peripheral* information should be categorized, and this is likely to affect the ability to draw stable conclusions in this research area. Below we first review the research literature regarding how researchers have classified eyewitness statements into different types of information. In this context, we discuss the extent to which different classification systems of eyewitness reports are likely to aid researchers in understanding the eyewitness ability to answer the two types of questions police ask at the beginning of investigation. These questions are henceforth called *the police questions*. Next, we briefly review the relevant research literature on how communicating an experienced event to others influences memory and meta-memory realism (i.e., how well participants confidence in their recall match the actual accuracy in their recall) of forensically central and forensically peripheral information. We then discuss how multiple retellings of an event influence the realism in ones confidence judgments for central and peripheral information. Finally, hypotheses for the research are presented.

Classification of Eyewitness Statements into Central and Peripheral Categories

There is a general consensus in the eyewitness research literature that eyewitnesses remember the central information about a witnessed event better than 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 & Var-

toukian, 1996; Saywitz, Goodman, Nicholas, & Moan, 1991). However, the definitions of central and peripheral information vary between researchers. For example, the color of the suspects or victims hair or shirt is considered central information by Brown (2003), Christianson (1992), Christianson and Loftus (1987, 1991), Memon and Vartoukian (1996), Parker and Carranza (1989) and by Wessel and Merckelbach (1997). The same information is considered peripheral by Candel et al. (2004), Hershkowitz and Terner (2007), Heuer and Reisberg (1990), Orbach et al. (2000), Roebers, Moga, and Schneider (2001) and by Roebers and Schneider (2000). It is somewhat surprising that both set of authors agree that central information is remembered best while categorizing some of the same information into different categories. The present research does not aim to solve this conflict. Our aim is to merely identify a way to divide eyewitness statements into categories that will help us understand eyewitnesses capacity to answer the two types of questions police have at the beginning of the crime investigation.

The classification models used by different researchers can be grouped into the *Visual attention model*, the *Plot relevancy model*, the *Mixed models*, and the *Empirically based models*. Below we briefly describe these models and discuss their potential use for our purpose.

The Visual attention model

The idea behind the Visual attention model is Easterbrooks (1959) hypothesis claiming that arousal results in the narrowing of attention. As a result of arousal, people increase the processing of central information, but this is at the cost of reduced processing of peripheral information. According to this model, information that is at the focus of attention or is the source of arousal is considered central, such as the offenders act of shooting the victim and the description of the offender (e.g., the color of his shirt or the height of the suspect). In contrast, information that is not at the focus of attention or is irrelevant to the source of arousal (e.g., a car parked on 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).

The Plot relevancy model

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

Mixed models

Burke, Heuer, and Reisberg (1992) and Ibabe and Sporer (2004) proposed more comprehensive models to divide the information into different categories. They used two or more criteria to sort out information into different categories. Burke et al. (1992)

first divided the test questions into central and peripheral questions according to the Plot relevancy model (Heuer & Reisberg, 1990). In the second stage, central questions were divided into gist and basic level questions about visual information, whereas peripheral questions were further divided into questions regarding central details and background details by following the Visual attention model (Christianson & Loftus, 1991). Ibabe and Sporer (2004) first divided the information 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.

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 at large, or specific categories of people, consider as central (important) and peripheral (unimportant) information. Heath and Erickson (1998) asked participants to rate the importance of 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 then four participants were considered peripheral. Roberts and Higham (2002) used four police officers and one crown counsel to classify information on videotape used as a stimulus into correct relevant, correct peripheral, errors, and confabulations. Finally, 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. Unfortunately, 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.

The Visual attention model and the Empirically based models cannot help researchers understand witness ability to answer the two types of police questions because no distinction is made between action details and descriptive details. That is, the same detail can, in different studies, be classified as central or peripheral depending on what the witness is attending to. Since these models allow both action and descriptive details to be present in both the central and the peripheral category, it is hard to know if action details are remembered better than descriptive details. In contrast, the Plot relevancy model is useful for understanding witnesses ability to answer the two types of police questions because this model basically divides the information into action details and descriptive details. The reason is that only a change of action details causes the alteration in the story while a change in descriptive information does not cause alteration in the story.

The classification model used in this research

For the reasons stated above, the plot relevancy model by Heurer and Reisberg (1990) was used in this research but with a modification as described in the Method section below.

Effects of Retelling on Memory Accuracy

Our understanding of witness memory performance in court is complicated by the fact that eyewitnesses tend to share the witnessed event with other people before they testify, often multiple times (Paterson & Kemp, 2006). How an event is shared affects the memory of the event (Hyman, 1994; Marsh, 2007), and discussing the event may affect memory and meta-memory performance differently for different types of forensic information. This is investigated in the first experiment in this study.

Eyewitnesses often incorporate into their memories information supplied by their discussion partners. This may decrease eyewitnesses memory performance (Nourkova, Bernstein, & Loftus, 2004; Wright, Self, & Justice, 2000). Consequently, how a witnessed event is retold and discussed can affect the memory of the event in following recalls (Hyman, 1994; Marsh, 2007; Skowronski & Walker, 2004).

One may think that discussing the details of a witnessed event is like rehearsing them because in discussions the information is also repeated, but in fact simple retelling of an event is different from discussing it. Merely retelling an event can be considered the same as *actively repeating* the learned materials without access to the original material, which is similar to being tested. Active repetition tends to improve accuracy in later recall (Roediger III & Karpicke, 2006a). Roediger and Karpicke called this effect *the testing effect* (Cull, 2000; Roediger III & Karpicke, 2006a, 2006b). Research shows that repetitions (i.e., tests taken) without access to the original material results in superior performance on the final test as compared to the same amount of repetitions (rehearsals) based on reading the original material (Karpicke & Roediger III, 2007, 2008). Moreover, testing can also improve memory of material that was not tested but is related to the tested material (Chan, McDermott, & Roediger III, 2006).

In some cases, multiple retellings can also hamper the retrieval of information because of *retrieval induced forgetting* (Coman, Manier, & Hirst, 2009; MacLeod, 2002). The retrieval induced forgetting effect is most likely to hold when information in the learned material is not well integrated. Thus, when the encoded content is well integrated there may be less or no retrieval induced forgetting (Anderson, 2003; Anderson & McCulloch, 1999).

Given this property of retrieval induced forgetting, the testing effect and the retrieval induced forgetting effect are likely to depend on the type of information being repeated. In the case of Forensically central information, the testing effect seems more relevant. The reason is that actions in an event are well integrated such that retrieval of the first action acts as a cue to the associated action and so on. Forensically peripheral information (i.e., descriptive details of individuals and things) is likely to be characterized by looser integration or association between the different pieces of information. For example, the colour of hair may have only a very weak association with the colour of shirt or with the height of person. Consequently, retrieval induced forgetting may affect the memory of Forensically peripheral information (Coman et al., 2009).

Realism in Confidence Judgments: Central vs. Peripheral Information

Surveys of police, prosecuting and defence attorneys, and jury-eligible samples have singled out eyewitness confidence as the most influential factor when evaluating the correctness of eyewitness statements (Brewer, Potter, Fisher, Bond, & Luszcz, 1999). The confidence expressed by eyewitnesses (henceforth called *confidence judgments*) is often used as a cue to validity by the courts when evaluating the correctness of eyewitness statements (Brewer & Burke, 2002; Wells & Bradfield, 1999; Wells, Lindsay, & Ferguson, 1979).

Ibabe and Sporer (2004), and Migueles and Garcia-Bajos (1999) used focused questions and compared mean confidence scores for central and peripheral information using the plot relevancy model. Their results show that confidence was higher for central information than for peripheral information. However, the authors did not report the relation between level of confidence and level of accuracy of the memory reports. As noted, these studies only used focused questions and not open free recall. However, open free recall and focused questions differ in cognitive demand because focused questions test recognition memory and free recall test participants actual recall. Consequently, participants meta-memory skills may well differ for forensically central and peripheral information.

Roberts and Higham (2002) investigated childrens accuracy and confidence for the different types of information in the context of the cognitive interview. Participants were interviewed in four phases. In first three phases, after context reinstatement, participants reported the event details in natural order, reverse order, and with changed perspective. In the fourth phase they used focused questions. They used an empirical approach to distinguish between central and peripheral information. The results were that confidence in central action details was higher than confidence in peripheral action details, and confidence in central descriptive details was higher than in peripheral descriptive details. Moreover, central information accuracy was associated with higher confidence, and peripheral information accuracy decreased with the increase in confidence level. Roberts and Higham (2002) explored the confidence accuracy relationship by collapsing the data for the four phases. This makes it difficult to understand which types of recall cause the effects because different types of recall use different cognitive resources and consequently different meta-memory processes.

An effect of repeated post-event questioning may be increased confidence without a corresponding change in the accuracy of memories recalled. One reason for this could be that the ease of information retrieval influences confidence judgments; that is, the feeling of retrieval *fluency*. Fluency has been reported to increase with repeated retrievals of the same information (Shaw, 1996; Shaw & McClure, 1996; Shaw, McClure, & Dykstra, 2007). This increase in retrieval fluency may be expected to affect Forensically peripheral information more than Forensically central information because peripheral information is likely to be less well integrated then Forensically central information. When less well-integrated information becomes more associated as an effect of repeated recalls, fluency is likely to increase.

In this study, three measures of meta-memory realism will be used as indicators of participants meta-memory: calibration, over-/underconfidence and slope. These measures are further described in the Method section below. The reason for us-

ing calibration methodology is that these measures give a more differentiated and informative understanding of realism in confidence judgments compared to simply computing the confidence-accuracy correlation, which is the traditional measure in lineup research (Juslin, Olsson, & Winman, 1996; Weingardt, Leonesio, & Loftus, 1994; Wells, Olson, & Charman, 2002).

Overview of the Present Study

This study consists of two experiments. Experiment 1 (the main experiment) is a part of a larger project. The overall findings of participants performance have been reported elsewhere (Sarwar, Allwood, & Innes-Ker, 2010). In the present context, we are focusing on participants memory and meta-memory realism for Forensically central, Forensically peripheral, and Non-forensic information using free recall and focused questions. According to the authors knowledge, participants meta-memory realism for Forensically central, Forensically peripheral, and Non-forensic information has not been investigated in the context of open free recall. These issues have also not been investigated for open free recall after retellings and discussions. Experiment 1 investigated the memory and meta-memory realism for different types of information, and how discussing an experienced event with non-witnesses may influence both memory and meta-memory realism for different types of information in the following free recall and focused questions. Participants first watched a short film about a kidnapping then participated in four conditions (described below). After three weeks, their memory for events in the film was tested, and 3 to 4 days later participants gave confidence judgments of their recall.

The four conditions were: 1) The *Lab-discussion condition*, where participants retold (5 times) the witnessed event to an unfamiliar person (each time new). The person then posed questions about the event which were answered by the witness. 2) The *Family discussion condition*, participants retold (5 times) the witnessed event to their family and friends (each time new). Family and friends then posed questions about the event which were answered by the witness. 3) The *Retell condition*, participants retold (5 times) the witnessed event to a person but no questions were asked. 4) The *Control condition*, where neither retelling nor discussion took place.

The Lab-discussion condition can be seen as the critical condition. However, due to the very controlled nature of this condition, the Family discussion condition was created to increase the ecological validity and reduce the impact of strict lab conditions. To tap the effects of discussion, the Retell condition was introduced to control for the effects of merely retelling the event.

Hypotheses

Hypothesis 1

The participants accuracy (i.e., proportion correct recalled information of all recalled information) and meta-memory realism were predicted to be better for Forensically central information than for Forensically peripheral information. The reasoning is that people in general are better at describing actions than giving descriptions (Migueles & Garcia-Bajos, 1999) and, consequently, the memory of the central information will be better.

Hypothesis 2

For Forensically central information, participants in the Lab- and Family discussion conditions were expected to have lower accuracy and poorer calibration than participants in the Retell condition. The reason is that the exchange of information in the two discussion conditions can be expected to introduce information that might decrease the accuracy of these participants (Nourkova et al., 2004; Wright et al., 2000). In contrast, participants in the retell condition were simply expected to benefit from the retellings because of the testing effect (Roediger III & Karpicke, 2006a). The confidence of the participants in the two discussion conditions would be inflated due to increase in retrieval fluency (Shaw, 1996; Shaw & McClure, 1996) caused by the post-event discussions in addition to the increase in fluency that may have taken place as an effect of the mere retelling in all three experimental conditions. Consequently, participants in the two discussion conditions were expected to show poorer meta-memory than those in the retell condition. All together this would result in making the participants in the Retell condition more accurate and better calibrated for Forensically central information than the participants in the discussion conditions.

Hypothesis 3

For Forensically peripheral information, participants in the three experimental conditions were expected to show poorer calibration than the Control condition. The reason is that the three conditions would not differ in accuracy because of a floor effect, 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). Moreover, confidence would be inflated due to an increase in retrieval fluency (Shaw, 1996; Shaw & McClure, 1996) caused by the post event retellings. Consequently, with poor accuracy and inflated confidence, the participants in three experimental conditions would show poor calibration. In contrast, though the participants in the Control condition would also have poor memory for forensically peripheral information, their confidence judgments were expected to be more realistic since they neither repeated nor discussed the forensically peripheral information. Consequently the participants in the control condition would show better calibration than the participants in the three experimental conditions.

Experiment 1

Method

Participants. The participants were 89 students (62 women) from Lund University with the mean age of 25 (18-47 years). At first, 92 participants were recruited with the aim to allocate 23 participants to each condition. The dropouts from the Lab-discussion condition, Family discussion condition, the Retell and the Control condition were 7, 6, 4, and 1 respectively. Dropouts during the study were all replaced except for the three who dropped out at the end. Participants who successfully completed the experiment received a movie ticket worth 90 SEK (approximately US$ 12).

Design

A between-subjects design with four conditions was used. The four conditions were: Lab-discussion (n = 21; one participant only attended 4 out of 5 sessions), Family discussion (n = 23), Retell (n= 23) and Control condition (n= 22).

Materials

Videotape

A film about two men kidnapping a woman from a bus stop was used. 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 (e.g., Allwood, Ask, & Granhag, 2005; Granhag, 1997).

Focused questions about the film

The questionnaire consisted of 44 questions about the film. Each question had one correct and one incorrect answer alternative. The questions were about different details of the peoples appearance, clothes, ages, as well as the surrounding environment such as letter boxes, cars, busses and, specifically, the offenders car. These questions have been used in other work as well (Allwood, Innes-Ker, Holmgren, & Fredin, 2008 ; Allwood, Jonsson, & Granhag, 2005).

Confidence judgment scales

An 11-point scale was used for the participants judgments of the detailed parts of the free recalls. This scale started at 0% (Completely sure that I remember wrong) and increased in steps by 10 (10, 20, 30) to 100% (Completely sure that I remember correct). This scale was inserted below each statement to be confidence judged. A 6-point scale was also used for participants judgments of their answers to the 44 focused questions on the film. Here the probability to randomly choose the right answer was 50%, and this confidence scale went from 50% (Guessing), in steps by ten to 100% (Completely sure).

Procedure

The participants were received in the lab in small groups between 4 to 8 participants. Participants were told: We are investigating human perception in different forensic situations. First, participants watched the video film, then they were randomized to one of the four conditions. Participants in the three experimental conditions received a time schedule for the five occasions over a 20-day period when they would retell the film. In each of the five sessions, the participants in the Lab-discussion and Retell conditions visited the laboratory. Participants in the Lab-discussion condition discussed the events in the film with a confederate. The confederate asked spontaneous questions about the film after listening to the story of the film. 105 confederates were recruited solely for the discussion purpose, and each confederate participated only in one discussion. The participants in the Retell condition simply told the story about the film to the experimenter. Participants were instructed to tell whatever they remembered about the film. The experimenter posed no question to the participants.

All discussions and retellings were recorded on a MP3 recorder. Participants in the Family discussion condition did five discussions at home with their family or friends, each time with a new individual. The participants carried out their discussions on the scheduled dates and times. Further, participants confirmed their discussions to the experimenter by 7:00 pm by sending an SMS (Short Message Service). If any participant failed to send the SMS by 7:00 pm on the discussion day, the experimenter called the participant and reminded him/her of the task. In the Control condition, participants were instructed to not tell the contents of the film to anybody.

All participants returned to the lab on the 21st day for a memory test. The participants first typed whatever they remembered about the events they witnessed in the film (free recall test) and then answered 44 focused questions. Finally, all participants returned to the laboratory on the 24th or 25th day to give their confidence judgments. First, participants gave their confidence judgments about their free recall statements (which had been prepared as described below). Finally, the participants gave their confidence judgments for the 44 forced-choice questions.

Preparation of material for participants confidence judgments

In order to record participants confidence judgments for each piece of information, each participants free recall was broken down into single pieces of information using the method developed by Allwood, Ask, and Granhag (2005). The principles used were: (1) statements about actors and actions carried out were considered a single piece of information. For example the woman looked at the timetable was considered as a single piece. (2) An object with one associated characteristic was considered as one piece. For example the yellow letterbox was considered as a single piece. (3) If an object was associated with more than one characteristic, the additional characteristics were regarded as separate pieces of information. For example the bald old guy was considered as two pieces. To help participants recall the context in which they mentioned each piece of information, one or two sentences adjacent to that piece were attached. These reference items were enclosed in brackets while the items to be confidence judged were underlined. Finally, an 11-point confidence scale was placed under each piece of information.

Division of memory reports into forensic and non-forensic sub-categories

In preparation for the ensuing analyses, participants memory reports were classified into different information categories in two steps. In the first step, to separate out the irrelevant information, participants statements were divided into Forensic and Non-forensic statements. Information such as houses, roads, surroundings, etc., was seen as Non-forensic because it is either not needed to solve the crime or can be collected from the crime scene without the help of an eyewitness. In the second step, the forensic statements were divided into Forensically central and Forensically peripheral categories by following the plot relevancy model by Heurer and Reisberg (1990). This model is useful since it divides the information into action details and descriptive details.

Measurements

We used calibration, over-/underconfidence and slope to measure the realism in participants confidence judgments. *Calibration* and *over-/underconfidence* measure the relationship between a persons level of confidence and accuracy. Calibration compares a persons accuracy and the corresponding confidence for each confidence level and represents the persons squared deviance from perfect calibration at all confidence levels (depending on the type of confidence scale used, e.g., 50, 60, 70,). Over-/underconfidence is the difference between the persons average confidence and his/her percentage of correct items. The value zero represents perfect realism for both calibration and over-/underconfidence. For the over-/underconfidence measure, positive and negative values show overconfidence and underconfidence respectively. Slope is a measure of the persons ability to separate his/her correct and incorrect answers by means of the level of his/her confidence judgments. It is computed by subtracting the persons mean confidence level for incorrect items from the mean confidence level of the correct items. Yates (1994) provides the specific formula for calibration and a more thorough discussion of these measures.

Results

Free Recall

We first present the analyses for the number and proportion of recalled items of the different types (i.e., Forensically central, peripheral, and Non-forensic items), and the total number of correct recalled items for each information type and for each condition. Next we present results regarding participants meta-memory realism for the Forensically central, Forensically peripheral and Non-forensic information (Hypothesis 1) and the effect of discussions on memory and meta-memory realism for each condition for the Forensically central information (Hypothesis 2), Forensically peripheral information (Hypothesis 3) and Non-forensic information. Last, the results for the focused questions are presented.

Number of recalled items by forensic category and condition

The total amount of items reported (both correct and incorrect) indicates whether the different conditions affected the *quantity* of information volunteered. In order to investigate the number of items recalled in the different items categories and the effect of conditions on the number of items recalled in each category, the data were subjected to 4x3 mixed ANOVA. The between-subjects factor was condition (Lab-discussion, Family discussion, Retell and Control condition) and the within-subject factor was information type (Forensically central, Forensically peripheral, and Non-forensic). The means and standard deviations are shown in Table 1.

The results showed that there was a significant main effect of information type, Wilks' Lambda= .19, $F(2, 84) = 178.3$, $p < .001$; $\eta^2 = .62$. Contrasts with Bonferroni corrected significance levels revealed that Forensically central items were reported significantly more often than Forensically peripheral items $F(1, 85) = 185.3$, $p < .001$;$\eta^2 = .65$, and Non-forensic items, $F(1, 85) = 328.2$, $p < .001$; $\eta^2 = .77$.

There was also a significant main effect of the four conditions, $F(3, 85) = 6.34$, $p < .001$; $\eta^2 = .18$. Contrasts for the four conditions and the type of information with Bonferroni corrected significance levels revealed for Forensically central information that the Lab-discussion condition, $t(41) = 1.90$, $p < .001$; Cohen's $d = 1.5$, the Retell condition, $t(43) = 3.84$, $p < .001$; Cohen's $d = 1.2$, and the Family discussion condition, $t(43) = 2.75$, $p < .01$; Cohen's $d = .83$, reported significantly more items as compared to the Control condition. No differences between the conditions were found for the Forensically peripheral and Non-forensic information.

The results also showed that there was a significant interaction effect for condition and information type, Wilks' Lambda= .83, $F(2, 86) = 2.74$, $p < .01$; $\eta^2 = .04$. Consulting Table 1, it appears the interaction is driven by the greater difference in frequency between Forensically central and peripheral information in the Retell condition, compared to the other conditions.

Proportions of reported Forensically central, Forensically peripheral and Non-forensic information

In order to get an impression of the importance of the Forensically central, Forensically peripheral, and Non-forensic information recalled in the different conditions, we first look at the *proportion* of the recalled items that can be classified into the various categories of forensic relevance. Proportions were calculated by dividing the number of each type of information with the total number of items reported and then multiplying by one-hundred to derive percentages. A 4x3 mixed ANOVA with a similar construction as above was conducted. The proportions are shown in Table 1. The results showed that there was a significant main effect for information type, Wilks' Lambda= .23, $F(2, 84) = 138.8$, $p < .001$; $\eta^2 = .61$. Contrasts with Bonferroni corrected significance levels revealed that the proportion of Forensically central items was significantly higher than the proportion of Forensically peripheral items, $F(1, 85) = 170.04$, $p < .001$; $\eta^2 = .52$, and the proportion of Non-forensic items, $F(1, 85) = 257.0, 1$, $p < .001$, $\eta^2 = .57$.

Number of correct items for different kinds of information

The number of correct items (here used as a measure of completeness) is an indicator of the quality of the recalled information. A 4x3 mixed ANOVA with a similar construction as above was conducted. The number of correct items is shown in Table 1. The results showed that there was a significant main effect of information type, Wilks' Lambda= .19, $F(2, 80) = 170.9$, $p < .001$; $\eta^2 = .69$. Contrasts with Bonferroni corrected significance levels revealed that significantly more correct Forensically central items were recalled as compared to the Forensically peripheral items, $F(1, 81) = 293.04$, $p < .001$; $\eta^2 = .75$, and the Non-forensic items, $F(1, 81) = 260.5$, $p < .001$; $\eta^2 = .74$. The contrasts also revealed that the number of correct Non-forensic items was also significantly higher than the number of correct Forensically peripheral items, $F(1, 81) = 11.91$, $p < .001$; $\eta^2 = .12$.

There was also a significant main effect of condition, $F(3, 81) = 6.45$, $p < .001$; $\eta^2 = .19$. Contrasts with Bonferroni corrected significance levels revealed that the participants in the Lab-discussion condition, $t(41) = 4.52$, $p < .001$, Cohen's $d = 1.4$, and Retell condition, $t(43) = 3.62$, $p < .001$, Cohen's $d = 1.1$, recalled a significantly

higher number of correct items then the participants in the Control condition. Further contrasts for each type of information revealed that the participants in the Lab-discussion recalled significantly more correct items than the Control condition for Forensically central information, $t(41) = 5.16$, $p < .001$, Cohen's $d = 1.6$, and Non-forensic information, $t(39) = 2.33$, $p < .03$, Cohen's $d = .75$. The participants in the Retell condition also recalled significantly more correct items then the participants in the control condition on the Forensically central information, $t(43) = 4.63$, $p < .001$, Cohen's $d = 1.4$, and the Non-forensic information, $t(39) = 2.10$, $p < .05$, Cohen's $d = .67$.

The results showed that there was a significant interaction effect for condition and information type, Wilks' Lambda= .82, $F(2, 80) = 2.88$, $p < .01$; $\eta^2 = .04$. By inspection of means, it appears that the interaction is driven by the greater difference between Forensically central and Forensically peripheral information in the number of correct items recalled in the Retell condition, compared to the other conditions.

Meta-memory realism: Calibration curves

Figure 1 shows the calibration curves for the free recall for Forensically central, Forensically peripheral, and Non-forensic information for all participants. The x-axis shows the confidence levels (from 0 to 100%) and the y-axis shows the percent of correct answers. The numbers inside the graph give the percent of answers for each confidence level in each condition. The diagonal shows perfect calibration. As can be seen in the Figure, the calibration curves for the Forensically central, Forensically peripheral, and Non-forensic information show very low percentages for the confidence scores between 0 and 50. Consequently, these levels will not be further discussed. For confidence scores 50 and above, the scores for Forensically central and Non-forensic information are clustered close to the diagonal, indicating that confidence is fairly realistic for these two types of information. For the Forensically peripheral information, the calibration curve shows overconfidence from 60% to 100% confidence. In addition, for both the Forensically central and Non-forensic information, about half of the confidence judgments are at the 100% level. For Forensically peripheral information, the confidence judgments are fairly evenly distributed between 50% and 100%. For all information types the participants show some overconfidence at the 100% confidence level.

Table 1.
Experiment 1: Results for the Open free recall. Mean Number, Proportions, and Correct items of Recalled Items for Different Types of Information for the Four Conditions

	Lab-discussion	Family discussion	Retell	Control
Mean number of items				
Forensic central	31.24(8.3)	25.83(9.0)	28.48(8.9)	18.86(7.9)
Forensic peripheral	13.81(8.5)	10.56(9.6)	7.13(6.7)	9.23(7.3)
Non-forensic	12.71(7.5)	10.69(7.6)	10.34(8.2)	7.40(5.2)
Mean proportions				
Forensic central	56.05(11.4)	56.85(14.9)	64.15(15.1)	55.26(15.5)
Forensic peripheral	22.69(11.7)	20.85(17.6)	14.74(13.4)	23.58(17.5)
Non-forensic	21.26(10.5)	22.29(12.6)	21.11(12.1)	21.16(12.3)
Correct items				
Forensic central	27.43(7.9)	22.00(8.2)	26.95(8.8)	16.75(6.3)
Forensic peripheral	8.52(5.5)	6.52(6.4)	4.71(4.9)	5.50(5.2)
Non-forensic	10.81(6.9)	8.91(6.1)	10.3(6.7)	6.75(3.6)

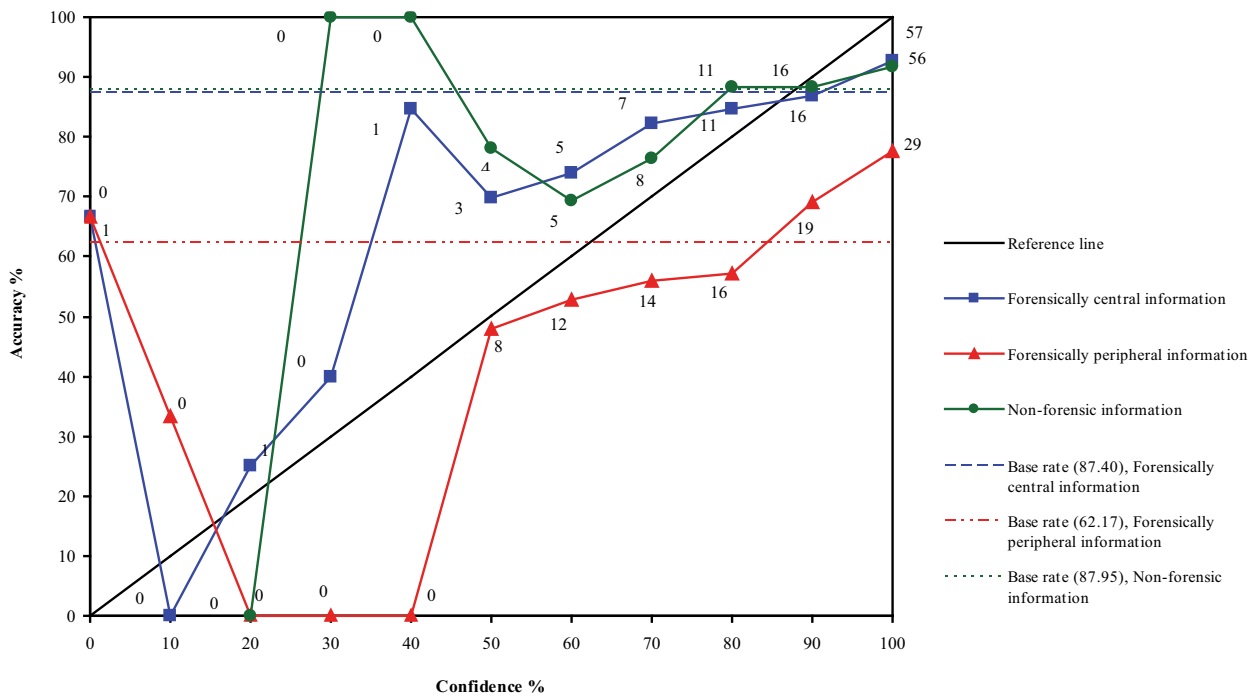

Figure 1: *Experiment 1: Calibration curves for the free recall for the different information types (Forensically central, Forensically peripheral, and Non-forensic) collapsed over the four conditions. Digits at each point show the % of confidence judgments that used this confidence level at that particular accuracy level.*

Accuracy, confidence and realism in confidence for Forensically central, peripheral and non-forensic information and for the four conditions

The next analysis addressed the first hypothesis, which stated that the participants would have a higher accuracy (proportion correct) and better meta-memory realism for Forensically central information as compared to Forensically peripheral information. To compare the participants performance in the four conditions for the three information types, the data were subjected to 4x3 mixed ANOVAs with the same construction as above, separately for accuracy, confidence, calibration, over-/underconfidence, and slope. The means and standard deviations are shown in Table 2.

Accuracy. The results showed a significant main effect of information type, Wilks' Lambda= .43, $F(2, 66) = 43.56$, $p < .001$; $\eta^2 = .50$. Contrasts with Bonferroni corrected significance levels revealed that the participants were significantly more accurate for Forensically central information then the Forensically peripheral information, $F(1, 67) = 73.52$, $p < .001$; $\eta^2 = .25$. The participants were also more accurate for the Non-forensic information than the Forensically peripheral information, $F(1, 67) = 85.68$, $p < .001$; $\eta^2 = .30$.

Confidence. The results showed a significant main effect for confidence for information type, Wilks' Lambda= .43, $F(2, 66) = 44.64$, $p < .001$; $\eta^2 = .46$. Contrasts with Bonferroni corrected significance levels revealed that the participants were significantly more confident for the Forensically central information than the Forensically peripheral information, $F(1, 67) = 79.60$, $p < .001$; $\eta^2 = .25$. The participants were also more confident for the Non-forensic information as compared to the Forensically peripheral information, $F(1, 67) = 75.64$, $p < .001$; $\eta^2 = .28$.

Calibration. The results showed a significant main effect for calibration for information type, Wilks' Lambda= .65, $F(2, 83) = 22.02$, $p < .001$; $\eta^2 = .27$. Contrasts with Bonferroni corrected significance levels revealed that the participants were significantly better calibrated for Forensically central information then the Forensically peripheral information, $F(1, 84) = 44.15$, $p < .001$; $\eta^2 = .16$. The participants were also better calibrated for Non-forensic information as compared to the Forensically peripheral information, $F(1, 84) = 29.04$, $p < .001$; $\eta^2 = .14$.

Over-/underconfidence. The results showed a significant main effect for over-/underconfidence for information type, Wilks' Lambda= .75, $F(2, 83) = 14$, $p < .001$; $\eta^2 = .19$. Contrasts with Bonferroni corrected significance levels revealed that the participants showed significantly less over/underconfidence for Forensically central information than for Forensically peripheral information, $F(1, 84) = 26.22$, $p < .001$; $\eta^2 = .12$. The participants also showed less over/underconfidence for Non-forensic information as compared to the Forensically peripheral information, $F(1, 84) = 24.18$, $p < .001$; $\eta^2 = .11$.

Slope. There was no significant difference for slope between the three types of information.

The focused questions

Separate between-subjects one-way ANOVAS were conducted for each of the dependent measures for the focused questions (see Table 3). There were no significant effects for any of the measures.

Table 2.
Experiment 1: Results for the Open free Recall. Forensically Central, Forensically Peripheral, and Non-forensic Information: Means (and SDs) for Completeness, Accuracy, Confidence, Calibration, Over/underconfidence and Slope

	Lab-discussion	Family discussion	Retell	Control
Accuracy				
Forensic central	87.60(10.1)	84.20(10.0)	92.56(5.2)	85.10(12.9)
Forensic peripheral	61.10(21.2)	63.91(22.0)	68.15(25.9)	53.87(27.9)
Non-forensic	86.15(13.8)	91.32(11.3)	93.84(6.8)	86.18(13.1)
Confidence				
Forensic central	87.96(6.7)	90.45(5.9)	90.17(4.9)	86.86(7.7)
Forensic peripheral	75.71(10.4)	80.60(12.1)	80.44(7.2)	74.92(12.8)
Non-forensic	89.76(6.8)	91.43(8.3)	90.53(7.1)	86.57(11.9)
Calibration				
Forensic central	.04(.03)	.05(.03)	.03(.02)	.05(.03)
Forensic peripheral	.11(.09)	.12(.11)	.11(.12)	.12(.10)
Non-forensic	.05(.05)	.05(.07)	.02(.02)	.06(.06)
O/U confidence				
Forensic central	.00(.10)	.05(.11)	-.02(.07)	.01(.13)
Forensic peripheral	.14(.19)	.15(.24)	.12(.26)	.14(.22)
Non-forensic	.03(.14)	.01(.13)	-.03(.08)	.02(.16)
Slope				
Forensic central	8.11(9.9)	16.63(14.9)	.04(8.2)	7.10(7.8)
Forensic peripheral	7.15(9.2)	7.27(13.1)	11.25(6.1)	13.88(14.7)
Non-forensic	10.79(16.0)	10.63(18.3)	21.98(11.9)	11.64(7.9)

Note. O/U confidence = Over-/underconfidence

Table 3.
Experiment 1: Results for Focused Questions. Means (and SDs) for Accuracy, Confidence, Calibration, Over-/underconfidence and Slope

	Lab-discussion	{Family discussion	Rehearsal	Control	F	η^2
Accuracy	53.68(6.68)	55.34(7.14)	56.13(5.83)	54.03(6.54)	0.67	0.02
Confidence	62.97(4.45)	65.55(6.76)	64.33(6.97)	62.28(5.57)	1.29	0.04
Calibration	.05(.02)	.05(.03)	.05(.02)	.05(.03)	0.32	0.01
O/U confidence	.09(.08)	.10(.08)	.08(.07)	.08(.09)	0.33	0.01
Slope	4.02(4.37)	3.50(4.94)	4.22(5.73)	3.64(5.46)	0.09	0

Note. O/U confidence = Over-/underconfidence

Discussion

The results for the open free recall showed that participants performed significantly better for the Forensically central information as compared to the Forensically peripheral and Non-forensic information on mean number of items, proportion, and correct items reported in each information category. The participants better performance for central information is in line with previous research (Roebers et al., 2001). The lower number of peripheral items could, at least partly, be due to repetition. Hershkowitz and Terner (2007) found that in the second recall attempt, participants recalled less peripheral details as compared with the first attempt. It is difficult to say whether the low number of peripheral details recalled was due to the second recall attempt or because of the other factors that might have influenced the memory of peripheral information during the time between first and second recall. A partial explanation for the lower number of Non-forensic information units may be that participants exert control over what to share on the basis of their expectation of what is expected of them (Grice, 1975; Russell & Schober, 1999). Consequently, the participants may only have shared such information that they considered relevant to the forensic event.

There were also differences between the conditions for the different types of information. The results showed that the participants in the Retell and the Lab-discussion conditions benefited from the retellings and discussions respectively, and they reported significantly more Forensically central items as compared to the Control condition. This benefit of retellings and discussions did not translate into a benefit for the two conditions as compared to the Control condition for the number of peripheral items and correct peripheral items recalled.

In the first hypothesis we expected that the participants accuracy and meta-memory realism would be better for the Forensically central information as compared to the Forensically peripheral information. Our results indeed showed that participants had better memory, higher confidence, better calibration, and over-/underconfidence for the Forensically central information as compared to the Forensically peripheral information. However, the slope measure did not differ between the information types.

Participants superior performance for the Forensically central information as compared to the Forensically peripheral information is in line with the previous findings where focused questions were used (Ibabe & Sporer, 2004; Migueles & Garcia-Bajos, 1999). Large effect sizes showed that there was a strong effect of information type. The present findings extend the previous findings to eyewitness open free recall. Moreover, the results of the meta-memory measures further add to our knowledge in that the participants were realistic in their confidence for Forensically central information. Interestingly, the slope results showed that participants ability to discriminate between correct and incorrect responses did not differ for the Forensically central and peripheral information types.

According to the second hypothesis, Forensically central information in the two discussion conditions was expected to show lower accuracy and meta-memory realism than in the Retell condition. The results did not support this hypothesis and showed no difference between the conditions for memory, confidence, calibration, over/underconfidence, and slope. This result should be taken with caution because the mixed ANOVA is a bit stringent, and 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. This issue should be explored further in future research with a different design and with more participants.

The results did not support the third hypothesis that for Forensically peripheral information, participants in the three experimental conditions would show poorer calibration than the Control condition. The reason is that Forensically peripheral information is difficult to remember and any possible effect might have been buried in a floor effect.

The four conditions performed poorly on 44 focused questions and did not differ between themselves. One possible reason is that the focused questions were only about Forensically peripheral information. Since there were no differences between the conditions for the Forensically peripheral information in the open free recall, findings for focused questions further strengthen the open free recall results that Forensically peripheral information is difficult to remember accurately. Since the 44 focused questions in this Experiment were only about Forensically peripheral information, comparison of the participants performance for central and peripheral information was not possible for the focused questions. To further explore this issue we conducted Experiment 2. Given that the results for the focused questions did not differ between the four conditions, we choose not to include the different conditions of Experiment 1 in Experiment 2.

Experiment 2

Method

Participants

Seventy-seven students from Lund University participated. The mean age of the participants was 26.5 years, ranging from 20 to 65 years. Each participant was given a lottery ticket worth 25 SEK (approximately US$ 4).

Design

A within-subject design was used. The within-subject factor was the two types of focused questions, Forensically central and Forensically peripheral.

Materials

Videotape

The same video was used as in Experiment 1.

Focused questions about the film

The questionnaire consisted of 63 questions about the film. Eighteen questions were about forensically central details of the events shown in the film. Forty-five questions were about forensically peripheral details of the events. The different number of questions in the two categories was because there were different numbers of these

details in the film. Questions were arranged in the order events happened in the film. Each question had three alternative answers where one was always correct. The greater number of options was chosen to lower the guessing level as compared to the two options used in the 44 focused questions in Experiment 1. The position of the correct answer was randomized for the questions.

Confidence judgment scale

An 11-point confidence scale from 0% (Completely sure that I remember wrong) to 100% (Completely sure that I remember correct) confidence was used. This scale was placed under each question.

Procedure

The participants first watched the same film as in Experiment 1. In order to avoid covert rehearsal, participants then read a one and half page long article in English as a filler task for 10 minutes. Next, participants answered the questionnaire and made a confidence judgment immediately after answering each question.

Results

Calibration curves

Figure 2 shows the calibration curves for the Forensically central and peripheral information. The Forensically central information shows very low percentages for the confidence levels between 0 and 60. The calibration curve for the Forensically peripheral information shows that confidence scores are almost equally spread from 0% to 90%. The graph suggests that participants were relatively well calibrated for the Forensically central information as compared to the Forensically peripheral information. The calibration curve for Forensically peripheral information is fairly flat. This shows that participants had difficulty in separating correct from incorrect answers by means of their confidence judgments.

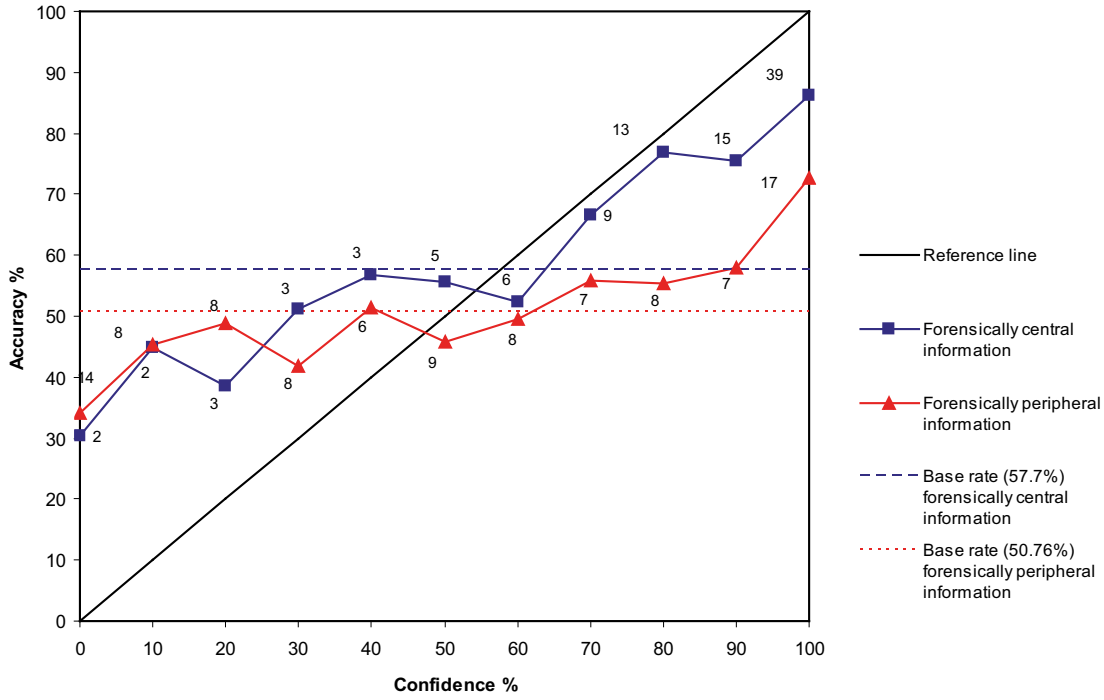

Figure 2: *Experiment 2: Calibration curves for the focused questions for the Forensically central and Forensically peripheral information. Digits at each point show the % of confidence judgments that used this confidence level at that particular accuracy level.*

Accuracy, confidence, over-/underconfidence, and slope

Paired sample t-tests were conducted to compare the participants scores for accuracy, confidence, calibration, over-/underconfidence, and slope for Forensically central information and Forensically peripheral information (Table 4). The outcome was that participants showed significantly higher accuracy, $t(76) = 16.46$, $p < .001$, Cohen's $d = 2.19$, confidence, $t(76) = 25.13$, $p < .001$, Cohen's $d = 2.36$, had better calibration, $t(76) = -3.06$, $p < .001$, Cohen's $d = -0.55$, and steeper positive slope, $t(76) = 2.73$, $p < .008$, Cohen's $d = .39$, for the Forensically central information as compared to the Forensically peripheral information. However, participants showed less overconfidence, $t(76) = 3.96$, $p < .001$, Cohen's $d = .40$, for the Forensically peripheral information as compared to the Forensically central information.

Discussion

Experiment 2 was conducted to investigate if participants would show better memory performance and meta-memory realism for the Forensically central information as compared to the Forensically peripheral information for focused questions. The results showed that the participants had better accuracy, higher confidence, better calibration, and larger slope for Forensically central information. In contrast, participants showed less over-/underconfidence for the Forensically peripheral information as compared with the Forensically central information. The results for accuracy and confidence are in line with previous studies using open-ended questions (Yuille & Cutshall, 1986), focused questions (Parker & Carranza, 1989), and both focused and open-ended questions (Ibabe & Sporer, 2004) that show people are better in remembering Forensically central information as compared to Forensically peripheral information.

As noted, the results showed that the participants had lower overconfidence for Forensically peripheral information as compared to Forensically central information. This result was unexpected because the participants were expected to show less overconfidence for the Forensically central information due to better memory for this type of information. Moreover, the Forensically peripheral information is difficult to remember and is supposed to cause overconfidence as compared to the Forensically central information because of the hard-easy effect (Merkle, 2009). This result is difficult to explain. However, the Calibration curve in Figure 2 shows that the Forensically peripheral items demonstrate underconfidence for all confidence levels under 50% and overconfidence for all confidence levels over 40%. Here it is also noteworthy that the participants used the confidence class 0 for 14% of the Forensically peripheral items whereas their performance for these items was at chance level. In total, in the over-/underconfidence measure, these two tendencies (under- and overconfidence at the different halves of the scale) balance out. This is also illustrated in that the calibration measure showed worse calibration for the Forensically peripheral items as compared with the Forensically central items. Finally, it should be noted that the Forensically peripheral items showed greater overconfidence than the Forensically central items for all confidence levels above 40%. In brief, these observations suggest that the better overconfidence score for the Forensically peripheral items compared to the Forensically Central items should be viewed with caution.

In general, the focused questions, summarizing over the two question types, showed

Table 4.
Experiment 2: Means (and SDs) for Accuracy, Confidence, Calibration, Over-/underconfidence and Slope

	Forensically central information	Forensically peripheral information	t	Cohen's d
Accuracy	72.01(11.05)	51.60(7.25)	16.46*	2.18
Confidence	76.97(10.27)	51.41(11.49)	25.13*	2.34
Calibration	.09(.05)	.12(.06)	-3.06*	-0.54
O/U confidence	.05(.12)	-.00(.13)	3.96*	0.4
Slope	20.79(15.60)	15.50(11.61)	2.73*	0.38

Note. O/U confidence = Over-/underconfidence
*$p < 0.01$

a lower level of overconfidence than has been observed in previous research (e.g Allwood et al., 2008 ; Allwood, Knutsson, & Granhag, 2006). A possible reason might be that the three answer alternatives may have induced more uncertainty (Robinson, Johnson, & Robertson, 2000), especially for the more difficult questions (i.e., the Forensic peripheral).

General Discussion

This research investigated eyewitness capacity to report and give realistic confidence ratings of their recall of Forensically central and peripheral information. The issue was investigated in the context of eyewitness early open free recall reports and their answers to focused questions. For the open free recall, we also looked at Non-forensic information.

The results for the open free recall showed that the number of items reported and the number of correct items reported was higher for the Forensically central information as compared to the Forensically peripheral and Non-forensic information, while participants accuracy both for the Forensically central and the Non-forensic information was higher than the Forensically peripheral information. These results are in line with the previous findings that for Forensically central information participants report more details (Jelicic et al., 2006; Yuille & Cutshall, 1986) and have higher accuracy (Ibabe & Sporer, 2004; Migueles & Garcia-Bajos, 1999) than Forensically peripheral details. This suggests that people can describe the course of an event in a forensic situation quite well as compared with their description of individuals and objects. Hence, the police need to be more careful when taking into account the description of the offender provided by the eyewitnesses compared with the witness description of the information of the event.

The results for the meta-memory measures for the free recall show that confidence for the Forensically central information (in contrast to the Forensically peripheral information) can be fairly realistic. These results indicate that confidence may be used as a reliable indicator of accuracy for Forensically central information.

This research also addressed the impact of retellings, discussions, and the type of memory question on the memory and meta-memory realism of different types of information. The results showed that the participants in the two discussion conditions reported significantly more Forensically central information (and the participants in the Lab-discussion condition more correct Forensically central information items) than the participants in the control condition. However, this superior performance did not translate into better accuracy compared with the control condition. Moreover, the four conditions did not differ in terms of accuracy, confidence, calibration, and over-/underconfidence and slope. These results are relevant for the courts decision making since courts place more trust on eyewitnesses who provide more information than eyewitnesses who provide less information (Bell & Loftus, 1988; Heath, Grannemann, Sawa, & Hodge, 1997). This suggests that courts need to be careful when using the amount of information reported as a criterion for the credibility of an eyewitness. In this context, results of slope measure can be useful. These results showed that the participants in the four conditions (and for both types of forensic information) did not differ in their capacity to discriminate between correct and incorrect items. Comparing the items that were assigned high confidence with the items that were

assigned low confidence may be helpful in separating correct from incorrect information. This comparison should be done separately for each type of information. The reason is, as discussed above, that forensically central items received high confidence judgments as compared to forensically peripheral items. Consequently, comparing the confidence level within the Forensically central and within the Forensically peripheral items might help to separate correct from incorrect information.

The results showed no difference between the participants performance in the four conditions for the Forensically peripheral information. However, a limitation of the results concerning the open free recall in this study was that the information sorted as Forensically peripheral had a quite low frequency and for this reason differential effects may have been buried in a floor effect.

The results for focused questions from the two experiments are in line with other empirical findings using the Plot relevancy model that people, in general, are better at remembering Forensically central information (Hershkowitz & Terner, 2007; Roebers & Schneider, 2000; Yuille & Cutshall, 1986).

The results for focused questions from Experiment 2 show that people also have better meta-memory realism for the Forensically central information than for the Forensically peripheral information. An exception to this was that the participants showed no overconfidence for the Forensically peripheral information and, in this context, better realism for the peripheral questions than for the Forensically central questions. However, as discussed above, the absence of over- or underconfidence for the peripheral items seems to be due to underconfidence at one end of the scale and overconfidence at the other. The specific reason for this pattern is not clear and needs to be further investigated in future research.

In brief, our results indicate that although repeated retellings and discussions risk introducing specific errors into the testimony, the overall effect of these activities seen in this study was positive (compared to the control condition) for memory performance for Forensically central information. In contrast, there was less effect for Forensically peripheral information. When, as often may be the case, focused questions concern Forensically peripheral information our results from Experiment 1 suggest that the effect of repeated retellings and discussion over a short time period of about three weeks will not be very noticeable.

References

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

Allwood, C. M., Innes-Ker, Å., Holmgren, J., & Fredin, G. (2008). Accuracy and realism in confidence in childrens and adults event recall in response to free recall and focused questions. *Psychology, Crime & Law*, 14, 529-547.

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.

Allwood, C. M., Knutsson, J., & Granhag, P. A. (2006). Eyewitnesses under influence: How feedback affects the realism in confidence judgements. *Psychology,*,

Crime & Law, 12(1), 25-39.

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., & 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.

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.

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.

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

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., & 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.

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

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.

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.

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.

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

Hyman, I. E., JR. (1994). Conversational remembering: Story recall with a peer versus for an experimenter. *Applied Cognitive Psychology*, 8(1), 49-66.

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.

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.

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

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.

Merkle, E. C. (2009). The disutility of the hard-easy effect in choice confidence. *Psychonomic Bulletin & Review*, 16(1), 204-213.

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

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.

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

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.

Robinson, M. D., Johnson, J. T., & Robertson, D. A. (2000). Process versus content in eyewitness metamemory monitoring. *Journal of Experimental Psychology. Applied*, 6(3), 207-221.

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 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 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.

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.

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., & Bradfield, A. L. (1999). Distortions in eyewitnesses' recollections: Can the postidentification-feedback effect be moderated? *Psychological Science*, 10(2), 138-144.

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

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.

Content analysis of eyewitnesses repeated recalls and discussions

Paper III

Content analysis of eyewitnesses' repeated recalls and discussions

Farhan Sarwar[1], Carl Martin Allwood[2], and Åse Innes−Ker[1]
[1]Department of Psychology, Lund University, Sweden
[2]Department of Psychology, University of Gothenburg, Sweden

Correspondence concerning this article should be addressed to Farhan Sarwar, Department of Psychology, Lund University, P.O. Box 213, SE-22100 Lund, Sweden. Tel.: +46 46-222 05 72; fax: +46 46-222 42 09. E-mail address: farhan.sarwar@psychology.lu.se

Abstract

Eyewitnesses retell and discuss the witnessed event with different people. It is of forensic importance to understand the change in eyewitness memory over successive discussions and how such discussions affect confidence levels for retrieved information. Forty-four participants watched a kidnapping film clip and were subsequently randomized to two conditions: 1) Retell, where participants retold the film events 5 times over 3 weeks, 2) Lab-discussion, where participants retold and discussed the film events 5 times over 3 weeks with different confederates. Finally, all participants at day 21 gave an open free recall and, three days later, confidence judged their free recall. Lab-discussions and Retellings were recorded and transcribed. The results showed that retellings and discussions cause people to recall previously unrecalled information (reminiscence) and recall more correct items (hypermnesia). The most repeated information was more often reported in the final test. Moreover, increase in confidence over repetitions (with accuracy constant) was found for forensically peripheral information but not for forensically central information (the reiteration effect). A general conclusion is that forensically central information was found to be less vulnerable to memory distortions and the reiteration effect.

Keywords: Eyewitnesses, Confidence, Reminiscence, Hypermnesia, Free recall, Retellings, Central information, Peripheral information

Introduction

Eyewitness testimony has a paramount importance in the criminal justice system. However, both accuracy and confidence in the correctness of recalled information are vulnerable to distortions due to factors that affect both of them as well as factors that affect each of them independently. This means that accuracy can be altered without any change in corresponding confidence and vice versa (Leippe & Eisenstadt, 2007). One of the challenges for forensic professionals is to evaluate the credibility of the testimony. An eyewitness confidence in the accuracy of his/her statements is the most commonly used yardstick by the legal system to assess credibility (Wells & Bradfield, 1999; Wells, Lindsay, & Ferguson, 1979). In cases where no circumstantial evidence is available to cross-check the eyewitness claims, confidence may be the only option available to evaluate the credibility of the eyewitness statements. For this reason it is important to understand the degree to which the level of confidence corresponds to the level of accuracy. The level of correspondence between accuracy and confidence is called the *realism* in eyewitness confidence judgments.

There are many factors that can affect an eyewitness accuracy and confidence in his or her memory of the experienced event. Examples of such factors are eyewitness exposure to the media coverage of the witnessed event (Ost, Granhag, Udell, & Hjelmsäter, 2008), and how often an eyewitness retells and answer questions about the event (Lane, Mather, Villa, & Morita, 2001). An example of such a situation is when an eyewitness discusses the witnessed event with their family and friends. These discussions can potentially distort eyewitness confidence and accuracy. This factor is important because eyewitnesses often share the witnessed event with their family and friends many times (Paterson & Kemp, 2006). The main aim of these discussions is to share whats new (Skowronski & Walker, 2004). Consequently, the way an event is retold and discussed may distort the eyewitness memory (Lane et al., 2001) and confidence (Hertwig, Gigerenzer, & Hoffrage, 1997).

The aim of this paper is to study the quantitative and qualitative change in eyewitness memory over successive retellings and discussions of an experienced event, and the impact these discussions may have on the level and realism in the corresponding confidence judgments done at the testing phase. In this introduction, we first review research on the impact of multiple discussions with other persons on the eyewitness memory of forensic episode. Then we discuss the implications of these multiple discussions for confidence judgments. Finally, we present the hypotheses of this research.

Impact of multiple retellings and discussions on eyewitness memory

Two phenomena of relevance in the context of multiple retellings and discussions are *reminiscence* and *hypermnesia*. *Reminiscence* refers to the phenomenon that each retrieval attempt causes people to recall previously unrecalled information, both correct and incorrect (La Rooy, Pipe, & Murray, 2005; Mulligan, 2006; Payne, 1987). *Hypermnesia* refers to the phenomenon that the number of correct items recalled increases with repeated recall attempts (Bornstein, Liebel, & Scarberry, 1998; La Rooy, Pipe, & Murray, 2007; La Rooy et al., 2005; Mulligan, 2001, 2006; Payne,

1987).

Research support for reminiscence has been consistent across studies (e.g. La Rooy et al., 2005; Mulligan, 2006; Turtle & Yuille, 1994). In contrast, the support for hypermnesia has been mixed. A common method used to study hypermnesia was developed by Erdelyi and Becker (1974). In this method, participants are presented with pictures or words, and later participants are asked to recall a fixed number of items and they are allowed to make guesses. Studies using this method have shown that the number of correct items retrieved increases with repeated recall attempts (hypermnesia, (e.g. Henkel, 2004; Scrivner & Safer, 1988). Studies using real life events have also found support for hypermnesia. For example, Bluck, Levine, and Laulhere (1999) interviewed participants eight months after the televised verdict of the O. J. Simpson case. The results showed that both absolute and cumulative number of correct items retrieved by the participants increased from the first interview to the third interview. In contrast, La Rooy et al. (2005) and La Rooy et al., (2007) using a procedure similar to Bluck et al. (1999), could not find support for hypermnesia. Other studies using different methods also failed to find support for hypermnesia (e.g. Turtle & Yuille, 1994). Research has also shown that though people recall more correct items in successive recalls, the number of incorrect details also increases with repeated recall attempts (Henkel, 2004).

Multiple *discussions* are likely to cause higher reminiscence than mere retellings. The reason is that in discussions three factors are likely to facilitate reminiscence. First, just as for retelling, an eyewitness would show reminiscence simply because of the repeated recall attempts (La Rooy et al., 2005; Payne, 1987; Turtle & Yuille, 1994). Second, probing questions asked by discussion partners can make an eyewitness recall information that was not recalled by him/her earlier when describing the forensic event to the discussion partner. Third, discussions are interactive in nature where exchange of information takes place (Pasupathi, 2001). Eyewitnesses may then incorporate the information contributed by their discussion partners into their memories (Loftus, 1979), and this information may be communicated to the next listener (Gabbert, Memon, & Allan, 2003; Marsh, 2007; Tversky & Marsh, 2000). This would increase reminiscence in eyewitness reports over successive discussions. Since hypermnesia depends both on reminiscence and recall of previously recalled information (La Rooy et al., 2005; Payne, 1987), increase in reminiscence is likely to cause hypermnesia as well.

Two other relevant phenomena in the context of multiple retellings and discussions are the *testing effect* (Roediger III & Karpicke, 2006a) and *retrieval induced forgetting* (Coman, Manier, & Hirst, 2009). The *testing effect* hypothesis claims that testing improves memory (Cull, 2000; Roediger III & Karpicke, 2006a, 2006b). Research has shown that the number of times a test taken is associated with better memory performance as compared to rehearsing the same content with access to the original material the same number of times (Karpicke & Roediger III, 2007, 2008). In addition, results show that the testing effect may improve memory of related content even if that content is not tested for (Chan, McDermott, & Roediger III, 2006).

According to the *retrieval induced forgetting* hypothesis, multiple retrieval attempts can cause forgetting of unrecalled information (Coman et al., 2009; MacLeod, 2002). However, retrieval induced forgetting may have little or no effect if the learned content is well integrated (Anderson, 2003; Anderson & McCulloch, 1999).

The testing effect is especially relevant in the case of mere retellings. The reason is that retellings are like active repetitions without access to the original material (Roediger III & Karpicke, 2006a, 2006b). Moreover, new information recalled (information that was not brought up in earlier retellings) during the successive retellings is likely to be correct (hypermnesia). The reason is that the testing effect is also known to have memory benefits for the material that is related to the tested material but was not retrieved earlier (Chan et al., 2006). Consequently, mere retellings are likely to result in better memory accuracy.

Retrieval induced forgetting could affect eyewitness memory both in retellings and multiple discussions (Coman et al., 2009; MacLeod, 2002). However, research has shown that the effect of retrieval induced forgetting is limited to the material that is not well integrated (Anderson, 2003; Anderson & McCulloch, 1999). So, to understand the impact of retrieval induced forgetting, it is important to look at the type of content investigated. A forensic event includes two types of information: 1.) Information about the action details (i.e. who did what in the event, below called *forensically central information*), and 2.) Descriptions of persons and objects involved in the event (i.e. height of the suspect, type of the weapon used, etc., below called *forensically peripheral information*; (Heuer & Reisberg, 1990; Sarwar, Allwood, & Innes-ker, 2010). Action details tend to be well connected with each other since one action will lead to the next action. For this reason, action details may not be affected by retrieval induced forgetting (Anderson, 2003; Anderson & McCulloch, 1999). However, the action details are likely to benefit from mere retellings because of the testing effect (Roediger III & Karpicke, 2006a).

In contrast, the descriptive details may not be especially well connected to each other. For example, there is no obvious relation between the height of a person and the color of his clothes. Consequently, the descriptive details would probably be more vulnerable to retrieval induced forgetting (Coman et al., 2009; MacLeod, 2002).

Consequences of Multiple Discussions and Retellings for the Confidence Judgments

Whenever memories are shared they can be assumed to have been confidence judged as part of the process of reporting them (Koriat & Goldsmith, 1996). Partly these feelings of confidence are based on a cue that the source of information is known (Kelley & Lindsay, 1993). However, confidence judgments can be distorted by other cues that mediate the feeling of confidence; for example, how swiftly the information was accessed (Leippe & Eisenstadt, 2007) or an individuals belief about his memory ability (Perfect, 2004).

In the case of eyewitness mere retellings and multiple discussions with family and friends, the eyewitness feeling of confidence can be changed because of the *reiteration effect* (Hertwig et al., 1997). According to the reiteration effect hypothesis, multiple assertions of a statement cause the confidence to inflate without any change in corresponding accuracy (Hertwig et al., 1997). The basic cause behind the reiteration effect has been suggested to be increased retrieval fluency (i.e. subjective feeling of how easily the information was recalled;(Shaw, 1996; Shaw & McClure, 1996). However, the research support for the reiteration effect has been mixed, for detail discussions see Shaw, McClure, and Dykstra (2007) and Leippe and Eisenstadt (2007). We speculate

that the reiteration effect may depend on the nature of the content that has been confidence judged. If the content is well integrated (i.e. action details), it may not cause a reiteration effect. The reason is that additional retrievals may not increase the retrieval fluency for well-integrated content. Note that Hertwig et al. (1997) found that the reiteration effect increased most for the first retrievals compared to later retrievals.

Overview of the Present Study

This research investigated the quantitative and qualitative changes in memory and confidence over five successive retellings and discussions. The participants first watched a short film. Then they had five sessions over 3 weeks involving only retelling (Retell condition) or retelling and discussion (Lab-discussion condition), before they had the final recall test. Last, the participants confidence judged their statements after 3 to 4 days.

Participants retellings and lab-discussions were recorded and later transcribed. In this study, these transcripts are analyzed together with the final recall and confidence judgments.

Differential effects of retellings and discussions over time on accuracy and confidence levels are expected for different types of forensic information. Furthermore, we also investigate the *reiteration effect* hypothesis within the context of mere retellings and discussions for different types of information. The Lab-discussion was the critical condition as it was intended to investigate how discussions may impact the quantity and quality of recall and confidence. The purpose of the Retell condition was to set a control for the effects of mere retelling on the quantity and quality of memory and confidence. In this way, the additional effects of the discussion aspect of the Lab-discussion condition could be better understood. This study uses data collected in the context of a large study (Farhan Sarwar, Carl Martin Allwood, & Åse Innes-Ker, 2010), but the present data and analyses have not been reported earlier.

Hypotheses

Hypothesis 1

We predicted that the amount of reminiscence and hypermnesia would increase more over the successive discussions in the Lab-discussion condition as compared with the Retell condition. One reason is that the questions posed by the discussion partner would cause participants to recall more new information in the Lab-discussion condition as compared with the Retell condition.

Hypothesis 2

We expected that information that is more often discussed or retold in the experimental sessions is more likely to be reported in the final test. One reason for this prediction is the testing effect reported by Roediger III and Karpicke (2006b).

Hypothesis 3

We also expected that the confidence would be higher for information repeated more often as compared with the confidence for information that was repeated less often (accuracy kept constant) because of the reiteration effect (Hertwig et al., 1997). However, as discussed above, we expected that the reiteration effect would be limited to the forensically peripheral information because it might be less well integrated than the forensically central information. We speculated that the retrieval fluency of less well integrated information might increase more over successive retrievals due to the information becoming more associated with each retrieval.

Method

Participants

Forty-four students (37 women) from Lund University participated in the study. The mean age of the participants was 25 years (18-47). On successful completion of the experiment, each participant was rewarded with a movie ticket worth 90 SEK (approximately US$ 12). Initially, we recruited 46 participants (23 for each condition). There were 7 and 4 dropouts from the Lab-discussion condition and the Retell condition, respectively. These participants were replaced. There were two more dropouts from the Lab-discussion condition at the end who were not replaced.

Design

Two between-subjects conditions were used in this study. These conditions were: Retell (n= 23) and Lab-discussion (n = 21; one participant only attended 4 out of 5 sessions)[1].

Material

Videotape

A 3 minute and 50 second long color film was used. This film showed a woman arriving at a bus stop and few minutes later being kidnapped by two men from the buss stop. This film has been used in previous research (e.g., Allwood, Ask, & Granhag, 2005; Granhag, 1997).

Confidence judgment scales

Participants used a confidence scale to rate their confidence for the detailed parts of the free recalls (low-level statements). This scale had 11 levels beginning at 0% (Completely sure that I remember wrong) and then 10%, 20%, 30%, to 100% (Completely sure that I remember correct).

[1]We analyzed the data with and without this participant. The results did not change. Therefore this participant was included in the final analysis.

Procedure

The first session of the experiment was conducted in small groups of between four to eight participants. The participants were told that the experiment was about human perception in forensic situations. After signing the consent form, participants watched the short film and were randomized into the conditions. The experimenter then scheduled five sessions over a 20-day period with participants individually.

In each of the five sessions, participants in the Retell condition simply told the events of the short film to the experimenter. The participants were instructed to tell whatever they remembered about the film. The experimenter posed no questions to the participants.

The participants in the Lab-discussion condition first retold the events in the film to a confederate (each time new and unknown to the participant). Then the participants discussed the events of the film with the confederate. The confederate asked spontaneous questions about the film after listening to the events of the film. For this purpose, 105 confederates were recruited. Each confederate participated in only one discussion. All the discussions and retellings were recorded on a MP3 recorder.

All participants took the two memory tests on the 21st day. First, the participants typed in Microsoft Office Word whatever they remembered about the events of the film (open free recall test). Second, they answered 44 focused questions, each with two answer alternatives, on the film. On the 24th or 25th day, the participants made their last visit to the lab to confidence judge their free recall statements and focused questions. Participants were then debriefed and dismissed.

Preparation of Material for Participants Confidence Judgments

In order to prepare for the confidence judgments, the free recall of each participant was broken down into single units of information. We followed the principles described by Allwood, Ask, and Granhag (2005) when dividing the free recall into single units (pieces of information). These were as follows: (1) statements that were about actors and actions carried out were rendered as one unit. For example a car passed by was used as a single unit. (2) An object with one attached attribute was used as one unit. For example the car was blue was used as one unit. (3) If an object was described by more than one attribute, the additional attributes were rendered as separate units. For example the tall blond woman was rendered as two units. For the purpose of reminding participants about the context in which they mentioned each unit, one or two short sentences related to that item were attached to that specific item. The items to be confidence judged were underlined while the reference items were put into brackets. Finally, the 11-point confidence scale was inserted directly below each single item.

Preparation of the Material for Analysis

The recordings of the participants five discussion and retelling sessions were transcribed in full detail (whatever was uttered by the participants and the confederates). Only information from the participants was analyzed. The information in these transcriptions was also broken down into single units of information using the same tech-

nique described above. Next, this information was scored for correct information, incorrect information, and mean accuracy. Then the information provided by participants was traced from session 1 to session 6. Each piece of information was traced to the particular session it was retold or discussed. We ended with 31 combinations with respect to when information was discussed or retold during the respective five sessions and correspondingly reported at session 6. These session combinations were (1), ..(1, 2),..(1, 2, 3), (2, 3, 4, 5). (1, 2, 3, 4, 5). Then there was information that was not discussed or retold during any of the sessions but that was reported at the testing session (session 6). How these combinations were used is explained below.

Classification of Information into Different Types of Information

In order to investigate how the retellings and discussion would affect the participants memory and confidence for different types of information, the information in the transcriptions and the free recalls was divided into different categories. First, in order to separate out the irrelevant information, participants statements were divided into *Forensic* and *Non-forensic* statements. In a forensic context it is useful to separate out *Non-forensic* information such as information about houses, roads, surroundings, etc., because it is either not needed to solve the crime or it can be collected from the crime scene without the help of an eyewitness (However, it can still be used to evaluate the credibility of the witness).

Second, the Forensic statements were divided into *Forensically central* and *Forensically peripheral* categories by following the plot relevancy model by Heurer and Reisberg (1990). *Forensically central* information was defined as the information that cannot be changed without changing the story of the event. This category basically separates out the action details because the change in actions causes the story to change. For example, the suspect stabbed the victim with a knife. *Forensically peripheral* information was defined as the information that can be changed without changing the story of the event. This category identifies the descriptive details because the change in descriptive details does not basically change the story (Heuer & Reisberg, 1990). Examples are the suspects height and the color of his shirt.

Results

Number of Correct items, Incorrect items, and Accuracy Over the Six Sessions and for the Two Conditions

In order to investigate the difference in the number of correct items, incorrect items, and accuracy over the six sessions for the two types of information (Forensically central and Forensically peripheral) and between the two conditions (Retell condition and Lab-discussion condition) the data were subjected to 6x2x2 mixed ANOVA. The within-subjects factors were session (6 sessions, where participants in five sessions either retold or discussed the witnessed event and where the sixth session was the testing session) and information type (Forensically central and Forensically peripheral), and the between-subjects factor was condition (Retell, and Lab-discussion). The analysis was performed separately for the three dependent measures: correct items, incorrect items, and accuracy. The means and standard deviations are shown in Table 1.

Table 1. Means (and SDs) for Correct items, Incorrect items, and Accuracy, for Forensically central, Forensically peripheral, and Non forensic information

	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6
Correct items						
Forensic central						
Retell	17.9(5.5)	17.4(5.7)	16.4(5.5)	16.0(4.5)	17.0(5.4)	24.6(5.3)
Lab-discussion	16.9(4.4)	17.0(4.6)	19.7(4.5)	17.7(3.4)	18.5(4.0)	29.7(7.8)
Forensic peripheral						
Retell	4.5(3.9)	4.1(2.5)	5.0(3.3)	4.4(3.3)	4.9(3.0)	7.5(4.7)
Lab-discussion	5.6(3.4)	7.5(4.3)	6.5(3.7)	7.3(4.2)	6.8(3.6)	10.4(4.4)
Incorrect items						
Forensic central						
Retell	1.3(1.1)	1.6(1.7)	1.7(2.1)	1.5(1.8)	1.6(2.1)	1.9(1.5)
Lab-discussion	1.3(1.2)	1.2(1.3)	1.2(1.2)	1.2(1.0)	1.7(1.1)	3.9(3.5)
Forensic peripheral						
Retell	1.0(1.7)	1.3(1.6)	1.5(1.7)	1.5(1.9)	1.1(1.8)	2.4(2.4)
Lab-discussion	2.5(2.6)	2.9(2.4)	2.4(1.3)	2.9(2.2)	3.8(2.3)	5.0(3.7)
Accuracy						
Forensic central						
Retell	.91(.04)	.90(.10)	.88(.14)	.90(.10)	.89(.13)	.93(.05)
Lab-discussion	.93(.07)	.92(.08)	.94(.04)	.94(.05)	.90(.09)	.90(.08)
Forensic peripheral						
Retell	.80(.20)	.67(.17)	.70(.18)	.65(.24)	.76(.24)	.74(.19)
Lab-discussion	.67(.26)	.67(.26)	.68(.17)	.68(.20)	.57(.21)	.65(.14)

Correct items

The results showed that there was a significant main effect for session, Wilks' Lambda= .18, $F(5, 23) = 20.97$, $p < .001$; $\eta^2 = .06$. The contrasts revealed that the participants reported significantly more correct items in sessions 6 as compared to session 1, $F(1, 27) = 65.46$, $p < .001$; $\eta^2 = .14$, session 2, $F(1, 27) = 71.74$, $p < .001$; $\eta^2 = .12$, session 3, $F(1, 27) = 70.62$, $p < .001$; $\eta^2 = .11$, session 4, $F(1, 27) = 116.68$, $p < .001$; $\eta^2 = .13$, and session 5, $F(1, 27) = 92.09$, $p < .001$; $\eta^2 = .12$. The five experimental sessions did not differ in the number of correct items reported.

A main effect for information type showed that the participants reported more correct items classified as Forensically central than Forensically peripheral, $F(1, 27) = 189.36$, $p < .001$; $\eta^2 = .04$. A main effect for condition also showed that the participants in the Lab-discussion condition reported more correct items than the participants in the Retell condition, $F(1, 27) = 4.15$, $p < .05$; $\eta^2 = .13$.

There was a two-way interaction effect for session and information type, $F(5, 23) = 5.91$, $p < .001$; $\eta^2 = .14$. It showed that the number of correct items reported in Forensically central and Forensically peripheral categories over the sessions were different. Inspection of means showed that over the sessions the reported number of correct Forensically central items increased more as compared to the number of correct Forensically peripheral items.

There was also a three-way interaction effect for session, information type, and condition, $F(5, 23) = 3.91$, $p < .01$; $\eta^2 = .18$. It showed that the number of correct Forensically central and Forensically peripheral items reported over the sessions differed between the conditions. Inspection of means showed the increase in the reported number of correct Forensically central and Forensically peripheral items over the sessions was greater in the Lab-discussion condition as compared to the Retell conditions.

Incorrect items

The results showed that there was a significant main effect for session, Wilks' Lambda= .58, $F(5, 35) = 5.18$, $p < .001$; $\eta^2 = .12$. The contrasts revealed that in session 6 participants reported significantly more incorrect items than in session 1, $F(1, 39) = 19.26$, $p < .001$; $\eta^2 = .12$, session 2, $F(1, 39) = 13.27$, $p < .001$; $\eta^2 = .10$, session 3, $F(1, 39) = 20.56$, $p < .001$; $\eta^2 = .11$, session 4, $F(1, 39) = 20.30$, $p < .001$; $\eta^2 = .12$, and session 5, $F(1, 39) = 12.67$, $p < .001$; $\eta^2 = .07$. The five experimental sessions did not differ in the number of incorrect items reported.

A main effect for the information type showed that the participants reported significantly more incorrect Forensically central items than incorrect Forensically peripheral items, Wilks' Lambda= .86, $F(1, 39) = 6.59$, $p < .01$; $\eta^2 = .18$. A main effect for condition showed that the participants in the Lab-discussion condition reported significantly more incorrect items than the participants in the Retell condition, $F(1, 39) = 7.47$, $p < .01$; $\eta^2 = .16$.

The results showed a significant two-way interaction effect for session and condition, Wilks' Lambda= .73, $F(5, 35) = 2.62$, $p < .04$; $\eta^2 = .15$. Inspection of the means showed that the increase in the number of incorrect items reported over the sessions in the Lab-discussion condition was greater as compared with the incorrect items reported over the sessions in the Retell condition. The results also showed a

Table 2.
Mean number of New Items Reported (Both Correct and Incorrect) by Each Participant in Sessions 2, 3, 4, 5, and 6 in Retell and Lab-discussion Conditions

New Items	Session2	Session3	Session4	Session5	Session6
Forensic Central					
Retell	4.26	2.04	2.22	0.91	1.83
Lab-discussion	5.43	3.81	1.86	1.19	1.05
Forensic Peripheral					
Retell	1.26	0.87	0.39	0.3	2.04
Lab-discussion	4.14	2.19	2	1.43	1.57

two-way interaction effect for information type and condition, Wilks' Lambda= .80, $F(1, 39) = 9.68$, $p < .01$; $\eta^2 = .17$. Inspection of the means showed that the participants in the Lab-discussion condition reported more incorrect Forensically central and Forensically peripheral items than the participants in the Retell condition.

Accuracy

A main effect for information type showed that participants were more accurate for Forensically central information as compared to Forensically peripheral information, Wilks' Lambda= .28, $F(1, 27) = 69.38$, $p < .001$; $\eta^2 = .06$. There was also a significant two-way interaction effect for session and condition, Wilks' Lambda= .64, $F(5, 25) = 2.61$, $p < .05$; $\eta^2 = .15$. Inspection of means 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.

New Items Retold or Discussed (Both Correct and Incorrect)

To investigate Reminiscence, the number of new items reported by each participant in each session was calculated for both conditions. The means are shown in Table 2. Using a small program, which was written in Microsoft Excel, we collected the new items in each session. This program collected the new items for each session but not for each participant; therefore, it was not possible to compare the sessions by use of significance tests. A 2x2 mixed ANOVA was conducted where the within-subject factor was information type and a between-subjects factor was condition. A significant main effect showed that participants reported more new items regarding Forensically peripheral information than the Forensically central information, Wilks' Lambda= .58, $F(1, 8) = 5.79$, $p < .04$; $\eta^2 = .38$. Although we could not do a significance test for the sessions, inspection of means showed that the participant in the Lab-discussion condition reported more new items in sessions 2-5 than the Retell condition. In contrast, participants in the Retell condition reported more new items in session 6 than the participants in the Lab-discussion condition.

Effects of the Number of Times an Item is Retold or Discussed

To test if the number of times an item was retold or discussed was related to the reporting of that item at the testing phase, we compared the items that were retold or discussed in 1 or 2 sessions with the items that were retold or discussed in 3 or 4 sessions. The information retold or discussed during all the five sessions was not used. There were two main reasons for this. First, the information that was retold or discussed in all the five sessions was overwhelmingly reported in session 6 and this made this combination an outlier. Second, including this combination in the group where information was discussed or retold 3 to 4 times does not change the results. By excluding items that were retold 5 times we also avoided the problem of whether the items that were retold 3 times should be counted as belonging to the group with items retold a low or a high number of times.

In this context, the data were submitted to 2x2x2 mixed ANOVA where the within-subjects factors were information type (Forensically central and Forensically peripheral) and number of repetitions (1-2 repetitions, and 3-4 repetitions), and the between-subject factor was condition (Retell condition and Lab-discussion condition). The means and standard deviations are shown in Table 3.

Reported items

A main effect for information type showed that the Forensically central items were reported more than the Forensically peripheral items, Wilks' Lambda= .55, $F(1, 22) = 17.79$, $p < .001$; $\eta^2 = .09$. A main effect for the number of repetitions showed that the items that were reported 3 to 4 times were reported more at session six, Wilks' Lambda= .77, $F(1, 22) = 6.61$, $p < .02$; $\eta^2 = .06$. The results also showed an interaction effect of the information type and the number of repetitions, Wilks' Lambda = .63, $F(1, 27) = 13.05$, $p < .002$; $\eta^2 = .21$. It showed that the number of repetitions had a different effect on the two types of information in terms of their chance to be reported at session 6. Inspection of means showed that most of the Forensically central items that were reported at session 6 were repeated 3-4 times in the five experimental sessions and this difference was significant, $t(14) = -2.29$, $p < .04$, Cohen's $d = -.32$. In contrast, the Forensically peripheral items that were reported at session six were not affected by how many times they were repeated in the five experimental sessions.

Non-reported items

A main effect for information type showed that the amount of Forensically central information was significantly higher than the Forensically peripheral information for the information that was repeated during the five sessions but not reported in session 6, Wilks' Lambda= .41, $F(1, 14) = 20.06$, $p < .001$; $\eta^2 = .06$.

A main effect for the number of repetitions showed that the information that was not reported at session six was such information that was repeated less than three times during the five sessions as compared to the information that was repeated 3 to 4 times, Wilks' Lambda= .45, $F(1, 14) = 17.21$, $p < .001$; $\eta^2 = .27$.

The results also showed a two-way interaction effect for information type and condition, Wilks' Lambda= .64, $F(1, 14) = 7.86$, $p < .01$; $\eta^2 = .02$. Inspection of means showed that Forensically central information that was not reported at session

Table 3.
Means (and SDs) for the Number of Times Items Repeated and Discussed Less than Three Times or More Than Three Times During the Experimental Phases and Finally Reported at the Testing Session

	Reported information		Not reported information	
	Up to three repetitions	More than three repetitions	Up to three repetitions	More than three repetitions
Forensically central information				
Retell	28.10(1.89)	71.90(9.56)	83.20(11.45)	16.80(2.39)
Lab-discussion	34.03(2.95)	65.97(9.03)	78.49(11.60)	21.51(1.97)
Forensically peripheral information				
Retell	43.10(1.84)	56.90(2.54)	82.50(3.78)	17.50(.55)
Lab-discussion	52.66(2.79)	47.34(2.50)	87.25(11.19)	12.75(1.08)

6 was repeated more during the five experimental sessions in the Retell condition than in the Lab-discussion condition. Forensically peripheral information that was not reported at session 6 was repeated more during the five experimental sessions in the Lab-discussion condition than the Retell condition. The results also showed a three-way interaction for information type, number of repetition, and condition, Wilks' Lambda= .75, $F(1, 14) = 4.77$, $p < .05$; $\eta^2 = .08$. It showed that the forensically central and the forensically peripheral information that was repeated during the five sessions but was not reported in the session 6 were affected by the number of repetitions and the also by the two conditions differently.

Effects of Earlier vs Later Retelling and Discussions of Information

To test if early or late retelling or discussion affects the reporting of an item at the testing phase we compared the items that were discussed or retold earlier with the items that was discussed or retold later in the experimental sessions.

In preparation for the analysis, the means of the session numbers of 30 of the session combinations were computed (see the Method section above for how the combinations were made). The 31st combination where the information was discussed or retold during all the five sessions was not used. The combinations with a mean score less than three were allocated to the group of earlier retellings or discussions. The combinations with a mean score greater than three were allocated to the group of later discussions or retellings. In this way each group (early or later) received 12 combinations of cases. There were six cases with the mean score of three. These cases were randomly divided into the two groups. Thus, finally each group had 15 cases each.

In this context the data was submitted to 2x2x2 mixed ANOVA where the within-subjects factors were information type (Forensically central and Forensically peripheral) and earlier vs. later repetitions (early repetitions, and late repetitions), and the between-subjects factor was condition (Retell condition and Lab-discussion condition). The means and standard deviations are shown in Table 4.

Reported items

The results showed that there was a main effect of information type, Wilks' Lambda= .62, $F(1, 20) = 12.47$, $p < .002$; $\eta^2 = .16$, which showed that the participants reported more Forensically central items than Forensically peripheral items. A main effect for earlier vs. later repetitions showed that the information that was on average repeated later during the five experimental sessions was reported significantly more at session 6 than the information that was on average reported earlier during the five experimental sessions, Wilks' Lambda= .82, $F(1, 20) = 4.34$, $p < .05$; $\eta^2 = .02$. The contrasts for the two types of information did not reach significance for the information retold or discussed earlier or later in the sessions.

Non-reported items

The results showed there was a main effect of the information type, Wilks' Lambda= .69, $F(1, 16) = 7.07$, $p < .02$; $\eta^2 = .12$. No other differences reached the significance

Table 4.
Means (and SDs) for the Number of Times Items Repeated and Discussed Earlier or Later during the Experimental phases and were Reported at the Testing Session

	Reported information		Not reported information	
	Earlier repetitions	Later repetitions	Earlier repetitions	Later repetitions
Forensically central information				
Retell	41.67(8.57)	58.33(9.60)	61.34(15.28)	38.66(5.28)
Lab-discussion	41.64(7.01)	58.35(7.87)	64.01(11.72)	35.99(5.58)
Forensically peripheral information				
Retell	35.55(1.51)	61.45(2.71)	61.39(4.68)	38.61(1.47)
Lab-discussion	48.47(1.85)	51.53(2.20)	60.56(10.59)	39.43(7.07)

level $p < .05$.

Accuracy and Confidence Judgments of the Information that was Retold or Discussed Earlier or Later in the Experimental Sessions

To test if early or late retelling or discussion was related to the accuracy and confidence level of the items at the testing phase we compared the items that were discussed or retold earlier with the items that was discussed or retold later in the experimental sessions. The means of 31 combinations were taken (see the method for how the combinations were made). In this analysis the information that was only reported in session 6 was also used. Consequently, there were 32 combinations. The combinations with the mean session score less than three were allocated to the group of earlier discussions or repetitions. The combinations with a mean session score higher than three were allocated to the group of later retellings or discussions. There were seven cases with a mean score of three. These cases were randomly divided between the two groups. In this way there were 16 combinations in each group.

The data was submitted to 2x2x2 mixed ANOVA where the within-subject factors were information type (Forensically central and Forensically peripheral) and repetition type (earlier vs. later repetitions), and the between-subjects factor was condition (Retell condition and Lab-discussion condition). The means and standard deviations are shown in Table 5.

Accuracy

The results showed that there was a main effect of information type, Wilks' Lambda= .39, $F(1, 23) = 36.06$, $p < .001$; $\eta^2 = .21$. There was no difference in accuracy between the items retold or discussed earlier or later during the experiment.

Confidence

The results showed that there was a main effect of information type, Wilks' Lambda= .50, $F(1, 24) = 23.71$, $p < .001$; $\eta^2 = .24$, in that confidence was higher for Forensically central information than for Forensically peripheral information. There was also a three way interaction between information type, repetition, and condition, Wilks' Lambda= .84, $F(1, 24) = 4.67$, $p < .04$; $\eta^2 = .04$. Inspection of means showed that the participants in the Retell condition had higher confidence for Forensically central information that was retold in the earlier sessions than the information that was retold in the later sessions. In contrast, for Forensically peripheral information the participants in the Retell condition had higher confidence for the information that was retold later than the information that was retold earlier. However, for the Lab-discussion condition this pattern was vice-versa for each type of information. However, the difference between the confidence levels for the items retold or discussed earlier or later during the experiment was not statistically significant.

Table 5.
Means (and SDs) for Accuracy and Confidence Reported for the Information Repeated Early or Late in the Retelling and Lab-discussion conditions

	Accuracy		Confidence	
Confidence	Earlier repetitions	Later repetitions	Earlier repetitions	Later repetitions
Forensically central information				
Retell	.98(.05)	.87(.14)	93.57(4.04)	89.85(4.23)
Lab-discussion	.97(.06)	.94(.13)	85.96(9.72)	90.48(5.58)
Forensically peripheral information				
Retell	.73(.26)	.78(.29)	82.28(11.11)	87.02(8.24)
Lab-discussion	.69(.24)	.64(.25)	79.14(8.45)	77.46(10.94)

Testing the Reiteration Effect

The presence of a reiteration effect was investigated by comparing the accuracy and confidence level for the items that were discussed or retold up to three times with the accuracy and confidence level for the items that were discussed or retold more than three times.

For this analysis, only the information that was retold or discussed during one or more of the five sessions and reported at the session 6 was used. The reason was that the confidence judgments were made in session 6 (testing session). In this case we had 32 possible combinations of sessions. Thirty-one of these combinations were the combinations where participants retold and discussed the information during one or more of the five sessions and in session 6. The 32nd combination was the information that was only reported at session 6. The 32 combinations were divided into two groups. One group included the combinations where the information was discussed or retold 1 to 3 times and the other group included the combinations where the information was discussed or retold 4 to 6 times in the six sessions. In the analysis the confidence levels for the two groups were compared in order to study if the average confidence was higher in the group where the items had been repeated more times (the reiteration effect). The accuracy levels for the two groups were also compared in order to study if the accuracy also increased with the increase in the number of times a piece of information was repeated. This is relevant since an analysis of the presence of the reiteration effect assumes that accuracy is controlled for.

In this context the accuracy and confidence scores were separately submitted to 2x2x2 mixed ANOVAs where the within-subject factors were information type (Forensically central and Forensically peripheral) and Number of repetitions (1-3 repetitions, and 4-6 repetitions), and the between-subjects factor was condition (Retell condition and Lab-discussion condition). The means and standard deviations are shown in Table 6.

Accuracy

The results showed that there was a main effect of the type of information type, Wilks' Lambda= .41, $F(1, 25) = 36.37$, $p < .001$; $\eta^2 = .24$, which showed that the participants had higher accuracy scores for Forensically central items than Forensically peripheral items. There w no difference between the accuracy of the items retold or discussed 1 to 3 times or 4 to 6 times.

Confidence

The results showed that there was a main effect of the type of information type, Wilks' Lambda= .57, $F(1, 26) = 19.77$, $p < .001$; $\eta^2 = .19$, in that the participants showed a higher confidence level for the Forensically central items than for the Forensically peripheral items. There was also a main effect for the number of repetitions, Wilks' Lambda= .75, $F(1, 26) = 8.63$, $p < .01$; $\eta^2 = .07$. More repeated items showed higher confidence. The results also showed a two-way interaction between information type and number of repetitions, Wilks' Lambda= .81, $F(1, 26) = 6.10$, $p < .02$; $\eta^2 = .05$. This interaction showed that the number of repetitions was differently related to Forensically central and Forensically peripheral information. Inspection of means

Table 6.
Means (and SDs) for Accuracy and Confidence For Items Reported Up to Three Times and More Than Three Times in the Retelling and Lab-discussion conditions

	Accuracy		Confidence	
Confidence	Up to three repetitions	More than three repetitions	Up to three repetitions	More than three repetitions
Forensically central information				
Retell	.93(.12)	.94(.11)	91.69(4.86)	91.59(3.61)
Lab-discussion	.92(.13)	.98(.05)	86.77(7.95)	88.77(9.13)
Forensically peripheral information				
Retell	.74(.22)	.82(.29)	74.60(25.17)	88.54(6.84)
Lab-discussion	.64(.19)	.66(28)	74.22(7.07)	82.26(10.28)

showed that there was no difference in the confidence level for the forensically central information with fewer or more repetitions. In contrast, the confidence level for Forensically peripheral information was higher for the items that were repeated more often.

Discussion

The purpose of this study was to investigate the change in number of correct items, number of incorrect items, and accuracy in eyewitness memory reports for the Forensically central and peripheral information over five retellings and discussions. Further, another purpose was to investigate if the eyewitnesses confidence for different types of information was higher for the items that were repeated more often (reiteration effect).

Our first hypothesis stated that the amount of reminiscence and hypermnesia would increase more over the successive sessions in the Lab-discussion condition as compared to the Retell condition. The reason for this hypothesis was that we expected that the questions posed by the discussion partner would cause the participants to recall more new information in the Lab-discussion condition as compared to the Retell condition.

The results for the main effects for the sessions showed that there was reminiscence and hypermnesia in both conditions in session 6. In both conditions the number of correct items and incorrect items regarding both Forensically central and peripheral information were higher in the testing session as compared with the five experimental sessions. The interaction effects showed that both correct and incorrect items were affected differently by the conditions over the sessions. The results also showed that participants in the Lab-discussion condition reported more correct and incorrect items in the sessions than the participants in the Retell condition. The results also showed that the participants in both the conditions recalled new items in each session. Amount of new information was higher in earlier sessions and in the testing session. These results supported our hypothesis that the Lab-discussion condition would show higher reminiscence and hypermnesia than the Retell condition. These results also support 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 (Henkel, 2004; Mulligan, 2006; Payne, 1987).

Further, the interaction effect for incorrect items between session and condition showed that the increase in the number of incorrect items was higher in the Lab-discussion condition than the Retell condition. One reason for this may have been that the participants when replying to the questions asked by the confederates, were making more retrieval attempts. These results also showed that the number of incorrect items increased with more recall occasions (Henkel, 2004). There was also an interaction effect between information type and condition for incorrect items. This showed that participants in the Lab-discussion condition recalled more incorrect Forensically peripheral items than did the participants in the Retell condition. These results are in line with the previous findings that peripheral information is more vulnerable to misinformation as compared to central information in the face of focused questions (Candel, Merckelbach, Jelicic, Limpens, & Widdershoven, 2004; Roebers & Schneider, 2000). This may be because of the fact, discussed in the intro-

duction, that forensically peripheral information can be assumed to be less integrated than the forensically central information that would make the forensically peripheral information more vulnerable to errors.

An implication of these findings is that discussions of the forensically peripheral information from the experienced event are likely to have negative consequences for eyewitness memory in that the eyewitnesses are more likely to provide a distorted description of persons and objects involved in the event than they are to provide a distorted description of action events.

The second hypothesis proposed that the information that is more often discussed or retold in the experimental sessions is more likely to be reported in the final test probably because of the testing effect (Roediger III & Karpicke, 2006b). The results partially supported our hypothesis and showed that the Forensically central information that was retold or discussed in three to four sessions was reported at the testing session significantly more than the Forensically central information that was retold or discussed in one to two experimental sessions. In contrast, for the Forensically peripheral information that was reported in session 6 there was no difference as an effect of if this information was repeated more or fewer times during the five experimental sessions. In addition, we did find that information that was mentioned in the later sessions was more likely to be reported in the testing session than items that was mentioned in the earlier sessions. The effect was rather weak, and when splitting up the information into Forensically central and Forensically peripheral information the effect no longer reached conventional levels of significance. The results indicate that both for Forensically central and peripheral information the *number of retellings* may be more important for the information to be reported at the testing session than if the information was repeated earlier or later in the previous sessions. An implication of this result in forensic situations is that eyewitnesses are likely to report such information to the crime investigators or in the court that was repeated more in the previous interactions with different people. Eyewitnesses may easily miss to report valuable information that happened not to be retold often. The analysis for only correct and only incorrect information separately (though not reported in the results section) showed the same trend as well. From this at least it can be concluded that it is not the correctness of information that mediate this effect. It is not clear what factors caused our participants to report some information and miss other information in the initial discussions or retellings but the items that are most often repeated are likely to be reported more in the future recall occasions.

The third hypothesis suggested that for Forensically peripheral information the confidence level would be higher for information that was repeated more as compared with the confidence level for the information that was repeated less (with accuracy controlled), that is, the hypothesis predicted the presence of a reiteration effect for the Forensically peripheral information (Hertwig et al., 1997). An explanation could be that more frequent retrieval of items creates a stronger feeling of retrieval fluency. We also expected that there would be no reiteration effect for forensically central information. The interaction effect for the information type and condition supported our hypothesis. This interaction effect showed that the confidence judgments increased over sessions for the reoccurring forensically peripheral information while the accuracy level (obviously) remained the same whereas this was not the case for the forensically central information where the results showed no difference in accuracy and confidence.

The suggested reason, as discussed earlier, is that the forensically peripheral information is not as well integrated with the other information and that there is thus more room for increase in retrieval fluency with additional recalls for the peripheral as compared with the central information (Shaw & McClure, 1996). The forensically central information can be assumed to be more well integrated and the results for this type of information, showing a lack of reiteration effect, may help to explain why confidence was found to be a better predictor for the accuracy of forensically central details.

One may think that retrieval of the information in later sessions would affect the confidence judgments because that information would probably be fresh in memory and easily available. However, the results showed that earlier or later retrieval of the information was not related to the confidence level. To conclude, this study has shown that when making a credibility evaluation of recalled information the nature of content should be considered because it may, with many recalls, affect the persons ability to monitor their meta-memory judgments.

References

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

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., & 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.

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

Bornstein, B. H., Liebel, L. M., & Scarberry, N. C. (1998). Repeated testing in eyewitness memory: a means to improve recall of a negative emotional event. *Applied Cognitive Psychology*, 12(2), 119-131.

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

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.

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.

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.

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.

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.

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

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.

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.

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. (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., Mather, M., Villa, D., & Morita, S. K. (2001). How events are reviewed matters: Effects of varied focus on eyewitness suggestibility. *Memory & Cognition*, 29, 940-947.

Leippe, M. R., & Eisenstadt, D. (2007). Eyewitness confidence and the confidence-accuracy relationship in memory for people. In R. C. Lindsay, D. F. Ross, J. D. Read & M. P. Toglia (Eds.), *Handbook of Eyewitness Psychology Vol. 2 Memory for people.*(pp. 377-425). Mahwah, NJ: Lawrence Erlbaum Associates.

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

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

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

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.

Ost, J., Granhag, P.-A., Udell, J., & Hjelmsäter, E. R. a. (2008). Familiarity breeds distortion: The effects of media exposure on false reports concerning media coverage of the terrorist attacks in London on 7 July 2005. *Memory*, 16(1), 76-85.

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.

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.

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 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.

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

Sarwar, F., Allwood, C. M., & Innes-ker, Å. (2010). Effects of repeated recall and discussions on eyewitness' accuracy and meta-memory realism for different types of forensic information. Unpublished Manuscript. Department of Psychology, Lund University, Sweden.

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

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. Memory for events.*(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.

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.

Wells, G. L., & Bradfield, A. L. (1999). Measuring the goodness of lineups: parameter estimation, question effects, and limits to the mock witness paradigm. *Applied Cognitive Psychology*, 13(S1), S27-S39.

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.