

Counting on the Details

Inquiring Into Past Events of Cooperative Interviewees

Counting on the Details

Inquiring Into Past Events of Cooperative Interviewees

Rebecca M. Willén

Doctoral Dissertation in Psychology
Department of Psychology
University of Gothenburg
(2016)

© Rebecca M. Willén
Cover layout: Ann-Sofie Sten
Editing: Kat Singer
Printing: Ineko, Gothenburg, Sweden, 2016
ISBN: 978-91-628-9864-9 (PDF)
ISBN: 978-91-628-9865-6 (Print)
ISSN: 1101-718X Avhandling/Göteborgs universitet, Psykologiska inst.
ISRN: GU/PSYK/AVH--341--SE
<http://hdl.handle.net/2077/43236>

You've got time

*Think of all the roads
Think of all their crossings*

*Everything is different
The second time around*

-Regina Spektor

Abstract

Willén, R. M. (2016). *Counting on the Details: Inquiring Into Past Events of Cooperative Interviewees*. Department of Psychology, University of Gothenburg, Sweden.

Amount and quality of detail in recollections of past events are often studied in legal psychology. What, how much, and how accurate does a witness typically recall? How can we facilitate witnesses' recollections in police interviews? How can we detect deception? The overall aim of this thesis was to employ research with high ecological validity to investigate amount and quality of detail in interviews with cooperative adults. Study I investigated details in true and false confessions by 30 cooperative offenders in a within-subject experiment. The confessions concerned one crime the respondent had conducted and been sentenced for (true confession) and one crime the respondent had never conducted nor been sentenced for (false confession). Studies II and III investigated details in interviews with 95 cooperative adults about repeated dental visits. Study II was an experiment employing a within-subject design to investigate the effect of context-specific cues on recollection of repeated events. Study III further studied the data by investigating how five factors (interviewee age, rehearsal, interviewer, number of experienced events, and unpleasantness) affected two different measures of amount and quality of detail. Study IV made all data from Studies II and III, supplemented with a codebook, freely available in an online repository. Surprisingly, Study I showed very few differences in amount and quality of detail between offenders' true and false confessions. However, three detail measures could distinguish the true and false statements, which is promising for future research on offenders' statements. Results from Study II suggested that context-specific cues may generate more details about repeated events than cues commonly used in police interviews. Thus, mnemonics such as context-specific cues may, in the future, be a positive addition to current mnemonic techniques employed in legal practice. Study III showed that the two different detail measures were affected differently by all five factors (e.g., who conducted the interview did have an effect on one of the two measures but had no effect on the second measure). The results highlight the need for standardization of how we measure amount and quality of detail in research on investigative interviewing. The data from Studies II and III can be reused as data from an experiment (including both interviews, as in Study II) or as single interview data (including data only from Interview I, as in Study III). The data and material can be used for research and educational purposes. This thesis contributes to the ongoing methodological revolution in psychology by practicing transparent reporting, publishing the data for Studies II and III, and by raising the problem of using non-standardized procedures and measures in research on investigative interviewing.

Keywords: Investigative interviewing; Amount of detail; Autobiographical and episodic memory; Open data; Open material; Repeated events; Statement analysis

ISBN: 978-91-628-9864-9 (PDF); ISBN: 978-91-628-9865-6 (Print); ISSN: 1101-718X
Avhandling/Göteborgs universitet, Psykologiska inst.; ISRN: GU/PSYK/AVH--341--SE;
<http://hdl.handle.net/2077/43236>

Rebecca M. Willén, Department of Psychology, University of Gothenburg, P. O. Box 500,
405 30 Gothenburg, Sweden. Phone: +46-31-3308005. Email: contact@rmwillen.info

Swedish Summary

Detaljrikiedom är ett begrepp av central betydelse i många rättsliga utredningar. I polisförhör och domstolsutredningar rörande sexuella övergrepp eller våld i hemmet är detaljrikiedom i målsägarens berättelse något man ofta bedömer och återkommer till under utredningens gång. Även i asylärenden görs en bedömning av detaljrikiedom i den asylsökandes berättelse. Det är därför naturligt att detaljrikiedom också har en central betydelse inom rättspsykologisk forskning där man till exempel studerar hur väl olika förhörstekniker och tillförlitlighetsbedömningar fungerar. Det kanske främsta måttet på hur väl en förhörsteknik fungerar är nämligen hur detaljrik (och korrekt) utsaga den förhörde ger. Enligt rättspsykologisk forskning är det vanligt att falska utsagor (dvs, en lögn) kännetecknas av brist på detaljrikiedom. I linje med detta är det vanligt i domstolsutredningar att en målsägares utsaga inte godkänns som bevis om den anses vara alltför detaljfattig.

Avhandlingens övergripande syfte

Denna doktorsavhandling består av tre empiriska arbeten och en artikel där all data från två av de empiriska studierna gjordes offentligt tillgängliga. Avhandlingens övergripande syfte var att, med så realistiska undersökningsmetoder som möjligt, studera detaljrikiedom i minnesrapporter från intervjupersoner som samarbetar med intervjuaren för att ge en så detaljerad utsaga som möjligt.

Metoder och resultat

I denna avhandling studerades detaljrikiedom i totalt 125 individers utsagor. I Studie I intervjuades 30 personer intagna vid svenska kriminalvårdsanstalter om brottsliga händelser så som misshandel eller stöld. Varje deltagare intervjuades två gånger med en kort paus mellan de två intervjuerna. I en intervju gav deltagaren ett sant erkännande av ett brott hen hade begått och dömts för, och i en annan intervju gav deltagaren ett falskt erkännande av ett brott hen aldrig hade begått eller dömts för. Vi studerade därefter detaljrikiedom i dessa erkännanden genom att tillämpa två olika kriteriebaserade verktyg: Criteria-Based Content Analysis (CBCA) och Reality Monitoring (RM). I linje med tidigare rättspsykologisk forskning predicerades det att de sanna erkännandena skulle vara mer detaljerade än de falska erkännandena.

Resultaten från studien visade dock på mycket få skillnader i detaljrikedom mellan de sanna och falska erkännandena. Det gick helt enkelt inte att använda CBCA eller RM för att skilja de sanna utsagorna från de påhittade. CBCA och RM består av sammanlagt 31 kriterier som på mer eller mindre olika sätt kan sägas mäta detaljrikedom. Endast 3 av dessa kriterier var användbara för att skilja de sanna utsagorna från de påhittade. De sanna utsagorna innehöll oftare beskrivningar av oväntade komplikationer (t.ex., *”när jag körde ut från gården hoppade en höna plötsligt ut framför bilen, så jag var tvungen att gå ut och lyfta undan den innan jag kunde fortsätta”*). De påhittade utsagorna däremot innehöll (oftare än de sanna) tvivel rörande den egna utsagan (t.ex., *”nu när jag hör mig själv berätta det här för dig så låter det helt otroligt att det verkligen skulle ha gått till på det här viset...”*), och nedvärderande eller ogillande uttalanden rörande den egna personen kopplat till händelsen (t.ex., *”det var så hemskt av mig att göra så mot henne”*). Resultaten från Studie I visar att sanna och falska erkännanden från personer med kriminell erfarenhet kan vara mycket svåra – men inte omöjliga – att särskilja.

I Studierna II och III intervjuades 95 personer om de tandvårdsbesök de hade gjort under den senaste 10-årsperioden. I Studie II intervjuades deltagarna vid två tillfällen med en kort paus emellan. I första intervjun ombads de försöka minnas så mycket som möjligt och berätta om detta. I andra intervjun ombads deltagarna berätta om den ytterligare information de eventuellt kommit ihåg under pausen. I pausen fick deltagarna ta del av sådant som eventuellt kan vara till hjälp för att minnas tandvårdsbesöken. Det fanns tre olika typer av sådant potentiellt minnesunderlättande material och varje deltagare fick ta del av en av dem: (1) Sådant som andra personer i samma situation sagt sig ha haft hjälp av för att minnas sina tandvårdsbesök. Ca 30 deltagare fick ta del av detta som enbart kategorier, t.ex. *”Tänk tillbaka på sådant du pratat om med personalen”* och *”Tänk tillbaka på orsaken till besöket”*. (2) En annan grupp av lika många deltagare fick ta del av kategorierna tillsammans med citat från personer som varit i samma situation (t.ex. *”Tänk tillbaka på sådant du pratat om med personalen”* och fem tillhörande citat, t.ex. *”Jag fick instruktioner om hur jag skulle förebygga hål”* och *”Jag lovade tandläkaren att sluta snusa”*). (3) En tredje grupp fick ta del av sådant som ofta används i polisintervjuer idag i syftet att underlätta för målsägaren att minnas mer. T.ex. *”Tänk tillbaka på om något tillfälle skiljer sig från de övriga”*, och *”Tänk tillbaka på senaste gången det hände”*. Denna typ av instruktioner gavs till en tredje grupp om ca 30 deltagare. I linje med tidigare forskning på området förväntade vi oss att de två förstnämnda materialen skulle generera mer detaljrikedom än sådant material som idag används i polisförhör.

I Studierna II och III mättes detaljrikedom på två sätt. Vi räknade antalet enskilda tandvårdsbesök som deltagaren lyckades nämna minst någon slags detalj om, och med vilken specificitet deltagaren beskrev sina

tandvårdsbesök. Specificiteten mätte vi genom att kategorisera allt deltagaren sa som antingen specifikt, specifikt-förlängt, eller generellt. Denna kategorisering är vanlig inom klinisk psykologisk forskning om autobiografiskt minne och mäter hur detaljrikt personen beskriver ett visst minne.

Resultaten från Studie II visade små statistiska skillnader i detaljrikedom mellan de tre grupperna, men i linje med prediktionen generade materialet baserat på andra tandvårdspatienters erfarenheter något högre detaljrikedom än materialet som idag ofta används i polisförhör. Resultaten innebär att förhörstekniker möjligen på sikt kunde vinna på att integrera någon typ av frågor som baseras på andra individers erfarenheter av liknande situationer.

I Studie III undersökte vi på ett explorativt sätt vad deltagarna berättade i sin första intervju (andra intervjun ingick således enbart i Studie II, inte i Studie III). Mer specifikt studerade vi hur fem olika faktorer påverkade de två måtten på detaljrikedom (dvs, antalet enskilda besök deltagaren nämnde och mängden specificitet i berättelsen). De fem faktorer vi undersökte var deltagarens ålder, hur många tandvårdsbesök personen hade gjort under de senaste 10 åren, hur mycket deltagaren hade tänkt på eller pratat om tandvårdsbesöken efter det att de ägde rum, hur obehagligt deltagaren fann tandvårdsbesök vara, och vem som genomförde intervjun. Resultaten visade att de två olika måtten på detaljrikedom aldrig påverkades på samma sätt av de fem faktorer vi testade. Till exempel ökade mängden specificitet om en viss intervjuare skötte intervjun, men antalet enskilda besök deltagaren nämnde påverkades inte av vem som genomförde intervjun. Ett annat exempel är att det inte spelade någon roll för mängden specificitet hur obehagligt deltagaren fann tandvårdsbesök vara, men däremot kom deltagarna ofta ihåg fler enskilda besök om de överlag tyckte att tandvårdsbesök är mycket obehagliga. Resultaten från Studie III visar på problemen med att det ännu saknas ett standardiserat sätt att mäta detaljrikedom på. Detta gör resultat från olika studier svåra att jämföra med varandra och försvagar forskningsfältets vetenskaplighet.

Datan från Studierna II och III publicerades i ett offentligt digitalt arkiv och beskrevs i Studie IV. Således är all data från dessa två studier fritt tillgängligt för vem som helst att återanvända. Även allt forskningsmaterial som utgjorde grund för Studierna II och III finns offentligt tillgängligt online. Datan kan återanvändas som data från ett experiment (vilket skulle inkludera datan från båda intervjuerna, som i Studie II) eller som vanlig intervjudata (vilket skulle inkludera data enbart från den första intervjun, som i Studie III). Både data och material kan återanvändas och modifieras för att användas i forskning eller utbildning.

Slutsatser

Denna doktorsavhandling bidrar till tidigare rättspsykologisk forskning om utredande förhör på minst tre sätt. Först och främst har avhandlingens empiriska studier alla en hög nivå av realism (s.k., ekologisk validitet) i sin utformning. Detta är relativt ovanligt i rättspsykologisk forskning om utredande förhör där detaljriktighet studeras. För det andra har såväl data som forskningsmaterial publicerats offentligt vilket gör det möjligt för andra forskare att bygga vidare på arbetet men också möjliggör kontroller av arbetets kvalitet. För det tredje lyfter denna avhandling upp problemet med bristen på standardisering av detaljriktighet inom rättspsykologin. Utan standardisering av ett såpass centralt begrepp som detaljriktighet hämmas fältet i sin kunskapsutveckling och vi rättspsykologer riskerar ge felaktiga rekommendationer till rättspraktiker, vilket i sin tur kan öka risken för (objektivt sett) felaktiga beslut i rättsliga utredningar.

Sedan 2011 pågår en omfattande metodologisk revolution inom psykologin och andra empiriska vetenskaper. Psykologin går med hastiga och beslutsamma steg mot ökad vetenskaplig transparens. Allt fler tidskrifter, finansierare och enskilda forskare kräver mer transparent (i det här fallet mer ärlig och fullständig) rapportering av metod och resultat, preregistrering av studierna innan dess att datainsamlingen påbörjats, och öppen och fri publicering av såväl forskningsartiklar som av data och material. Denna avhandling bidrar till den pågående revolutionen genom att praktisera transparent och ärlig rapportering och genom att data och material gjorts fritt tillgängligt i öppna digitala arkiv.

Preface

This thesis consists of the following four studies (referred to by their Roman numerals):

- I. Willén, R. M., & Strömwall, L. A. (2012). Offenders' uncoerced false confessions: A new application of statement analysis? *Legal and Criminological Psychology*, *17*, 346–359. <http://dx.doi.org/10.1111/j.2044-8333.2011.02018.x>
- II. Willén, R. M., Granhag, P. A., Strömwall, L. A., & Fisher, R. P. (2015). Facilitating particularization of repeated similar events with context-specific cues. *Scandinavian Journal of Psychology*, *56*, 28–37. <http://dx.doi.org/10.1111/sjop.12180>
- III. Willén, R. M., Granhag, P. A., & Strömwall, L. A. (2016). *Factors Affecting Two Types of Memory Specificity: Particularization of Episodes and Details*. Manuscript submitted.
- IV. Willén, R. M., & Granhag, P. A. (2015). Data from interviews with 95 respondents recollecting repeated dental visits. *Journal of Open Psychology Data*, *3*: e7. <http://dx.doi.org/10.5334/jopd.an>

Contents

Contents.....	i
Acknowledgements.....	iii
Introduction.....	1
An Introduction to Human Memory.....	4
Established Principles of Memory	5
Memory for Traumatic Events	8
Memory for Repeated Events	9
Memory for Events Questionable in Veracity	12
Eliciting Details.....	14
Investigative Interviewing	14
Specific Mnemonics	15
Memory Training	17
Analyzing Details.....	19
Statement Analysis	19
Autobiographical Memory Specificity	21
Counting the Details	21
Summary of the Included Studies.....	25
Disclosure Statement.....	27
Disclosure Statement for Study I	27
Disclosure Statement for Study II	28
Disclosure Statement for Studies III and IV	28
Study I.....	29
Aims	29
Methods	29
Results	30
Conclusions	30
General Methods for Studies II, III, and IV.....	32
Participants	32
Deriving the Memory Cues for Study II	32
The Interviews	35
Post-interview Procedure	35
Data Preparation	35
Study II.....	37
Aims	37
Methods	37
Results	37
Conclusions	38
Study III.....	39
Aims	39
Methods	39
Results	39
Conclusions	40
Study IV.....	41

Aims	41
Content	41
Reuse potential	41
General discussion	43
Eliciting Details from Cooperative Interviewees	43
Analyzing Details from Cooperative Interviewees	46
Research Designs for the Future	48
Limitations	49
Conclusions	51
References	53
Appendix	63
<i>Table 1</i>	7
<i>Table 2</i>	20
<i>Table 3</i>	34
<i>Table 4</i>	40

Acknowledgments

My main supervisor *Professor Pär Anders Granhag* trusted me enough to let me work independently, even in the geographical sense. He thereby gave me the freedom I needed to thrive both professionally and personally. For this I am and always will be sincerely grateful.

My second supervisor *Professor Leif Strömwall* inspired me to enter the PhD program and taught me the importance of handling the statistics and methods well. I am truly thankful for this. Almost as important, Leif also taught me to sigh, shrug, and move on when things just wouldn't work out.

My colleagues and peers in the research group for *Criminal, Legal, and Investigative Psychology (CLIP)* at the University of Gothenburg, Sweden, and in the *Psychology and Law Lab* at the Memorial University of Newfoundland, Canada, have all been invaluable resources of knowledge and support. I especially want to mention *Dr. Malin Hildebrand Karlén*, *Dr. Emma Roos af Hjelmsäter*, *Kirk Luther*, *Chris Lively*, and *Zak Keeping*. A special thanks to *Professor Brent Snook* who graciously hosted me in his lab and was very helpful with both practical and academic matters.

My present and future colleagues devoted to improving the empirical sciences have been very important to me during my last two years as a PhD student. I especially want to mention *Dr. Gustav Nilsson* at the Karolinska Institute and Stockholm University, *Professor Olle Häggström* at the Chalmers University of Technology, *Professor Gert Helgesson* at the Karolinska Institute, *Fil. Lic. Gró Einarsdóttir* and *Fil. Lic. Bodil Karlsson* at the University of Gothenburg.

Thanks to all the fantastic people in our department's administration, especially *Ann Backlund*, *Christina Wanner*, *Petra Löfgren*, *Ann Witte*, and *Ann-Sofie Sten*. Thank you also to *Kat Singer*, for an excellent proofreading of this thesis.

Thank you to everyone who made the studies included in this thesis possible. I especially want to thank *the Public Dental Care in the Region of Västra Götaland, District South*, and the director *Anders Ljungné*. I also want to thank all research assistants who performed a tremendous job with transcribing, coding, categorizing, and interviewing. I especially want to mention *David Lopes* who assisted me across all four studies included in the thesis; *Andreas Aspholmer* who was always very reliable and efficient and generated high-quality work; and *Thérèse Eriksson* whom I, without a doubt, would trust to finish off my thesis and defense (with better results than if I did it myself) in the case I would not be able to make it.

My family and friends give me the love, happiness and cries I need to thrive. First and foremost, I want to thank my daughter, *Tova Lova Sara-Li*, as well as *Daniel*, my dear co-parent and always my very best friend. I am extremely grateful and proud to have you both in my life – *I am because we are*. My mum, *Lena Bertilsson*, and her husband, *Johan Bertilsson*: I thank you both for being such important parts of my life. My dear friends, *Johan Rudolfsson*, *Anna Hedenus*, *Amanda Duregård*, *Karolina Lindén*, *Pauline Nylin*, *Frances Sprei*, and *Erika Näsholm* – for open hearts, open minds, and open-hearted talks that mind. Thanks also to *Carl Jakobsson* who kindly picked up my McNally book and a new computer in Sweden and brought them to me in Indonesia where I finished this thesis. And thank you, *Cuk with family*, for inspiring me to open new windows in my life.

I want to end by thanking two people who gave me the inspiration I needed to significantly change my life. *Ewa Asplund*, for your endless patience and strong conviction that change is always possible. I cannot grasp who and where I would have been if it was not for you – thank you always. And my very beloved dad, *Ingemar Willén*. While I slowly won my battle, you just as slowly lost yours. I know you would have been here if you could, proud, happy, and filled with enthusiasm. I do thank you with all my heart, for I am because you were.

This research was financially supported by *the Swedish Crime Victim Compensation and Support Authority* (Grant No: 252107824).

Introduction

Think back to yesterday and try to recall the day. How and when did it begin? How did it proceed? What happened at which time? How and when did it end?

Calling particular events into mind, such as recalling the yesterday, may bear some resemblance to projecting a video or a series of still photographs to the inner mind (although our memory is far from a recording of the events). However, while some parts may come to mind easily and without much deliberate effort, other parts will remain blank regardless of how much we try. Yet other parts of the events may be recollected only during the right circumstances, such as the right questions being asked in the right way.

Telling in words to someone about a memory of events would mean putting into words a complex collection of stimuli. The words may be descriptions of something seen, heard, tasted, smelled, sensed, felt, or thought. The descriptions can be more or less detailed because of factors such as personality, cognition, motivation, social and communicative dynamics, or environmental and physical conditions. Some people really like to talk, others prefer not to. Some people really like listening to what others have to say, others do not. Sometimes you are just too tired to concentrate or not really motivated to tell the story once again. Other times you might find it difficult describing events in detail simply because you misremember or, for any (good or bad) reason, are making things up.

Variations in amount and quality of detail provided when recollecting past events are studied by psychological researchers for different reasons. For instance, in clinical settings, individuals suffering from depression or post-traumatic stress may need memory training to increase the chances of rehabilitation; increased recollection of details can moderate symptoms of depression. It can also be important in health settings to know in what detail – and how accurately – people usually recall past events, for instance when inquiring about medical history.

Variations in amount and quality of detail have also been frequently investigated within legal and forensic psychology. Several research questions within this area have often been studied by measuring variations in the details. What, how much, and how accurate does a witness typically remember and report? How can we facilitate witnesses' recollection during police interviews? How can we detect deception? The scope of the current thesis covers all three questions with an overall focus on amount and quality of detail reported by cooperative interviewees. It consists of three empirical

studies and one data article (hereafter, all four papers will be referred to as “studies”). Study I investigated details in true and false confessions by cooperative offenders. Studies II, III, and IV regarded details in interviews with cooperative individuals about repeated events.

The initial aim of this thesis was to conduct a series of studies on memory for repeated events and how recollection of such memories can be facilitated in police interviews. However, the data collection for Studies II, III, and IV turned out to be more time consuming than expected. The scope of the thesis was therefore adjusted to include Study I.

The main post hoc aim of the present thesis was to employ research designs with high ecological validity to investigate the amount and quality of detail in interviews with cooperative interviewees. Thus, the four included studies share the following three features: (1) variations in amount and quality of detail were the primary outcome measures; (2) the interviews were made with cooperative individuals (i.e. interviewees willing to tell about the events); (3) the experienced events took place under natural circumstances and thus were not planned, induced, or controlled by any researcher.

In light of the current methodological revolution in psychological science (De Groot, 1956/2014), the shift in focus from memory for repeated events to the amount and quality of detail in interviews with cooperative interviewees generated new and interesting perspectives. Perspectives interesting not only for this thesis but also for future research on investigative interviewing.

How do researchers measure the amount and quality of detail? What does it mean for overall conclusions that amount and quality are operationalized differently in different studies? Do the many alternative measures (of details) available to researchers inflate the risk of questionable research practices being employed? One example is primary outcome measures being swifited, e.g., new codings being made if the first did not result in statistically detectable results (Simmons, Nelson, & Simonsohn, 2011). And consequently, how can we improve our measures of amount and quality of detail? Issues like these are in the core of science because wonky measures will generate wonky research findings. In turn, in the field of investigative interviewing, wonky research findings could inspire ineffective or even harmful legal practices. The importance of dealing with how we measure details can thus not be underestimated. Therefore, a secondary aim of the thesis was to briefly introduce and discuss issues related to counting details in the research of investigative interviewing.

The present thesis begins with a short introduction to human memory, giving special attention to three themes of core relevance to the present thesis: memory for traumatic events, memory for repeated events, and

memory for events questionable in veracity (specifically deceptive reports and false confessions). The literature on how to successfully elicit details from cooperative individuals will be summarized, including an outline of how the details are typically measured and analyzed. These introductions will be followed by summaries of the four studies. Finally, the observed results and issues are discussed and suggestions are made for future research.

An Introduction to Human Memory

There is a large body of knowledge concerning the human memory function, although there is still a lot we cannot yet explain. One thing we know little about is the actual basics of memory: exactly where it is and how we can model it. The two most frequently employed psychological perspectives of memory are the multiple systems view (e.g. Tulving, 1983) and the process view (e.g. Kolers & Roediger, 1984). While the first view models memory as consisting of several different systems, for example procedural, semantic, working, and episodic memories of which each has its own unique capacities, qualities and prerequisites, the processing view argues that memory capacity and quality is rather a result of encoding and retrieval processes – not of different structures. Both views have strengths and weaknesses. Surprenant and Neath (2009) summarize the current knowledge gaps as follows (p. 25):

“We do not know how many memory systems there are or how to define what a memory system is. We do not know how many processes (or components of processing) there are or how to distinguish them. [S]hort-term memory and long-term memory seem to differ in some ways, as do episodic memory and semantic memory, but are they really fundamentally different?”

Although consensus is lacking on the number of memory systems and how to define them, there seems to be some consensus on what we generally mean when we talk about the major systems. Major memory systems central to the present thesis are briefly described below.

Autobiographical memory. Baddeley, Eysenck, and Anderson (2009) described this as a system for information regarding ourselves and relations and events that we experience during our lifetime. It can be episodic as well as semantic (see below; Baddeley et al., 2009), although some view it as mainly episodic (Fivush, 2011).

Episodic and semantic memory. According to Tulving (2002), episodic memory enables us to relive our past and make plans for the future, and it makes it possible for us to have concepts about time (e.g. how long it will take until we meet our friend again). Episodic memory is believed to have developed out of semantic memory. Semantic memory refers to general knowledge of the world in absence of time. For example, we know that the capital of Germany is Berlin, that there once was a group of animals called dinosaurs, and that fire can be dangerous. It is, however, unlikely that we will recall how or when we learned this. In contrast, episodic memory refers to events that we have experienced, such as a holiday spent in Berlin, a visit to a

museum where we saw models of dinosaurs, or a time when we burned ourselves on a match. Episodic and semantic information is processed in both short-term and long-term memory.

Short- and long-term memory. Long-term memory is an uncontroversial system which researchers agree exist; individuals would otherwise be unable to hold memories and knowledge derived from the past (Cowan, 2008). In contrast, short-term memory and working memory are disputed and there is little agreement on how to define them, their functions, or whether or not they exist (Cowan, 2008; Neath & Surprenant, 2008). In spite of this, short-term memory and working memory are frequently referred to as established terms whenever cognitive performance is an issue.

The main scope of the present thesis is verbal reports of past events. More specifically, recollections of autobiographical episodic memories from long-term memory. It should however be noted that retrieval from long-term memory prerequisites activation of short-term memory. The short-term memory is continuously refilled with information stored in long-term memory (Cowan, 2000). Furthermore, short-term memory is needed for tasks such as planning, thinking, controlling the language and choosing retrieval cues (Shiffrin, 1993). Disruptions in short-term memory can therefore have a negative impact on our recollection and reporting of experienced – and non-experienced – events. Thus, capacity and functions of short-term (and working) memory may also be relevant for retrieval of autobiographical and episodic memories from long-term memory.

Established Principles of Memory

A different way to introduce human memory is to focus on what we do know about its function. Surprenant and Neath (2009) presented seven principles (see Table 1) that describe “*fundamental empirical regularities*” (p. 25) in memory that today are supported by extensive evidence and are valid regardless of how we choose to model or explain memory. The purpose of listing established principles was to spur theory development concerning the basics of memory, and the authors emphasize that the list is not exclusive and may not include the most important principles. However, each of their seven principles is fundamental to the present thesis and thereby constitutes a meaningful introduction to human memory function.

First, whenever we remember something it will always start with a cue (Tulving, 1974). The cue can, for example, be a question, a thought, a visual stimulus, an emotion, a smell, or a taste. There is always something that

triggers remembering and that something is called a cue (Surprenant & Neath, 2009).

Second, the encoding-retrieval principle (e.g., Tulving & Thomson, 1973) states that whether or not you will recall something depends on the match between conditions during encoding (i.e., the point in time where you experience the to-be-remembered stimuli) and retrieval (i.e., the point in time where you try to recall the experienced stimuli). No or weak association between the two conditions will have little potential to trigger remembering. For example, it is unlikely that pictures of vegetables will help you recall cities in Asia. According to Surprenant and Neath (2009), a consequence of this principle is that items, processes, and cues do not have an intrinsic, mnemonic value. That is, whether or not a retrieval strategy actually facilitates recollection depends on the encoding conditions and the match between retrieval strategy and encoding conditions.

Third, cues will be more effective if they do not have too many associations (Watkins & Watkins, 1975). The word “Swede” may be an effective cue for someone outside of Sweden to recall a certain person, but it will likely be rather ineffective for most people living in Sweden.

Fourth, memory is reconstructive (e.g., Bartlett, 1932). It is created through a dynamic process that does not end with the to-be-remembered stimuli. We reshape our memory to better fit our expectations, beliefs and understandings of the world around us. This is central whenever discussing memory and recollection. Every time we call a particular event into mind, the memory for this particular event is at risk of being somewhat altered (reconstructed). The next time we recall it, it is the altered version we recall which again is at risk of being somewhat altered. Thus, as the next principle states, memory is impure and far from a recording of the objectively true events.

Fifth, the impurity principle (Surprenant & Neath, 2009) is a consequence of the reconstruction principle (i.e., the fourth principle) and states that it is not possible for researchers to measure a particular memory system or process. Memory is not pure, and it is therefore impossible for outcomes to be pure whatever measure the researcher uses.

Sixth, we will better recollect items that, for any reason at the time of retrieval, stand out from the others (e.g., Fisher, 1981; Surprenant & Neath, 2009). An emergent and painful dental visit will likely stand out and therefore be recollected (as long as there were no additional such visits). However, even if the item did not stand out during encoding it might do so during certain retrieval conditions. A bundle of keys on the kitchen table may be encoded without having any particular meaning to you until you reach the

office and realize that you cannot get in. The locked door becomes the cue that makes the memory stand out – the keys are on the kitchen table.

Seventh, the specificity principle states that specific information is more vulnerable to memory errors than general information (Surprenant & Neath, 2009). The specific information may involve details about events or items, such as colors, forms, names, dates, or single episodes in a series of events. In contrast, general information regards the overall picture or typical features, for example how something usually looks or how a certain type of event usually occurs.

From the seven principles of memory (Surprenant & Neath, 2009) we can draw at least three important conclusions. One, cues are needed for recollection – but not any cue will do (Principles 1–3). Two, the recollections will be more or less distorted and it is therefore not possible for a researcher to measure memory in exact ways (Principles 4–5). Three, some details are easier to remember than others, but overall, details are vulnerable to memory errors (Principles 6–7).

Table 1. Seven principles of memory.

1	The cue-driven principle	Recollection will always begin with a cue.
2	The encoding-retrieval principle	Memory depends on the match between conditions at encoding and retrieval.
3	The cue-overload principle	Cues are less effective when they have too many associations.
4	The reconstruction principle	Memory is the result of many things – not only by information present at the time of encoding.
5	The impurity principle	It is impossible to measure a particular system or process – memory is always impure.
6	The relative distinctiveness principle	Items that stand out from the others will be well recalled.
7	The specificity principle	Specific information is more vulnerable than general information.

Notes. Table adapted from Surprenant and Neath (2009; p. 7–8).

Memory for Traumatic Events

Memory for traumatic events has been a disputed topic ever since the 1980s when repressed memories of sexual abuse were suggested to be the cause of diverse symptoms of psychological distress such as insomnia, nightmares, and anxiety. While some claim that memories for traumatic events are often less detailed than emotionally neutral events – sometimes even repressed or forgotten (see e.g., Alpert, Brown, & Courtois, 1998a) – others claim memories of traumatic events to be even more detailed than memories of emotionally neutral events (e.g., Ornstein, Ceci, & Loftus, 1998). In general, the first standpoint has been represented by clinical psychologists and psychotherapists while the latter standpoint has been represented by experimental psychologists (McNally, 2005).

Experimental psychologists prefer to give more weight to evidence from empirical research (Ornstein, Ceci, & Loftus, 1998) while clinical psychologists tend to give more weight to case studies and experiences from clinical practice (Alpert, Brown, & Courtois, 1998b). Relying on evidence from empirical research, the general conclusions are usually that experienced events that were strongly, emotionally charged are remembered well (for exceptions and a different perspective, see e.g., Deffenbacher, Bornstein, Penrod, & McGorty, 2004), although details that were peripheral to the rememberer during encoding might not be recalled (Christiansson, 1992; McNally, 2005; Ornstein, Ceci, & Loftus, 1998).

Traumatic events evoke post-traumatic stress disorder (PTSD) in only a minority of individuals exposed to trauma (McNally, 2005). PTSD is a psychiatric diagnosis characterized by, for example, re-experiences of the traumatic event, avoidance of stimuli that somehow reminds of the trauma, and emotional arousal (e.g., sleep disturbances or self-destructive behavior; American Psychiatric Association, 2013). Neuro-imaging studies typically show differences between individuals diagnosed with PTSD and trauma-exposed individuals without PTSD in brain activation when presented with trauma-related stimuli (Sartory et al., 2013). It is thus important to distinguish between individuals exposed to trauma and individuals diagnosed with PTSD. Exposure to trauma is a prerequisite for the PTSD-diagnosis according to the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013), but having experienced traumatic events is not the same as showing signs of PTSD.

Individuals who have been exposed to trauma (Ono, Devilly, & Shum, 2016), especially if they have developed PTSD (Moore & Zoellner, 2007; Ono, Devilly, & Shum, 2016), tend to report generic, autobiographical memories (e.g., “*I feel happy when I read books*”), even when they are

explicitly asked to recall specific memories (e.g., “*I felt happy the day I could not stop reading a book by Isabel Allende*”). This phenomenon is commonly referred to as overgeneral memory (OGM; Williams & Broadbent, 1986). Aside from trauma exposure with or without PTSD, depression (van Vreeswijk & de Wilde, 2004) and other emotional and affective disorders (Williams et al., 2007) are commonly associated with OGM.

Individuals with an overgeneral memory tend to face slow recovery from depression, and OGM tends to linger even after recovery (Raes, Williams, & Hermans, 2009). It has therefore been suggested that psychological treatment for OGM itself could reduce vulnerability to depression and increase the chances of successful recovery (Raes et al., 2009). For this purpose, a training program has been developed to enhance memory specificity in clinical populations (this training program is further discussed in a section below). It has been suggested that overgeneral responses may be a result of an inability to remain focused during retrieval (Williams et al., 2007).

OGM tends to be stable over time (Sumner, Mineka, Zinbarg, Craske, Vrshek-Schallhorn, & Epstein, 2014); however, individuals can be conditioned to provide more specific answers (Debeer, Raes, Williams, Craeynest, & Hermans, 2014). In contrast, it seems more difficult to condition generic responding. This finding may indicate that overgeneral responding is not learned but rather a dysfunction during retrieval that can be at least temporarily reversed. This can have important implications for legal investigations because it shows that recollection of specific memories really can be enhanced – even for individuals prone to overgeneral responding – through adequate support before or during retrieval.

Memory for Repeated Events

The term “episodic memory” was coined by Neisser (1981) to describe episodic recollection of repeated events that has an overall correctness although the specific details may be ascribed to the wrong event. The concept is closely related to – albeit different from – the terms “schemata” (Bartlett, 1932) and “script” (Schank & Abelson, 1977). The meaning of schemata is any general knowledge about the world around us: about events, people, things, places, and actions. We may have schematic knowledge about our own close relatives or family members in general (e.g., the role of mothers, fathers, siblings), and about places we may have visited or not (e.g., that in France you can visit the Eiffel tower, smoke Gauloise, and taste nice wines).

In contrast, script refers only to schematic knowledge about events (Baddeley et al., 2009), for example, how a restaurant visit, pocket theft, or

dental visit usually occurs. Script is thereby a key concept when discussing memory for events, although other types of schematic knowledge may influence our memory of particular details related to an event (e.g., script is relevant to our memory of a restaurant visit, although our memory of the waiter may be more shaped by other general knowledge we have). Hence, episodic memories (Neisser, 1981) can be influenced by the scripts we have for different types of events (Schank & Abelson, 1977) as well as other schematic knowledge (Bartlett, 1932).

Repeated events tend to be recalled in a general manner (e.g., Roberts & Powell, 2001). We recall how it usually happened (the gist) but face difficulties when trying to specify details or describe individual episodes (verbatim memory). Gist-based information takes less effort to process and it may be stored independently of the details (Reyna & Brainerd, 1995). In addition, it is retrieved differently in the sense that different retrieval cues seem to evoke mainly gist or details (Brainerd & Reyna, 2004). For example, thematic cues as “*food*” tend to evoke gist memories, while specific cues as “*potato*” tend to generate detailed memories. The gist of encoded information is usually remembered well while the specific details are more vulnerable to errors (Surprenant & Neath, 2009).

According to an overview by Reyna and Brainerd (1995), specific details fade more rapidly than gist memory, and there is a higher risk that details will be inaccurately recalled. For instance, gist information can be highly influenced by schematic knowledge (e.g., Koriat, Goldsmith, & Pansky, 2000) which in turn may lead to the inaccurate remembrance of details. A frequently cited example of this is that when people are presented with a list of related words (e.g., table, sit, legs, sofa, rest, desk), they tend to falsely recognize another related word (e.g., chair) as present (Roediger & McDermott, 1995). In addition, it can be difficult to trace the origin of details, for instance, pinpointing when a particular topic was discussed during a holiday with the family (a source monitoring problem; Johnson, Hashtroudi, & Lindsay, 1993). Hence, people can often remain accurate by reporting gist-based information instead of specific details (Reyna & Brainerd, 1995). This implies that it should not always be preferable to extract specific information since sooner or later it will most likely be on the cost of accuracy. If one does however want to enhance recollection of specific details one should employ cues that are specific (Brainerd & Reyna, 2004).

Specificity in recollections of a unique and single event refers to a detailed account about the event, for example, what happened, when it happened, and who was involved. However, when discussing repeated events, the reported memories can be specific in at least two different ways. One can be specific

by providing details in general about the episodes (e.g., “*I’ve had the same dentist since I moved to town 10 years ago. Her name is Paula. She is very gentle and I like her a lot, but sometimes she’s a bit confused and forgets things*”), and one can be specific by providing details about a particular episode (e.g., “*On one occasion my dentist misplaced her glasses and I had to wait while she looked for them, and on a different occasion we had to cancel the appointment because she had scheduled me on the wrong day*”). Although there are many details in the former example, there is no specificity of individual events as there is in the latter example. Both types of specificity are relevant when discussing recollections of repeated events.

Although also relevant to other areas (e.g., survey methodology, Means & Loftus, 1991), most research on repeated events has been conducted within the area of legal and forensic psychology. The largest body of literature so far has concerned repeated abuse of children, more specifically child sexual abuse (e.g., Connolly & Gordon, 2014; Connolly, Price, Lavoie, & Gordon, 2008; Powell, Roberts, & Guadagno, 2007; Roberts & Powell, 2001). Reports of repeated events and abuse may, to some extent, face the same or similar difficulties regardless of age. However, research on adults’ memories for repeated events has, so far (and to the best of my knowledge), not concerned abuse. When adult asylum seekers’ statements about repeated events are discussed (e.g., UNHCR, 2013) it is instead literature on survey methodology (e.g., Cohen & Java, 1995) that is cited. In addition, Neisser (1981) presented a case study on John Dean’s written and verbal testimonies of the Watergate events. John Dean was a key witness in the case against President Nixon in the 1970s, and the testimonies regarded repeated and similar meetings between Dean and Nixon.

Adults’ memories for repeated events have also been investigated within a new line of research that aims at developing techniques for interviewing detainees (i.e., individuals who are held at military bases and suspected of terrorism or suspected to hold information about such activities). For instance, Leins, Fisher, Pludwinski, Rivard, and Robertson (2014) studied how to enhance recollection of individual episodes from a series of repeated and similar meetings that detainees might have attended, for example political or religious meetings during which activists and combatants are sometimes recruited for guerrilla warfare.

Thus, the study of memory for repeated events has great implications for legal matters (repeated abuse against children and adults, and repeated events of various kinds experienced by asylum seekers, detainees, or witnesses). Except from the research on survey methodology (e.g., Means & Loftus, 1991) and Neisser’s (1981) case study, a common feature of the previous research is that the respondents/participants do not belong to the population

for which the research is generalized. This means that conclusions about memories of abuse or detainees' recollections are uncertain. We may conclude that memories of repeated events in other populations suggest that it is in a certain way, but we should be careful when discussing implications for other populations or types of events.

Although it is reasonable to believe that our memory for repeated events has some basic features that are common across event types, research on related topics needs to be taken into account as well (e.g., how our memory is affected by trauma and psychopathology, interrogation/interview methods, effects of long periods of isolation, and humiliating treatment in custody).

Memory for Events Questionable in Veracity

Accounts may be inaccurate in different ways and for different reasons. Sometimes they are inaccurate because of honest mistakes, for example, because of the reconstructive nature of memory (Bartlett, 1932) or due to external factors such as social influence (e.g., implicit or explicit suggestive questions and misinformation; e.g., Loftus, 2005; Loftus & Palmer, 1974; Schooler, Gerhard, & Loftus, 1986). Sometimes we form memories, so-called false memories, for events that we in fact never experienced (e.g., Ost, Granhag, Udell, & Roos af Hjelmsäter, 2008; Sjöden, Granhag, Ost, & Roos af Hjelmsäter, 2009). Sometimes, such a memory can persist in an individual even after s/he learned that the memory was false (for a recent review on non-believed memories, see Otgaar, Scoboria, & Mazzoni, 2014).

Other times, an account is inaccurate because of deliberate deceit. Such deceit may in turn occur because of different reasons. Among the obvious and possible reasons is that the person wants to conceal a wrongdoing. Less obvious reasons may be that the truth is shameful or could harm someone. Occasionally, in legal settings, a witness or suspect provides a deceitful confession, a so-called false confession, of a crime to cover for someone else's wrongdoing.

Empirical population-based survey research from Iceland showed that about 12% of the participating prisoners had confessed falsely during a police interview at least once in their lifetime (Gudjonsson & Sigurdsson, 1994). Most commonly, the false confessions had been provided to protect someone else, but another common reason was to escape a very uncomfortable and pressing police interview. Similarly, about 30% of exonerated, wrongfully convicted prisoners had falsely confessed to the crime (or made incriminating statements against themselves), usually because of coercion and harsh interview methods (Innocence Project, 2016; Kassin, 2015). Today, it is a

well-established fact that there is an increased risk of eliciting false confessions if coercive interview techniques are employed (Kassin, Drizin, Grisso, Gudjonsson, Leo, & Redlich, 2010). However, false confessions also do occur in the absence of harsh techniques. For instance, an individual might be motivated to falsely confess to protect someone else (Gudjonsson & Sigurdsson, 1994), maybe even having been paid to do so.

Arguably, uncoerced, false confessions are typically provided by individuals willing to give a (more or less) complete narrative about the supposedly experienced events. In that sense, such confessions could be viewed as deceitful accounts made by cooperative suspects. This is the main perspective we employed in Study I in which uncoerced, false confessions were analyzed and compared to (uncoerced) true confessions. Therefore, in this thesis as well as in the published article for Study I, the terms 'false confessions' and 'deceptive accounts' (or simply 'lies') are used interchangeably.

It happens that we ourselves are unsure of whether, for example, a memory from childhood is based on real experience or perhaps only imaginations based on photographs or stories told by our parents or even other people who were present at the time. For this specific purpose of making it possible for people to distinguish their own memories of imagined events from memories of events that truly were experienced, Johnson and Raye (1981) developed the Reality Monitoring Framework. During a period of several years, which resulted in a number of papers, they developed, refined and tested this framework. One of the findings during this period was that there are qualitative differences between memories for imagined events and memories for truly experienced events. Some factors can make them more difficult to distinguish, for example, the extent to which the events (imagined or not) have been rehearsed in memory. Talking or thinking much about the events have the tendency to make memories for imagined and experienced events more alike (Suengas & Johnson, 1988). In general, though, memories of truly experienced events tend to include more details of various kinds, for instance more visual, auditory, perceptual, temporal, and spatial details (Johnson, Foley, Suengas, & Raye, 1988). The Reality Monitoring Framework would later be further developed into a criteria-based technique (simply referred to as Reality Monitoring; Alonso-Quecuty, 1992) for distinguishing truthful statements from deceptive statements. There are additional criteria-based techniques with the same purpose and fundamental starting point. There are differences in the amount and quality of detail described when reporting about events that were only imagined (or made up) and events that were truly experienced. We will return to this subject in a later section (see the section *Statement Analysis*).

Eliciting Details

Investigative Interviewing

Verbal accounts from interviewees are the most common evidence in criminal investigations. However, according to surveyed police investigators, the amount of detail provided by the interviewees is often insufficient (Kebbell & Milne, 1998). Since the 1980s, experimental psychologists have therefore set out to develop interview techniques that are based upon scientifically generated knowledge from cognitive and social psychology. The most tested and widespread scientifically developed interview technique is the Cognitive Interview (CI; originally developed by Geiselman, Fisher, MacKinnon, & Holland, 1985). CI is a good starting point to discuss investigative interviewing, especially in this thesis, because of its strong base in memory psychology.

CI originally consisted of four mnemonic techniques aimed at facilitating detailed recollections: context reinstatement (the interviewee is guided to close eyes and think back to the events), an instruction to report everything, an instruction to recall the events from a variety of perspectives, and an instruction to recall the events in different temporal orders. A few years later, Fisher and Geiselman (1992) published an enhanced version of CI (ECI) in which a more general approach to investigative interviewing was introduced, aimed at optimizing the prerequisites for accurate and thorough recollections. ECI included rapport building, not interrupting the witness, and posing open-ended questions. Although CI and ECI were originally two different techniques, ECI comprises what we today usually refer to as CI. In line with this, unless otherwise stated, ECI (and other modified versions of CI) is included in what I hereby refer to as CI.

About 30 years of testing have resulted in at least 70 studies that generally support the hypothesis that CI elicits more detailed accounts than a standard approach to interviewing witnesses (Memon, Meissner, & Fraser, 2010). Crucial content in CI, such as rapport building, eliciting a free recall, employing open questions, and asking the witness to report everything, is also recommended in other scientifically based interview techniques such as the NICHD protocol (specifically developed for interviewing children, La Rooy et al., 2015). CI techniques (Fisher, Milne, & Bull, 2011) and the NICHD protocol (La Rooy et al., 2015) are currently used in practice by police officers in several countries, for example in Norway, the UK, and Canada, when conducting interviews with witnesses.

When interviewing about events that have occurred repeatedly, one is faced with the difficult task of distinguishing between gist details and details belonging to a particular incident. In the scientific literature, the process of unwinding these types of memories is commonly referred to as *particularization* (Powell et al., 2007). Differently put, repeated events are often recalled as a general cluster of events which, in applied settings (e.g., legal cases), needs to be particularized. Child abuse is among the types of criminal events that typically occur repeatedly. The NICHD protocol therefore includes explicit instructions to the interviewer to continuously ask whether something described by the witness happened once or several times.

A few studies (Cohen & Java, 1995; Leins et al., 2014; Means et al., 1989; Means & Loftus, 1991; Rivard, Fisher, Robertson, & Hirn Mueller, 2014) have aimed at specifically testing the effects of CI or CI-inspired techniques on the particularization of adults' memories for repeated events. Cohen and Java (1995) found that although CI did increase recollections to some extent (an increase of 6%), it did so to a lesser extent than a recognition checklist (29%; partly derived from the participants' own notes about their experienced events). Rivard et al. (2014) taught CI to professional interviewers at the Federal Law Enforcement Training Center and compared it to the standard interview that is routinely taught there. The results showed that CI was the better technique for particularization, although the results were statistically significant only after excluding one of the participants. This is in line with the results from Leins et al. (2014) who found that CI was more effective for particularization than a control interview. Similarly, Means et al. (1989; 1991) found that a CI-based interview – in combination with a set of other individual mnemonics (e.g., construction of a timeline; see more about specific mnemonics in the section below) – increased the number of individual events recalled. Thus, it seems that CI is somewhat better for particularization than a comparison interview, although the results are not as clear cut as when it comes to interviews about single events.

Specific Mnemonics

During an investigative interview, specific mnemonics can be added with the hope of eliciting even more (accurate) information. As mentioned above, CI originally consisted of four mnemonics (Geiselman, et al., 1985): context reinstatement, report everything, recollection from a variety of perspectives, and a change of temporal order.

Additional mnemonics were introduced in ECI, for example instructions to hold a particular image in mind and zoom in on it, and to follow a moving

object in mind as if it had been lit up by a moving spotlight (Fisher & Geiselman, 1992). Since CI consists of packages of mnemonics that tend to differ more or less across studies, we know little about the effectiveness of the individual mnemonic techniques.

Similarly, Leins et al. (2014) tested the effects of a package of mnemonics on the recollection of repeated family events. The package included the following mnemonics: constructing a family tree, constructing a timeline, normative cues (e.g., naming holidays that are common for families to get together), derived cues (cues derived from other individuals who have experienced similar situations), self-generated cues (asking questions that can activate the interviewee's associative network and help her generate own cues, e.g., asking why family events occur), and sub-categorization (e.g., asking the interviewee to think about events involving certain family members).

Furthermore, the participants were asked about frequency, a technique also suggested by Brubacher, Powell, and Roberts (2014) for the purpose of interviewing child witnesses, and, as already mentioned, this technique is included in the NICHD protocol. The results showed that the package of mnemonics enhanced the recollection of individual events. Similarly, Means et al. (1989: 1991) found that their CI-inspired package of mnemonics increased recollection of individual events. Aside from CI-mnemonics (Fisher & Geiselman, 1992), their interview package also included the construction of a timeline and specific probes such as thinking about the first visit, the last visit, and whether or not any events or details were different in any way.

The latter mnemonic is also one of the techniques recommended by Brubacher et al., (2014) who summarized mnemonics they consider to be ready for use in police interviews with child witnesses. They recommended four specific mnemonics: episodic memory training (a practice in elaborate memory reporting conducted just before the interview starts), asking for the gist before asking for details, asking how many times something happened, and asking about differences between events and details.

Relatively few mnemonics have been tested separately, which makes it difficult to draw conclusions about the effectiveness of individual mnemonic techniques. Some may be strong cues for recollection (or maybe only during particular settings), while other may be less valuable or even superfluous. Among the mnemonic techniques that have been tested separately are derived cues (Philips & Fisher, 1998), the construction of a timeline (Gosse & Roberts, 2013), and a type of memory training with adults aimed at lowering the number of inaccurate reports and increasing resistance toward unanswerable questions (Scoboria, Memon, Trang, & Frey, 2013). Episodic memory training with child witnesses has also been empirically tested and is

recommended by Brubacher et al (2014). A similar type of memory training has been tested with adults for other purposes in clinical populations. Since similar training may also be useful in investigative interviews, I will briefly describe it in a separate section below.

Memory Training

Individuals with affective disorders tend to recall memories that lack specificity (Williams et al., 2007); they suffer from an overgeneral memory (OGM). Individuals with OGM tend to recover more slowly from depression (than individuals who recall specific memories), and after recovery from depression they often continue to have OGM (see Raes et al., 2009, who summarizes key findings from several studies). In addition, it has been reported across studies that individuals from non-clinical settings who have OGM are at a higher risk of developing emotional stress and depression (Raes, Hermans, Williams, & Eelen, 2007). Several training programs have therefore been developed during the past years, aiming at increasing memory specificity in individuals with different types of psychological disorders. The *MEemory Specificity Training* program (MEST; Raes et al., 2009) originally consists of four training sessions of one hour each. The program runs over four weeks and the sessions are held in groups led by an instructor. In the first session the instructor holds a lecture about memory function related to depression, and the meaning of OGM is explained to the participants using examples. In the subsequent three sessions, memory specificity is thoroughly practiced. Between sessions, participants have homework consisting of further practice in retrieving specific memories. The overall results, from across three conducted studies (Neshat-Doost et al., 2013; Moradi, Moshirpanahi, Parhon, Mirzaei, Dalglish, & Jobson, 2014¹; Raes et al., 2009), are that MEST successfully increases the number of specific memories recalled. The positive effects of the training remained after two (Neshat-Doost et al., 2013) and three months (Moradi et al, 2014).

Similar training programs have been conducted with older adults (Serrano et al., 2004; Serrano et al., 2012) and with individuals diagnosed with schizophrenia (Ricarte et al., 2012; Ricarte et al., 2014). Again, the overall findings are that memory training has a positive effect on the retrieval of specific memories. One training program was administered online to non-depressed college students (Maestas & Rude, 2012). The program consisted of training in expressing oneself in one of three different ways of writing. The

¹ The participants in Moradi et al.'s (2014) study were diagnosed with PTSD.

results showed that writing about one's feelings and thoughts, as well as writing vividly with a lot of details, increased memory specificity. Most surprising, though, was the largest increase found at a six-month followup and not immediately after training. In addition, training in mindfulness has proven successful in increasing memory specificity (Hargus, Crane, Barnhofer, & Williams, 2010; Heeren, Van Broeck, & Philippot, 2009; Williams, Teasdale, Segal, & Soulsby, 2000).

Only one training program so far has failed to increase memory specificity. Mogoșe, Brăilean, and David (2013) had dysphoric undergraduate students going through concreteness training (originally developed by Watkins, Baeyens, & Read, 2009, and Watkins & Moberly, 2009) which was administered online. The training consisted of reading two written scenarios each day for about a week (e.g., "*It is your birthday. Your family organized a great surprise party for you at home.*"). The participants were instructed to concentrate on each scenario for two minutes and try to picture them in thought as a movie. This training did not enhance more specific recall than in students who did not receive any training at all.

The above review of studies indicates that cognitive training might be a useful tool for enhancing recollection in forensic interviews, for example in cases where particularization of repeated events is an issue. However, such training has never been tested for this particular purpose. It is important to note, though, that none of the studies above has controlled the accuracy of retrieved memories. This is an overall weakness in the OGM/AMT research (Zlomuzica, Dere, Machulska, Adolph, Dere, & Margraf, 2014). Furthermore, most of the studies had very few participants (sometimes less than 10). The findings should therefore be interpreted with caution.

Analyzing Details

Statement Analysis

After having elicited as detailed statements as possible during the interviews, the statements need to be analyzed. In legal practice, the main purpose of the analysis is to judge the veracity of the statements. That is, making a professional judgment on the extent to which the interviewee reported the events in a truthful manner. Veracity judgments are basically made daily in legal investigations by, for example, police officers, prosecutors, judges, juror members, and immigration board personnel. Statement analysis in legal practice is rarely conducted in the systematic way that is preferred by researchers (see e.g., Kagan, 2003; Strömwall, 2010), and veracity criteria commonly employed by legal practitioners lack scientific support (Willén & Strömwall, 2012).

Scientifically based techniques for assessing the veracity of statements were initially developed in Europe (Germany and Sweden) during the 1960s and 1970s. These techniques were primarily developed for the purpose of analyzing children's statements in very difficult cases concerning child sexual abuse (although the Swedish technique, *Formal Structure Analysis*, had a wider scope of use and was employed in, for example, arson cases; Trankell, 1965). The Swedish technique never underwent any robust scientific testing and was prematurely implemented into Swedish legal practice. In 1992, the highly criticized technique was finally stopped from further use in Swedish courts (Strömwall, 2010). Meanwhile, the German technique was developed further and underwent a successively increasing number of scientific tests. The method, still in use in German courts for particularly difficult cases of child sexual abuse, is called *Statement Validity Assessment (SVA)*. The core part of it is 19 fixed criteria referred to as *Criteria-Based Content Analysis (CBCA)*. These 19 criteria concern different types of details (see Table 2) that are supposed to occur more frequently in truthful statements than in deceptive ones (Köhnken, 2004), which is generally in line with scientific findings (Amado, Arce, & Fariña, 2015; Amado, Arce, Fariña, & Vilariño, 2016; Vrij, 2015).

Based upon more solid theory on memory function – although not yet used in practice – is the statement analytic technique called Reality Monitoring (RM). The technique was originally developed during the 1980s and 1990s by Johnson and Raye (1981) and Alonso-Quecuty (1992). Similarly to CBCA, RM consists of a set of criteria supposedly useful to distinguish truthful statements from deceptive statements, which scientific

investigations generally support (Vrij, 2015). The RM criteria are still in need of standardization, and different researchers tend to employ different criteria (see Table 2 for the criteria employed in Study I). Accuracy rates for distinguishing truths from lies with CBCA or RM generally exceed 70% (Vrij, 2015). CBCA and RM were employed as measures of amount and quality of detail in Study I.

Table 2. Criteria-Based Content Analysis and Reality Monitoring.

CBCA CRITERION	RM CRITERION
1. Logical structure	Visual details
2. Unstructured production	Audio details
3. Quantity of details	Smell
4. Contextual embeddings	Taste
5. Interactions	Physical sensations
6. Conversations	Affective details
7. Unexpected complications	Spatial details
8. Unusual details	Temporal details
9. Superfluous details	Cognitive operations
10. Details misunderstood	Clarity
11. External associations	Reconstruction
12. Subjective mental state	Realism
13. Others' mental states	
14. Spontaneous corrections	
15. Lack of memory	
16. Doubts about own testimony	
17. Self-deprecations	
18. Pardoning the other	
19. Crime-specific details	

Notes. The individual criteria included in RM differ somewhat across studies. The RM criteria listed above are those employed in Study I.

Autobiographical Memory Specificity

In clinical research, amount of detail in the reports of autobiographical (episodic) memories is commonly investigated with the Autobiographical Memory Test (AMT), originally introduced by Williams and Broadbent in 1986. The purpose of the test is to investigate the extent to which respondents show symptoms of overgeneral memory (OGM). In the AMT, respondents are asked to recall specific memories associated with a number of cue words. A specific memory is defined as something which occurred within a one-day time frame (e.g., “*the day we spent in the woods on our last holiday*”).

Notably, this definition corresponds quite well with remembering individual episodes of recurring events, the scope of Studies II and III. The reported memories are then typically categorized, by the researchers, as either specific, specific-extended, or generic; a few additional categories exist, such as “error” and “semantic”, but not all researchers make use of the same additional categories (Griffith et al., 2012). Individuals with symptoms of depression (van Vreeswijk & de Wilde, 2004) and PTSD (Moore & Zoellner, 2007) tend more often to report generic memories – despite the instruction to recall specific memories – compared to individuals showing less signs of depression or PTSD (Williams et al, 2007).

The AMT categorization of reported memories has also been applied in other ways without conducting the proper test. For instance, narratives from interviews with children who had witnessed domestic violence were divided into short utterances and then categorized as either specific, specific-extended, or generic (Orbach, Lamb, Sternberg, Williams, & Dawoud-Noursi, 2001). Similarly, Hargus, Crane, Barnhofer and Williams (2010) employed this categorization procedure in narratives from individuals interviewed about self harm and suicide attempts. Employing the AMT categorization procedure on narratives has shown a positive correlation with the outcome from a properly conducted AMT (Sumner, Mineka, & McAdams, 2013). AMT categorization was used as a measure of amount and quality of detail (in narratives) in Studies II and III.

Counting the Details

It should be noted that, currently, it is only researchers who conduct a proper count of details included in a narrative. Legal practitioners instead tend to make a more general assessment of whether or not a particular statement is detailed enough (e.g., Willén & Strömwall, 2012), for instance, whether or not the statement contains the amount of detail one can expect it to contain

given the particular circumstances. Several researchers and practitioners have suggested the implementation of more structured analyses and counts (see, for example, Kagan, 2003), and lately handbooks (e.g., Gyulai, Kagan, Herlihy, Turner, Hárđi, & Udvarhelyi, 2013) have been published aiming at educating legal practitioners on a more scientifically based procedure for assessments of credibility and reliability in asylum cases. However, despite striving toward more objective measures, there are still no measurements used in practice for assessing amount of detail.

Researchers, on the other hand, do have methodological procedures for measuring amount of detail. Whether a statement analysis (such as CBCA or RM) or an AMT categorization of reported memories is conducted, researchers will go through at least three crucial steps when analyzing the narratives. How each of these steps is handled by the researchers can potentially change the study conclusions substantially, although the content of the narratives remains the same.

In the first step, the researcher must conduct the categorization of the verbal content. This type of categorization is essential when striving toward objectivity and away from general and overall assessments. Without a categorization into predefined groups of details, we will not know what is meant with a “detailed” account. For CBCA and RM, the categorization is conducted according to the criteria of classifications, for example, to which extent the narrative includes references to *superfluous details*, *unusual details* or *unexpected complications* (and all the other criteria). Similarly, in the AMT procedure, researchers categorize memory reports as either generic, specific-extended, or specific (and often into a few additional categories).

These categories may be theoretically distinguishable from each other. In practice, however, it is quite frequently the case that different researchers interpret the one and same criterion/category differently (e.g., Sporer, 2012; see also the AMT-categorization conducted in Study II). In addition, the same verbal content can be understood and thus categorized differently by different researchers. Low intercoder agreement (also called interrater agreement or interrater reliability) can seriously threaten the validity of a scientific investigation. It is thus crucial that researchers using non-standardized measures do measure the extent to which different raters came to the same conclusions on the same material. The interrater agreement must also be correctly and transparently reported in the article.

Unfortunately, the importance of this issue is too often neglected. It is, for instance, still common that interrater agreement over the AMT categorization is unsatisfactorily reported with simple percentage agreement, despite the fact that percentage agreement does not take into account chance agreement (i.e. agreement occurring by chance). Also highly relevant is that the presence of a

criterion, in criteria-based analyses such as CBCA and RM, is measured differently. Most studies employ Likert scales (e.g., 1–5) while others employ a binary measure (present vs. absent). As pointed out by Sporer (2012), the Likert scales can also have different ranges (e.g. three points or seven points) and different endpoints (e.g., zero and two, or one and seven). The type of measure employed is likely to affect the interrater reliability (i.e. fewer scale points will likely result in higher agreement). Choosing which categories are included in the calculation of interrater agreement is another researcher-made decision that can actually affect the interrater agreement – and thus also the study conclusion.

In the second step, it is time for the researcher to conduct the statistical analyses. The lack of standardization in the AMT procedure has not only resulted in the use of different categories in different studies, but also a lack of agreement across studies on which categories shall be included in the statistical analyses (Griffith et al., 2012). Neither is there agreement across studies on the primary outcome measure. For example, while some studies report analyses of the proportion of specific memories, others report the proportion of generic memories, and still others report the proportion of both types of memories (Griffith et al., 2012). The outcome measure is the core of any scientific study. No need to say that different outcome measures may result in substantially different conclusions. Crucial to the interpretation of findings is whether such decisions were made prior to or after the data collection was conducted, or at least before any analyses (e.g., the categorization) had been initiated. There is a substantial risk of questionable research practices (such as *p*-hacking) being employed when there is too much room for flexibility – and such practices can significantly change the outcome and conclusions (Simmons et al., 2011).

Thus, the third step regards how transparently the researcher decides to report the procedures. It is currently common practice to not make it explicit in the report *when* a particular decision was made, for example when it was decided which outcome measure to employ as the primary one. Lack of explicit and transparent reporting can lead colleagues astray and severely hamper knowledge accumulation (Nosek, Spies, & Motyl, 2012). Arguably, future research in AMT and statement analysis will benefit much from being preregistered, enabling other researchers to make more accurate decisions on which ways forward may be the most fruitful.

An additional methodological issue for the research on investigative interviews is that amount of detail is commonly measured separately for different parts of the narratives (e.g., free recall only, specific questions only, and for the narratives in whole; see e.g., Leins et al., 2014, for an example of this). In addition, researchers employ different interview protocols in

different studies. These factors together make comparisons of results across studies complicated. Thus, measuring the amount of detail reported by interviewees is far from a straightforward procedure. It surely has the potential of getting more straightforward by, for instance, employing preregistration, transparent reporting, and standardization of employed measures. This could in turn result in the development of more objective measures, with the potential for use also in legal practice. Currently, though, it is quite a mess, and this should be kept in mind when interpreting and comparing study results from these fields.

Summary of the Included Studies

The present thesis consists of three empirical studies (Studies I, II, and III) and one data paper (Study IV). As mentioned in the Introduction, the overall aim of the thesis was to employ research designs with high ecological validity to investigate the amount and quality of detail in interviews with cooperative interviewees. The three empirical studies included in the thesis were based on two data collections. Both data collections were experiments designed to prioritize ecological validity on the cost of laboratory control, resulting in designs commonly referred to as quasi-experiments. The ecological validity in both experiments were prioritized in the sense that participants were recruited because they had experienced certain types of events (criminal offenses or dental visits), rather than being a randomly selected sample recruited to experience a particular event controlled by the researcher. Experimental research on investigative interviewing typically employs the latter design, which is important in terms of establishing or outlining basic principles of psychological phenomena. Such research must however be supplemented with ecologically valid research designs. Thus, an essential contribution of the present thesis to previous studies in investigative interviewing is the providing of three studies with high ecological validity.

The primary outcome measures in the included empirical studies are measures of amount and quality of detail in the interviewees' verbal reports. Such outcomes tend to be operationalized differently in different studies, and a secondary aim of the present thesis was to bring this topic into the light by discussing potential consequences of the lack of standardization of these measures.

Each study included in the thesis is briefly introduced below, followed by more detailed summaries of aims, methods and results.

Study I investigated amount and quality of detail in true and false confessions told by offenders. Amount and quality of detail were measured by employing Criteria-Based Content Analysis (CBCA) and Reality Monitoring (RM).

Study II investigated the effectiveness on particularization of adding one interview after presenting interviewees with one of three types of memory cues (mnemonics). Specifically, we investigated the effect of context-specific cues derived from other respondents in an initial study. No previous studies had investigated the effects of specific mnemonics on the particularization of adults' memories for repeated events. In addition, there is no previous

(published) research on the effect of derived, context-specific cues on the particularization of repeated events. Study II therefore aimed at increasing our knowledge about the effectiveness of specific mnemonics – more specifically, the effectiveness of derived cues – on adults’ memories for repeated events. Amount and quality of detail in the memory reports were studied using different two measures of specificity: number of events referred to and amount of detail provided about the events.

In Study III, we continued exploring the data initially collected for the purpose of Study II. In this study, we elaborate on the importance of distinguishing between two measures of specificity when studying memory for repeated events: number of individual events referred to and amount of detail recalled about the events. The influence of five factors on these two measures of specificity were investigated.

Finally, Study IV is a data paper making the data from Studies II and III freely available online. Thus, the collected and coded data can easily be reused by others for any purpose, such as teaching or conducting additional research.

Disclosure Statement

Psychology is currently facing a reproducibility crisis (Open Science Collaboration, 2015). One important reason for this may be that published research too frequently has weaknesses that were never disclosed in the publication (John, Loewenstein, & Prelec, 2012; Simmons, Nelson, & Simonsohn, 2011). This is a very serious problem, and the need for complete and transparent reporting of methodological procedures and analytical decisions was recently highlighted by the world's biggest statistical organization, the American Statistical Association (ASA), which have never before during its 177 years of existence made an official statement of this kind (Wasserstein & Lazar, 2016). One way of practicing transparent reporting is to include a disclosure statement in the publication (Simmons, Nelson, & Simonsohn, 2012), where what is reported and not is made explicit. It would be a substantial waste of efforts and resources to continuously conduct research with no reproducible effects. In the field of legal and forensic psychology there is also an impending risk that individuals can get seriously harmed if legal decisions are based on non-robust science. Thus, the initiation of disclosure statements is hereby embraced.

Disclosure Statement for Study I

A rough sample size of about 30 participants was decided in advance. No formal power calculation was conducted and the participation rate was expected to be low. Initially, 36 offenders participated in the experiment. The narratives from six participants were excluded for different reasons. Three respondents were excluded because – on the respondents' initiative – the narratives were not focused on the purpose of the interview (i.e., the experiment). One was excluded because the false confession was so lengthy that the interviewer had to end the interview prematurely. One respondent was excluded because the narratives were severely incoherent to the extent that they were not possible to understand. One respondent was excluded because the interviewer had reason to believe that the event behind the supposedly true confession actually had taken place. All six respondents were excluded prior to the statistical analyses being made. Some of the excluded narratives were used by the research assistants/coders for training purposes.

In line with findings from Willén and Strömwall (2012), it was initially predicted that gender and interview experience would influence the outcome on CBCA and RM scores. These analyses were not statistically significant ($p > .05$). In line with common publication practice (John et al., 2012), these

predictions were therefore deleted from the report and the two variables instead included as covariates.

Finally, for the purpose of any future research aiming to replicate the study, it should be noted that the interviewer also had criminal experience and that all respondents were briefly informed about this (albeit not the nature of the experience) during the recruitment process. Before granting access, the prisons' heads of security did a thorough check-up to ensure that the interviewer was not familiar with any of the current prisoners (i.e., not only potential respondents).

All manipulations and measures are reported in the paper.

Disclosure Statement for Study II

A rough sample size of about 30 respondents in each condition was decided in advance. No formal power calculation was conducted.

There was no analysis plan prior to the data collection. It was decided after the data collection as to how the amount of detail was to be operationalized and statistically analyzed. The first statistical tests conducted were two-tailed ANCOVAs with a $p > .10$. After reviewing the predictions and the choice of statistical test, it was decided that contrast analyses would be more suitable tests of the predictions.

Some measures were omitted from the report of Study II. All measures are thoroughly reported in the data paper (Study IV).

The predictions concerning rehearsal and unpleasantness were made after the data collection.

All data exclusions and manipulations are reported in the paper.

Disclosure Statement for Studies III and IV

All made decisions as well as the process were completed just as they are reported in the papers.

Study I

Aims

Research on deception detection is typically conducted in a mock-crime scenario with undergraduate students. The main purpose of the present study was to study these issues in a more ecologically valid setting. We investigated the amount and quality of detail in true and false statements by criminally experienced respondents. Amount and quality were measured with Criteria-Based Content Analysis (CBCA) and Reality Monitoring (RM). In line with previous research on CBCA (Undeutsch, 1982; Vrij, 2015) and RM (Johnson & Raye, 1981; Vrij, 2015), we expected both techniques to generate higher scores in truthful statements than in deceptive ones.

Methods

Thirty offenders, serving time in prison at that time, provided one true and one false confession in an experiment conducted at three Swedish low-security prisons during 2009. The prisoners received written and verbal information about the study from the experiment leader/interviewer at two separate occasions: a couple of weeks prior to the study taking place and on the same day as the study was going to take place. Participation was anonymous and voluntary. Each respondent received compensation equivalent to about six euros.

Participants were interviewed one by one by the same interviewer. In one interview, the respondent gave a true confession about a crime having personally committed and been sentenced for. In another interview, the respondent gave a false confession about a crime having never personally committed or been sentenced for. Each interview lasted about 5–10 minutes and was preceded by a few minutes during which the interviewee prepared for the upcoming interview. The participants did not know the topic of their false confession until they were given it from a research assistant who selected it from a long, prepared list of crimes. The selection of topics was carried out to roughly match the type of crime concerned in the true confession. Participants were never given a violent lie scenario if their truth did not involve any violence at all. In addition, it was regarded whether the topic of the truthful account concerned one specific event or a series of events. For example, when the truth concerned an isolated incident, the

participant would not receive a lie scenario that concerned a series of events. The order of interviews were counter-balanced; the number of respondents giving the true confession in their first interview was the same as number of respondents giving the false confession in their first interview.

Each interview started with the interviewer pointing out that she did not know if the participant was going to give a true or false confession in that particular interview. The participant was then reminded to give each statement in a convincing manner irrespective of the objective veracity status. After a free recall, the interviewer used at least two and never more than five open questions (e.g., “*Do you remember anything more?*”) in order to elicit a complete statement. The number of open questions depended on how much the interviewee revealed during the free recall.

The interviews were audio taped and transcribed verbatim.

Four research assistants, all blind for the objective veracity status of the statements, were trained in either Criteria-Based Content Analysis (CBCA) or Reality Monitoring (RM). Each narrative was scored for each criterion on a five-point rating scale. The interrater reliability was excellent for both CBCA and RM.

Results

Overall, neither CBCA nor RM were successful techniques in distinguishing the true confessions from the false, giving no support for any of our two predictions. That is, there were no statistically detectable differences between the total CBCA scores or the total RM scores between the truthful and deceitful statements. No individual RM criteria showed differences between the true and false statements, but three individual CBCA criteria could statistically differentiate the statements. Surprisingly, there were more *self-deprecations* and *doubts about own testimony* in the false confessions than in the true confessions. In addition, there were more *unexpected complications* in the truthful confessions.

Conclusions

Previous research conducted with respondents who lack criminal experience has shown that CBCA and RM can be used to distinguish true accounts from deceptive ones. The observed results were therefore unexpected, and replications are important to reveal whether the (almost nil) effects with criminally experienced individuals will replicate.

Three individual CBCA criteria did however show differences – expected as well as unexpected – between the true and false confessions. Thus, further research could make it possible to differentiate offenders' true and false confessions, and different types of statements other than confessions, through a scientifically based technique for statement analysis.

General Method for Studies II, III, and IV

One data collection carried out in Gothenburg, Sweden, during the spring of 2012, forms the basis for Studies II, III, and IV. The project was approved (id: 1007-11) by the Regional Ethical Review Board, University of Gothenburg, Sweden.

Participants

The respondents were originally recruited through advertisements in the waiting rooms of 20 Gothenburg-area dental clinics which graciously allowed us access. Due to difficulties in recruiting the large number of participants we needed, we also emailed inquiries to individuals who had announced interest in participating in research at the Department of Psychology. We further asked personnel at a few of the 20 collaborating clinics to verbally inform some of their patients about our study, namely patients younger than 30 years because they were initially underrepresented in our sample, and respondents who have made a very large number of dental visits (≥ 20 visits during the past ten years). The criteria for participating (Swedish-speaking adults having made about 8 visits or more) were the same, however, and differences in the recruitment process are therefore not expected to have influenced the results in significant ways. Participation was voluntary and compensated with a gift card to each respondent (worth about 28 euros, valid at stores and restaurants in Sweden).

Ninety-nine respondents participated in the study, although four had to be excluded (because they did not bring their dental records and because of interviewer mistakes). The final sample consisted of 95 respondents (71 women) with a mean age of 43 years. They reported to have made on average about 17 dental visits each (ranging from 6 to 60) during the past 10 years.

Deriving the Memory Cues for Study I

Prior to the major data collection, an initial study was conducted to derive the cues to be used in Study I. Twenty-seven university students (27–54 years) answered a short questionnaire and were compensated with one lottery ticket each worth about three euros. They were asked to recall as many dental visits as possible made during the past ten years. For every visit they recalled they were additionally asked to answer what made them remember that particular visit. Their answers to the latter question were categorized to form two types

of context-specific derived cues (derived specific cues and derived categories). Eight categories (Table 3, left column) were identified by the first author and a research assistant. These categories were presented to participants in the derived categories condition when collecting data for the major study. To create the specific cues, subcategories within each of the eight categories were selected, five quotations within each main category (resulting in a total of 40 quotations). A selected quotation was either a representation of its subcategory or considered to have high associative potential as a memory cue. These 40 quotations together with the eight categories formed the memory cues in the derived specific cues condition. Both types of derived cues are presented in Table 3.

Comparison cues were selected with the aim of choosing cues that are commonly used in police interviews with plaintiffs and witnesses. The following six cues were selected:

- The last visit
- The second-to-last visit
- The first visit
- The second visit
- Visits that stand out from the others
- Visits that co-occurred with a holiday or similar

All cues in all conditions were printed on cards and started with an instruction to think back and see whether the cue (which was underscored) would help them recall additional information about their dental visits. For example: “Think back to times and dates” and “Think back to the last visit you made”. In the derived specific cues condition, the quotations immediately followed their category. An example of a card in the derived specific cues condition is found below (the cards without quotations were shorter).

Think back to conversations you had with the personnel.

“I promised the dentist to stop using snuff”

“I was instructed how to prevent caries”

“As usual, I asked if they could make my teeth more white”

“She said I had bad dental hygiene”

“I got calming information”

Does this help you recall additional visits?

Does this help you recall additional details about the visits you have already reported?

Table 3. The derived cues employed in Study II.

CATEGORIES	QUOTATIONS
Emotions and sense-experiences	<p>“Worried that there would be an invasive procedure”</p> <p>“It was unpleasant and painful”</p> <p>“Afterwards, I had speech difficulties from the anesthesia”</p> <p>“It smelled bad when the dentist drained the tooth”</p> <p>“Afterwards, my mouth felt clean”</p>
Conversations	<p>“I promised the dentist to stop using snuff”</p> <p>“I was instructed how to prevent caries”</p> <p>“As usual, I asked if they could make my teeth more white”</p> <p>“She said I had bad dental hygiene”</p> <p>“I got calming information”</p>
Times and dates	<p>“The last two times, I had to make several fillings”</p> <p>“I had to come back a couple of weeks after the routine visit”</p> <p>“My birthday was to come”</p> <p>“I usually get a dental recall card around summer time”</p> <p>“It took several visits to fix”</p>
Places and locations	<p>“Met with a specialist in a new department”</p> <p>“It was when I moved to Sweden”</p> <p>“Had to go in emergency to a different clinic than the usual one”</p> <p>“I still lived in my old flat”</p> <p>“First and only time since I moved to Gothenburg”</p>
Personnel	<p>“Heavy-handed dental hygienist”</p> <p>“I noticed the name of the dentist”</p> <p>“New dentist”</p> <p>“My first visit to a dental hygienist”</p> <p>“The dental hygienist was nice, explained well”</p>
Economy	<p>“I had income support from the social services and they were supposed to pay, but it got too expensive so they didn’t want to”</p> <p>“The quality of my teeth was downgraded”</p> <p>“First visit since I had to pay for it”</p> <p>“Cheaper than usual”</p> <p>“I forgot my bank account number and couldn’t sign for a dental insurance”</p>
Treatments	<p>“Impressioning”</p> <p>“It didn’t bleed when the dentist removed the calculus”</p> <p>“Had to sew in the mouth”</p> <p>“She took some x-rays”</p> <p>“The dentist’s drill”</p>
Reasons for visit	<p>“Recommendation from my sister”</p> <p>“Got problem with the same tooth again”</p> <p>“Destroyed a tooth while biting and lost a filling”</p> <p>“I had chipped one tooth”</p> <p>“It was a different reason than usual”</p>

Notes. Derived categories and specific cues. Both categories and quotations were presented in the derived specific cues condition. Only the categories were presented in the derived categories condition. All cues are translated from Swedish.

The Interviews

All participants were interviewed twice. Both interviews are included and compared in Study I, while only the first interview forms the basis for Study II.

Participants brought their dental records to the experiment in envelopes that had been sealed by the personnel at the dental clinics. Everyone signed an informed consent form before the experiment started. The experimental session lasted between 90 and 120 minutes for each respondent. The interviews were conducted by one of four interviewers who followed a structured interview protocol. The length of the first interview was usually around 30 minutes.

After the first interview, participants were presented with one of three sets of memory cues (derived specific cues, derived categories, or comparison cues). The cues were printed on cards. For each condition there were 6 or 8 cards. The presentation order of the cards was randomized. All participants received between 15 and 20 minutes with the cards, and all were offered an additional 5 minutes if they wanted to. Participants were encouraged to make notes on a paper to use as support during their second interview. They were subsequently interviewed a second time in which they were asked whether they recall anything that they did not report in their first interview. If their answer was positive they would receive the same questions as in the first interview. The second interview usually lasted about 10–15 minutes.

Post-interview Procedure

After the second interview, respondents answered a post-interview questionnaire about demographics and how they had experienced the interview. They subsequently opened their sealed envelopes and had some time to review their dental record if they wished. When they were ready they received instructions from the experiment leader on how to anonymize the dental record. The records were then copied, and the copy was later used in the studies to establish objective truth.

Data Preparation

Number of Events

One research assistant coded all statements and the corresponding dental records. He counted the frequencies for each measure that is included in both

studies (Studies I and II). Another research assistant conducted independently the same coding procedure for 21% of the material. Intraclass correlations were calculated and showed an excellent agreement for all measures (no measure had an agreement less than .94).

Type of Memories

Type of memories measured the incidence of generic, specific, and specific-extended memories (see below), and these variables were employed in both Study I and Study II.

All transcripts were broken down into short utterances similar to the procedure employed in Orbach, Lamb, Sternberg, Williams, and Dawoud-Noursi (2001). The work resulted in about 70,000 utterances. Each utterance was then categorized into one of four categories that are commonly used in the Autobiographical Memory Test (see e.g., Griffith, Sumner, Raes, Barnhofer, Debeer, & Hermans, 2012, and Hargus et al., 2010):

- *Generic* – Summaries of how something usually or typically occurs (e.g., “because I’m often very dry in my mouth when being stressed”)
- *Specific* – A memory of something particular which lasted less than one day (e.g., “they had a trainee there during that visit”)
- *Specific-Extended* – A memory of something particular which lasted more than one day (e.g., “During that period I had a lot of acne”)
- *Error* – Not related to the dental visits or not containing any information (e.g., “I don’t know,” “I can’t say which day it was”)

One assistant categorized all statements, and a second assistant categorized 21% of the statements. The interrater reliability was initially not impressive, but after a training session with both coders the agreement reached an acceptable level (.66; Cohen’s unweighted kappa).

In addition, each utterance in the three categories (generic, specific, and specific-extended) was verified by comparison with the dental records. This work was conducted by two other assistants. Each utterance was coded as confirmed, refuted, or unverified. Again the interrater agreement was initially considered too low and a training session did not improve the agreement enough. Instead, the two assistants worked together with each utterance they disagreed upon until they reached an agreement of 100%. The main coder was thereafter instructed to apply their common rules when coding the rest of the material alone.

Study II

Aims

Previous studies on the particularization of adults' memories of repeated events have either examined the effect of the Cognitive Interview (Cohen & Java, 1995; Rivard et al., 2014) or the effect of interview techniques involving a package of several mnemonics (Means et al., 1989; 1991; Leins et al., 2014). Few studies have examined the effects of single mnemonics on the particularization of adults' memories. Inspired by the unpublished work of Philips and Fisher (1998), the main aim of Study I was to investigate the effect of one particular mnemonic, namely derived cues.

Three packages of cues were compared: derived specific cues vs. derived categories vs. comparison cues. We predicted that derived cues would generate more individual events (Hypothesis 1a) and more detailed memories (Hypothesis 2a) than the cues that are commonly used in police practice. In addition, we expected that derived specific cues would generate more individual events (Hypothesis 1b) and more detailed memories (Hypothesis 2b) than the derived categories. We further expected rehearsal to be positively associated with participants' recollection (Hypothesis 3), and that higher levels of unpleasantness would be associated with better recall of the dental visits (Hypothesis 4).

Methods

This study was a quasi-experiment making use of the experimental manipulation – including one interview prior to the manipulation (Interview I) and one interview following the manipulation (Interview II) – described in the section *General Method for Studies II, III, and IV*.

Results

In line with Hypotheses 1a and 1b, derived cues tended to be somewhat more effective for the recollection of individual events and details about the events than the comparison cues. In contrast, there were no statistically detectable

differences between the two types of derived cues for any of the two measures.

Rehearsal of the memories by telling them to others was associated with an increased number of recollected events and details which supported Hypothesis 3. It mattered most how much the respondents had talked about the events and, to a lesser extent, how much the respondents had thought about their dental visits. Partly in line with Hypothesis 4, higher levels of unpleasantness predicted a small increase in the number of referred visits but not in the number of details about the recalled visits. Importantly, derived cues did not decrease accuracy compared to cues that are commonly used in police practice.

Conclusions

The findings suggest that derived cues might be a valuable contribution to packages of mnemonics that aim at facilitating the particularization of repeated events for adults. However, how specific the cues (derived specific cues vs. derived categories) need to be in order to be as effective as possible should be further investigated. Future research on the usefulness of single mnemonics may consider increasing statistical power by employing larger samples. The effects observed in studies of interview techniques that use several mnemonics cannot be used for drawing conclusions about the size of effects in studies of single mnemonics since the merged effects of several mnemonics likely are stronger than the effects of single mnemonics.

Study III

Aims

Study III extends Study II by further investigating two different measures of memory specificity: the number of individual events and the amount of detail provided about the events. The overall aim was to inspire new research questions and provide new understanding about employed memory specificity measures in research on repeated events. We did so by studying the influence of five factors (interviewees' age, number of experienced events, interviewer, perceived unpleasantness concerning the events, and how much the interviewee had rehearsed the memories) on the two specificity measures. The study was exploratory and therefore the direction of effects were kept open.

Methods

This study was based on the data collected for Study II. The present study did however only investigate data from Interview I, that is, the participants' ($n=95$) memory reports prior to given the memory cues. The transcribed narratives from Interview I consisted of over 40,000 utterances. These narratives and the corresponding dental records were studied for the purpose of the present study. For more details about the methods, please see Study II.

Results

As can be seen in Table 4, it was found that the two specificity measures were affected differently by all five factors. *Recollections of individual events* were positively affected by an increase in number of experienced events, high levels of unpleasantness, and if respondents had talked much about the events (prior to being interviewed). Contrasting these effects, *amount of detail* provided about the events was not affected by any of these three factors. Only the interviewer seemed to positively affect *amount of detail* provided about the events, and there was a decrease of details (but not events) with age. Thinking about the events prior to the interview showed no effect on either measure. Additionally, a particular side finding is worth mentioning: participants typically underestimated how many dental visits they had made.

Conclusions

Interviews about repeated events may become more effective if researchers and practitioners start to distinguish between two types of specificity: number of individual events recalled and amount of detail provided about the events. The results suggest that factors such as number of experienced events and the presence of trauma can affect these two memory specificity measures differently. Interview techniques may therefore be more effective if specialized toward generating either individual events or details about the events. Furthermore, failed interviewee motivation might occur sooner or later. In some cases, it might therefore be necessary for practitioners to prioritize whether to extract additional events or additional details about the events.

Table 4. Influence of five factors on two measures of specificity.

Factor	Referred events	Referred details
Age	No	Yes
Experienced events	Yes	No
Interviewer	No	Yes
Unpleasantness	Yes	No
Rehearsal	Partly ^a	No

Notes. ^a It did matter how much the respondent had talked about the events but not how much they had thought about them.

Study IV

Aims

Because of word limits frequently employed in scientific journals, the method sections in articles often have to be more concise than would be optimal for anyone wanting to replicate the research. Publishing a data paper accompanying the empirical research articles is currently an option to increase reproducibility.

The present study is a data paper, which is published in a (gold) open access journal, consisting of all extracted data and all employed material that formed the basis for Studies II and III. Thus, the purpose of the paper was to make all data and material from Studies II and III publicly and freely available for others to reuse.

Content

The methods for data collection and data extraction for Studies II and III are exhaustively described and accompanied with links to the online repositories where the data (with codebook) and material are stored in open formats. All the data and material are freely available to anyone for any purpose with appropriate citation.

Reuse Potential

The data can be reused as data from an experiment (including both interviews, as in Study II) or as single interview data (including data only from Interview I, as in Study III). The data can be used for research as well as for educational purposes. In addition, the material can be modified and reused for new research.

Not all variables were analyzed in Studies II and III, and the data paper highlights three such variables that could be analyzed in future research: The proportion of dental visits that were never mentioned by the respondents could sometimes exceed 90%. This variable is potentially interesting to investigate further from a memory and/or interview perspective. How many times the respondents had moved during the past ten years and how often they had visited the dentist were questions included in the questionnaire but

never analyzed. These variables could be used to further study, for example, the distinctiveness principle (Surprenant & Neath, 2009).

General Discussion

The three empirical studies comprising the present thesis span across several different research themes – from the detection of false confessions with statement analysis to factors influencing recollection of repeated events and mnemonics that may facilitate recollection of repeated events. Nevertheless, I began this thesis by highlighting three main commonalities between the studies. First, variations in amount and quality of detail were the primary outcome measures. Second, the interviews were conducted with cooperative interviewees. Third, the experienced events took place under natural circumstances and thus were not planned, induced, or controlled by any researcher. These commonalities will compose the structure for the discussion of main findings and implications for future research.

Eliciting Details from Cooperative Interviewees

The results from Studies II and III point out several difficulties of remembering repeated events, but they also demonstrate possibilities in findings ways to facilitate recollection. Study II showed that derived cues could be a meaningful mnemonic for enhancing recollection of individual events, while Study III highlighted outcome differences depending on how amount of detail is operationalized and measured when studying memory for repeated events. The difficulties and possibilities are, respectively, discussed below.

Study III showed that all five investigated factors had different influence on the two measures of specificity (namely, number of events recalled and amount of detail provided about the events). This finding has implications for methodological practice as well as for how techniques can be developed for improving recollections. Future study designs should consider both specificity measures, even if choosing not to study both in the same study, and particularization of repeated events could possibly become more effective if a clear distinction is made between the two measures. Differently put, mnemonics may become more effective if customized to enhance recollection of either events or details, instead of developing mnemonics with the general aim to enhance particularization.

A valid question is why the two measures are affected differently. One reason may be that it is more difficult to specify individual events than it is to describe details about them. Study II showed that the number of recalled events increased from six to eight in the second interview (i.e., after the

participants had taken part of the cues). Hence, the increase was only about two events on average. In contrast, the amount of detail (specific and specific-extended memories) was increased from about 300 to almost 400 utterances. Thus, there are likely more opportunities to trigger memories of details than to trigger memories of individual events. In line with this explanation, the findings in Study III regarding influence of age on recollection of individual events (no effect) and recollection of specific memories (a decrease with age, in line with previous research, Phillips & Williams, 1997; Rabinowitz et al., 1982) may be due to the fact that the number of referred events was relatively small for most participants.

A number of studies with children have shown that recollections of repeated events tend to be more generic and less specific than recollections of single events (e.g., Roberts & Powell, 2001; Schneider, Price, Roberts, & Hedrick, 2011). Somewhat in line with this, Means and Loftus (1991) found that adults' reports about repeated events were less complete than reports about single events. An implication of these results may be that accounts about repeated events will include a larger portion of generic memories as the number of experienced events increases. In Study III we did not find such a relationship. The proportion of generic memories did not increase significantly as the number of events increased. This indicates that the actual number of events is not very important to the proportion of generic memories as long as it regards a series of many events (our respondents had experienced between 6 and 80 events). Hence, in adding our findings to previous research we can conclude that an individual who has experienced a number of similar events – regardless of how many – is likely to provide less specific memories than someone who has experienced one single event.

Study III also showed how difficult it is for people to make accurate estimations of how many times they have experienced a particular type of event. The number of experienced events was frequently underestimated, a finding that is supported by previous research. Underestimations of repeated and similar events have been reported for adults by, for example, Thompson and Mingay (1991) and for children in a study by Sharman, Powell, and Roberts (2011). Estimations of event frequency are common in legal settings (Powell et al., 2007) as well as in surveys (Means et al., 1989; Thompson & Mingay, 1991), and it is important for researchers and practitioners to consider these findings when deciding how to interpret and weight the estimations.

Despite the many difficulties, Study II adds to the literature (Brubacher et al., 2014; Leins et al., 2014; Means et al., 1989; 1991) showing that successful particularization is possible. Specifically, it demonstrated that

derived cues aided particularization to a greater extent than the comparison cues.

Four specific mnemonics have been recommended for use in forensic interviews with children who report about recurring incidents (Brubacher et al., 2014): episodic memory training, asking for the gist before asking for details, asking continuously how many times something happened, and asking whether there were details or events that stood out from the rest. The last mnemonic, asking for differences, has not yet been tested empirically, according to Brubacher et al. (2014), while the others have. Study II is one of the first studies to empirically test the effect of specific mnemonics on the particularization of adults' memories. More research on adults is needed in order to enable similar recommendations to legal practitioners and fellow researchers.

Particularization of adults' memories is sometimes needed in legal settings (e.g., Leins et al., 2014; UNHCR, 2013) as well as in research designs where memories of repeated events are studied (e.g., Means et al., 1989). Different types of mnemonics may be suitable for different purposes. Asking certain questions, as those suggested by Brubacher et al. (2014; e.g., asking for differences), as well as employing a timeline (Gosse & Roberts, 2013; Leins et al., 2014; Means et al., 1989) may be suitable for both purposes. Several questions and a timeline may also be employed in studies conducted online.

The case of derived cues is a bit more complex since it is event specific, but there is nothing suggesting that it would not work in both settings. Other techniques, such as the Cognitive Interview (Fisher & Geiselman, 1992; Geiselman, Fisher, MacKinnon, & Holland, 1985), may be foremost suitable for forensic settings because interaction is needed, otherwise the technique would be so time demanding that online respondents would be fatigued. The development and refinement of interview techniques for legal settings may also gain much from turning to clinical research conducted on overgeneral memory (Williams & Broadbent, 1986) and training in retrieving specific memories (e.g., Raes et al., 2009). It is possible that a similar training could be worthwhile in legal cases concerning repeated events, or that specific and more succinct techniques could be elicited from the clinical training programs. It is important to note, though, that the clinical programs do not consider accuracy of retrieved memories but only how to increase the amount of specific memories. In legal settings, however (as well as in research methodology), accuracy is of key importance.

The distinctiveness principle (Surprenant & Neath, 2009) states that items are better remembered if they differ somehow from the other items. This distinctiveness must not be present during encoding but could be created at the time of retrieval. Events that have occurred repeatedly are, by definition,

less distinct simply because there are many items that are very similar to each other. We therefore need to help interviewees make their memories more distinct. We can do this by developing and testing mnemonics that are effective for this particular purpose. Study III suggests that such mnemonics should be aimed at *either* enhancing particularization of individual events *or* extracting details about the events. Recollections of repeated events can be inaccurate for several reasons (Bartlett, 1932; Johnson et al., 1993; Neisser, 1981; Schank & Abelson, 1977) and we therefore need a specialized toolbox of techniques for memories of recurring events. Studies II and III in the present thesis aimed at contributing to the development of such a toolbox.

Analyzing Details from Cooperative Interviewees

None of the criteria-based techniques employed for measuring amount and quality of detail in Study I could successfully distinguish the truthful confessions from the false. Quite surprisingly, the technique with the most solid theoretical base (Reality Monitoring; Johnson & Raye, 1981) turned out to be the one least successful in distinguishing the statements. Is this the result of a true non-effect? Is it a consequence of the employed research design, such as the fact that the participants told both of their confessions with only minutes between them? Or are the supposedly truthful and deceptive confessions so very alike simply because the respondents did not comply with the instruction to tell one false and one true confession? Differently put, maybe there are no differences because the within-subjects design failed. Unfortunately, none of these explanations, not even the latter one, can be ruled out. Only more research can give us enough evidence to estimate the probability of different explanations.

However, let us for a while pretend we knew for a fact that the respondents actually did follow the instructions; they did tell one truthful confession of a crime they had committed and been sentenced for, and one false confession of a crime they had never committed and never been sentenced for. Given this starting point we can draw the following additional conclusions: (1) the offenders either told very convincing lies *or* they told very unconvincing truths; (2) differences between truthful and deceptive confessions were statistically detectable only for three different measures of amount and quality of detail out of 28 separate measures (30 if also counting the total scores for RM and CBCA).

In the original, published report for Study I, the possibility of the offenders telling very convincing lies was in focus, while the contrasting possibility never really occurred to us. What if the offenders rather told

unconvincing truths? To bring more clarity into this, we can turn to a study (Strömwall & Willén, 2011) conducted with the same offenders on the same occasion as the data collection for Study I. Strömwall and Willén asked the offenders what they did to convince the interviewer that their deceptive confession actually was true. The single most commonly reported answer to this question was to keep the made-up story close to truly experienced events. A few even said outright that they aimed to keep both statements as similar to each other as possible, which gives some support to the above suggested explanation that the lack of differences were an artifact of the employed research design. However, an implication of this is also that the offenders might have deliberately chosen to make the truthful statement seem less convincing in order to hide supposedly deceptive features of their false statement.

Previous research has shown that criminally experienced individuals do tend to hold more accurate beliefs about signs of deception than the general public does (Granhag, Andersson, Strömwall, & Hartwig, 2004), but they also tend to agree that deception really is a cognitively demanding task (Strömwall & Willén, 2011). In addition, the offenders were very well aware of the fact that suspects are not the ones responsible for proving their innocence – it is up to the police officer to prove their guilt (Strömwall & Willén, 2011). Altogether, it may be a fruitful strategy – in real life as well as during Study I – to keep truthful statements less convincing (e.g., by keeping it succinct and less forthcoming) instead of wasting energy on the more demanding task of telling a convincing lie. This explanation is supported by an earlier study surveying criminally experienced and criminally naive respondents about which lie strategies they would use during a hypothetical police interview (Granhag, Clemens & Strömwall, 2009). It was found that the criminally experienced respondents provided significantly less information than the naive respondents, despite the fact that they (arguably) seemed less credible when obviously not cooperating. If these type of strategies are commonly employed by (criminally experienced) suspects and witnesses in real-life investigations, we might need to revise current psychological theories on strategies employed by innocent versus guilty suspects (Granhag & Hartwig, 2008). Unfortunately, it is not very likely that we will learn more about these issues in the near future since it is quite difficult to recruit real offenders to test a theoretical framework on suspects' strategies during police interviews.

Differences between truthful and deceptive confessions were statistically detectable only for three – out of 28 – measures (i.e., individual criteria included in CBCA and RM) of amount and quality of detail. These differences may have been a result of chance alone. Notably, inferences

drawn from this type of study tend to be that “there were no differences in amount or quality of detail between statements,” or the opposite, that “there were significant differences in amount or quality of detail.” But what if we had instead employed the AMT categorization as the outcome measure? What if we had instead used any other measure of details? Because details can be measured in a variety of ways, comparisons of findings across studies are difficult. This is one methodological weakness in previous and current research on investigative interviewing that needs to be addressed (more on this in the next section).

In legal practice, statement analysis is commonly used when assessing the veracity of plaintiffs' (Köhnken, 2004) and asylum seekers' statements (Kagan, 2003), but less often to assess statements by suspects. There may be judicial reasons for not doing so (suspects are not bound to speak the truth; Gregow, 1996), but there are certainly no psychological reasons to not also conduct a proper statement analysis of a suspect's claims. On the contrary, independent assessment of the statements from both parties would in the best case increase the amount of evidence pointing in a certain direction. In the worst case, it would result in contradictory results. In any case, it would certainly decrease the error rate.

Research Designs for the Future

Research on investigative interviewing typically employs a mock-crime paradigm, a role play where the typical participants are students or children pretending to be witnesses or suspects in a legal investigation. Obviously, this type of research does not score high on ecological validity. Study I and Study II contributed to the previous literature on investigative interviewing by employing designs with higher ecological validity, and more such research is definitely needed.

An additional concern with the mock-crime paradigm is that it is not standardized. Sometimes the “crime” concerns wiping out something from a blackboard (e.g., Vrij, Akehurst, Soukara, & Bull, 2002), sometimes it concerns a missing book from a store or library (e.g., Granhag, Strömwall, Willén, & Hartwig, 2013). There is little known, for example, about how strong the manipulation needs to be to generate measurable effects. This in turn – as is also the case with the lacking standardization alone – has the negative side effect of complicating calculations of the required sample size for future studies; we rarely know what to expect from our chosen study design. Lack of standardization also complicates comparisons of findings across studies.

As previously discussed, similar difficulties are faced in research that measures amount and quality of detail in autobiographical memory reports. Although AMT categorization, as conducted for Studies II and III, has the potential of being standardized (Griffith et al., 2012), no standardized instruments to measure amount and quality of detail in memory reports seem to exist at the moment. Future research should be conducted focusing specifically on developing a standardized measure of amount and quality of detail. Such research should consider scientific needs as well as practical significance. For example, by developing a measure that can detect minor differences while also being a measure that is relevant to legal practice where differences in detail currently are assessed without instruments. Crucially, all results from such research should be made freely available in an online repository.

Ideally, a standardized measure like this should also include guidelines on information to consider (and provide) when preregistering research that measures amount and quality of detail in memory reports. A current practice that seems common is to divide the memory reports into different units (e.g., free recall and answers to specific questions) and analyze them separately. Future guidelines should also recommend that a thorough analysis plan is provided (prior to the data collection beginning) with the preregistration to avoid these types of undisclosed questionable research practices. These are critical and urgent matters to deal with if we are serious about conducting effective and high-quality research that really contributes to knowledge accumulation and better legal practices. This is also important to deal with to ensure that forensic and experimental psychologists provide well-founded recommendations to legal practitioners and decision makers.

Limitations

One of the main strengths of Study I is also one of its main weaknesses. We did achieve quite a high ecological validity, but in this particular case it was on the cost of establishing objective truth. The results must therefore be interpreted with caution until more controlled research, preferably with larger samples, can replicate them.

Two important limitations of Studies II and III concern the design and sample size. While Study II was mainly confirmatory in its nature, the research questions in Study III were formed after the data collection and the study is thereby exploratory. Consequently, the grounds for drawing conclusions about the findings could be viewed as somewhat stronger in Study II than Study III. However, Study II suffered from low-observed power

(about 35%; the same low power that is observed in most psychological research, Bakker, van Dijk, & Wicherts, 2012) and the observed effects may therefore be inflated (Button et al., 2013). In addition, both studies were partly based on data with large variations. This was especially the case for types of memories and number of experienced events. It is thus important to consider these limitations when interpreting the findings and await future (high-powered and preferably preregistered) studies to see whether or not the results are replicated. Future research should specifically aim at recruiting respondents who have experienced a very large number of events, at least 30 visits, to enable in-depth analyses on the effects of number of experienced events.

Furthermore, as implicitly stated in an earlier section, *Disclosure statement*, it should be noted that neither Study I nor Study II were reported in a completely transparent fashion in the original publications. Clearly, there was additional information about the procedures that could have been reported in the articles. This fact can be viewed from different perspectives. One possible conclusion is that the conducted research must have been pretty flawed to motivate the omission or withholding of methodological and/or statistical information from the original report. A different conclusion could be that the conducted studies were not flawed and that the decisions to omit information was correct, for instance, because of word limits practiced by journals. Readers agreeing with this second perspective might in turn argue that it was the wrong decision to disclose the omitted information in the present thesis because it might wrongfully give the impression that the conducted research was more flawed than most research.

A third potential perspective is that times are currently changing. Psychological science is in a transforming state: between old (outdated and closed) practices and new (honest and open) ways. As Brian Nosek and Chris Chambers (2015) put it: “[F]uture generations [will] look on the term ‘open science’ as a tautology – a throwback from an era before science woke up. ‘Open science’ will simply become known as science, and the closed, secretive practices that define our current culture will seem as primitive to them as alchemy is to us.” Employing this perspective, one will soon realize that active researchers who were taught in the old-school ways could easily get stuck in the middle, being blamed for realizing too late that they were wrongly taught. Discouraged by the risk of facing blame, one might prefer to not disclose old sins nor embrace the new era, and instead hang on to the familiar practices. I cannot emphasize enough the importance of this not becoming reality, not only because it would hinder or at least delay the transformation process toward a better science, but also because it is not right. It is easy to be the one knowing more, knowing first, knowing better.

Less easy to admit wrongdoing. Less easy to disclose hidden weaknesses of a lifetime work. Open science is connected to the open source movement in which kindness is an integrated part. Let us not forget the kindness when our science opens up.

Conclusions

Suspects' accounts are rarely assessed in the same methodological manner as is done with statements by plaintiffs (Köhnken, 2004). This is unfortunate because independent assessments of both parties' statements may double the amount of evidence pointing in a certain direction, or, at worst, result in contradictory results. In any case, such a procedure would decrease the error rate. Study I tested the effectiveness of such assessments on true and false confessions made by cooperative interviewees. Contrary to the expected, the results showed very few differences between the truths and lies. However, some differences were identified to show that truthful and deceptive statements by cooperative suspects may be successfully distinguished in the future.

Research on the particularization of adults' memories for repeated events is still in its infancy. There are, however, many possibilities. For instance, established findings from studies with children (e.g., Brubacher et al., 2014) could be integrated, and the usefulness of memory specificity training (Raes et al., 2009) that is developed for clinical populations could be tested in new settings. Furthermore, the results from Study III suggested that we may be able to develop more effective mnemonics if we distinguish between recollection of individual events and recollection of details about the events. Study II indicated that derived cues may be one valuable mnemonic for increasing the recollection of individual events. However, there is a need to investigate why the two specificity measures are affected differently. An answer to this question could inspire the development of even more effective techniques.

The current thesis contributes to previous research on investigative interviewing in at least three ways. First, the studies' had a high ecological validity, which is rather unusual in the field of investigative interviewing where amount and quality of detail is studied. Second, the raw data from two of these studies have been made freely available online, making it possible for other researchers to delve deeper into the data. Third, the lack of standardization in experimental psychologists' measures of amount and quality of detail (in memory reports) is raised as an urgent issue.

In addition, the current thesis contributes to the ongoing methodological revolution in psychological science (De Groot, 1956/2014) in at least three ways. First, by practicing transparent and honest reporting throughout, including a disclosure statement for each study. Second, by publishing the raw data for Studies II and III. Third, by pointing out that counting details in memory reports today is characterized by non-standardized procedures with a jungle of possible measures and analyses to choose from. The lack of standardization and the tradition of dividing narratives into separate units (e.g., free recall and answers to specific questions) results in many opportunities for biased decisions and questionable practices. This needs to be considered whenever asked whether or not we can count on the details.

References

- Alonso-Quecuty, M.L. (1992). Deception detection and reality monitoring: a new answer to an old question? In F. Lösel, D. Bender, & T. Bliesener (Ed:s), *Psychology and Law: International perspectives* (pp. 328–332). Berlin: Walter de Gruyter.
- Alpert, J. L., Brown, L. S., & Courtois, C. A. (1998a). Symptomatic clients and memories of childhood abuse: What the trauma and child sexual abuse literature tells us. *Psychology, Public Policy, and Law*, 4, 941–995.
- Alpert, J. L., Brown, L. S., & Courtois, C. A. (1998b). Reply to Ornstein, Ceci, and Loftus (1998): The politics of memory. *Psychology, Public Policy, and Law*, 4, 1011–1024.
- Amado, B. G., Arce, R., & Fariña, F. (2015). Undeutsch hypothesis and Criteria Based Content Analysis: A meta-analytic review. *The European Journal of Psychology Applied to Legal Context*, 7, 1–10.
- Amado, B. G., Arce, R., Fariña, F., & Vilariño, M. (2016; in press). Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*. <http://dx.doi.org/10.1016/j.ijchp.2016.01.002>
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th ed.* Arlington, VA: American Psychiatric Association.
- Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2009). *Memory*. Psychology Press; East Sussex.
- Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, 7, 543–554.
- Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.
- Brainerd, C. J., & Reyna, V. F. (2004). Fuzzy-trace theory and memory development. *Developmental Review*, 24, 396–439.
- Brubacher, S. P., Powell, M. B., & Roberts, K. P. (2014). Recommendations for interviewing children about repeated experiences. *Psychology, Public Policy, and Law*. DOI: 10.1037/law0000011
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14, 365–376.
- Christianson, S. Å. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin*, 112, 284–309.

- Cohen, G., & Java, R. (1995). Memory for medical history: Accuracy of recall. *Applied Cognitive Psychology*, 9, 273–288.
- Connolly, D. A., & Gordon, H. M. (2014). Can order of general and specific memory prompts help children to recall an instance of a repeated event that was different from the others? *Psychology, Crime & Law*, DOI: 10.1080/1068316X.2014.885969
- Connolly, D. A., Price, H. L., Lavoie, J. A. A., & Gordon, H. M. (2008). Perceptions and predictors of children’s credibility of a unique event and an instance of a repeated event. *Law and Human Behavior*, 32, 92–112.
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185.
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, 169, 323–338.
- De Groot, A. D. (1956/2014). The meaning of “significance” for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, and Han LJ van der Maas]. *Acta Psychologica*, 148, 188–194.
- Debeer, E., Raes, F., Williams, J. M. G., Craeynest, M., & Hermans, D. (2014). Operant conditioning of autobiographical memory retrieval. *Memory*, 22, 171–183.
- Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior*, 28, 687–706.
- Fisher, R. P. (1981). Interaction between encoding distinctiveness and test conditions. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 306–310.
- Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview*. Charles C. Thomas, Publisher; Springfield.
- Fivush, R. (2011). The development of autobiographical memory. *Annual Review of Psychology*, 62, 559–582.
- Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*, 70, 401–412.
- Gosse, L. L., & Roberts, K. P. (2014). Children’s use of a ‘time line’ to indicate when events occurred. *Journal of Police and Criminal Psychology*, 29, 36–43.
- Granhag, P.A., Andersson, L.O., Strömwall, L.A., & Hartwig, M. (2004). Imprisoned knowledge: Criminals’ beliefs about deception. *Legal and Criminological Psychology*, 9, 103–119.

- Granhag, P. A., Clemens, F., & Strömwall, L. A. (2009). The usual and the unusual suspects: level of suspicion and counter-interrogation tactics. *Journal of Investigative Psychology and Offender Profiling*, 6, 129–137.
- Granhag, P. A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. *Psychology, Crime & Law*, 14, 189–200.
- Granhag, P. A., Strömwall, L. A., Willén, R. M., & Hartwig, M. (2013). Eliciting cues to deception by tactical disclosure of evidence: The first test of the Evidence Framing Matrix. *Legal and Criminological Psychology*, 18, 341–355.
- Gregow, T. (1996). Några synpunkter på frågan om bevisprövning och bevisvärdering i mål om sexuella övergrepp mot barn [Some standpoints on the issues of evidence testing and evidentiary assessment in cases concerning sexual abuse]. *Svensk Juristtidning [Swedish Law Review]*, 7, 509–523.
- Griffith, J. W., Sumner, J. A., Raes, F., Barnhofer, T., Debeer, E., & Hermans, D. (2012). Current psychometric and methodological issues in the measurement of overgeneral autobiographical memory. *Journal of Behavior Therapy and Experimental Psychiatry*, 43, S21–S31.
- Gudjonsson, G. H., & Sigurdsson, J. F. (1994). How frequently do false confessions occur? An empirical study among prison inmates. *Psychology, Crime and Law*, 1, 21–26.
- Gyulai, G., Kagan, M., Herlihy, J., Turner, S., Hárdi, L., & Udvarhelyi, E. T. (2013). *Credibility assessment in asylum procedures – A multidisciplinary training manual*. Budapest, Hungary; Hungarian Helsinki Committee. ISBN: 978-615-5215-12-4
- Hargus, E., Crane, C., Barnhofer, T., & Williams, J. M. G. (2010). Effects of mindfulness on meta-awareness and specificity of describing prodromal symptoms in suicidal depression. *Emotion*, 10, 34–42.
- Heeren, A., Van Broeck, N., & Philippot, P. (2009). The effects of mindfulness on executive processes and autobiographical memory specificity. *Behaviour Research and Therapy*, 47, 403–409.
- Innocence Project. (2016). False confessions or admissions. Retrieved on March 23, 2016, from <http://www.innocenceproject.org/causes-wrongful-conviction/false-confessions-or-admissions>
- Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, 117, 371–376.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3–28.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67–85.

- Kagan, M. (2003). Is truth in the eye of the beholder? Objective credibility assessment in refugee status determination. *Georgetown Immigration Law Journal*, 17, 367–81.
- Kassin, S. M. (2015). The Social Psychology of False Confessions. *Social Issues and Policy Review*, 9, 25–51.
- Kassin, S. M., Drizin, S. A., Grisso, T., Gudjonsson, G. H., Leo, R. A., & Redlich, A. D. (2010). Police-induced confessions: Risk factors and recommendations. *Law & Human Behavior*, 34, 3–38.
- Keibell, M., & Milne, R. (1998). Police officers' perception of eyewitness factors in forensic investigations. *Journal of Social Psychology*, 138, 323–330.
- Kemp, A., Rawlings, E. I., & Green, B. L. (1991). Post-traumatic stress disorder (PTSD) in battered women: A shelter example. *Journal of Traumatic Stress*, 4, 137–148.
- Kolers, P. A., & Roediger III, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior*, 23, 425–449.
- Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology*, 51, 481–538.
- Köhnken, G. (2004). Statement Validity Analysis and the 'detection of the truth'. In P.A. Granhag & L.A. Strömwall (Ed:s.), *Detection of deception in forensic contexts*, (pp. 41–63). Cambridge: Cambridge University Press.
- La Rooy, D., Brubacher, S. P., Aromäki-Stratos, A., Cyr, M., Hershkowitz, I., Korkman, J., ... & Stewart, H. (2015). The NICHD protocol: a review of an internationally-used evidence-based tool for training child forensic interviewers. *Journal of Criminological Research, Policy and Practice*, 1, 76–89.
- Leins, D. A., Fisher, R. P., Pludwinski, L., Rivard, J., & Robertson, B. (2014). Interview protocols to facilitate human intelligence sources' recollections of meetings. *Applied Cognitive Psychology*, 28, 926–935. DOI: 10.1002/acp.3041
- Linton, M. (1975). Memory for real-world events. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in Cognition* (pp. 376–404). San Fransisco: Freeman.
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12, 361–366.
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13, 585–589.
- Maestas, K. L., & Rude, S. S. (2012). The benefits of expressive writing on autobiographical memory specificity: A randomized controlled trial. *Cognitive Therapy and Research*, 36, 234–246.

- McNally, R. J. (2005). *Remembering trauma*. Harvard University Press; London, England.
- Means, B., & Loftus, E. (1991). When personal history repeats itself: Decomposing memories for recurring events. *Applied Cognitive Psychology*, 5, 297–318.
- Means, B., Nigam, A., Zarrow, M., Loftus, E. F. & Donaldson, M. S. (1989). Autobiographical memory for health-related events. National Center for Health Statistics. *Vital and Health Statistics*, 6, 1–37.
- Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law*, 16, 340–372.
- Mogoșe, C., Brăilean, A., & David, D. (2013). Can concreteness training alone reduce depressive symptoms? A randomized pilot study using an internet-delivered protocol. *Cognitive Therapy and Research*, 37, 704–712.
- Moore, S. A., & Zoellner, L. A. (2007). Overgeneral autobiographical memory and traumatic events: an evaluative review. *Psychological Bulletin*, 133, 419–437.
- Moradi, A. R., Moshirpanahi, S., Parhon, H., Mirzaei, J., Dalglish, T., & Jobson, L. (2014). A pilot randomized controlled trial investigating the efficacy of MEMORY Specificity Training in improving symptoms of posttraumatic stress disorder. *Behaviour Research and Therapy*, 56, 68–74.
- Neath, I., & Surprenant, A. M. (2008). Short- vs. long-term memory. In A. S. Benjamin, J. S. de Belle, B. Etnyre, & T. Polk (Eds.), *Human learning: Biology, brain, and neuroscience* (pp. 21–31). Amsterdam, Elsevier.
- Neisser, U. (1981). John Dean's memory: A case study. *Cognition*, 9, 1–22.
- Neshat-Doost, H. T., Dalglish, T., Yule, W., Kalantari, M., Ahmadi, S. J., Dyregrov, A., & Jobson, L. (2013). Enhancing autobiographical memory specificity through cognitive training: An intervention for depression translated from basic science. *Clinical Psychological Science*, 1, 84–92.
- Nosek, B., & Chambers, C. (2015). *The first imperative: Science that isn't transparent isn't science*. The Guardian, 2015-06-25. Retrieved on November 6, 2015, from: <http://www.theguardian.com/science/headquarters/2015/jun/25/the-first-imperative-science-that-isnt-transparent-isnt-science>
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7, 615–631.
- Ono, M., Devilly, G. J., & Shum, D. H. K. (2016). A meta-analytic review of overgeneral memory: The role of trauma history, mood, and the presence of posttraumatic stress disorder. *Psychological Trauma: Theory, Research, Practice, and Policy*, 8, 157–164. <http://dx.doi.org/10.1037/tra0000027>

- Orbach, Y., Lamb, M. E., Sternberg, K. J., Williams, J. M. G., & Dawud-Noursi, S. (2001). The effect of being a victim or witness of family violence on the retrieval of autobiographical memories. *Child Abuse & Neglect*, *25*, 1427–1437.
- Ornstein, P. A., Ceci, S. J., & Loftus, E. F. (1998). Comment on Alpert, Brown, and Courtois (1998): The science of memory and the practice of psychotherapy. *Psychology, Public Policy, and Law*, *4*, 996–1010.
- Ost, J., Granhag, P. A., Udell, J., & Roos af Hjelmsäter, E. (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*, 76–85.
- Otgaar, H., Scoboria, A., & Mazzoni, G. (2014). On the existence and implications of nonbelieved memories. *Current Directions in Psychological Science*, *23*, 349–354.
- Philips, & Fisher, R. P. (1998). *Enhancing cooperative suspects' memories of crime: A cued recall approach*. Poster presented at the 106th annual convention of the American Psychological Association, San Francisco, USA.
- Phillips, S., & Williams, J. M. G. (1997). Cognitive impairment, depression and the specificity of autobiographical memory in the elderly. *British Journal of Clinical Psychology*, *36*, 341–347.
- Powell, M. B., Roberts, K. P., & Guadagno, B. (2007). Particularisation of child abuse offences: common problems when questioning child witnesses. *Current Issues in Criminal Justice*, *19*, 64–74.
- Rabinowitz, J. C., Fergus, I. M., Ackerman, B. P. (1982). A processing resource account of age differences in recall. *Canadian Journal of Psychology*, *36*, 325–344.
- Raes, F., Hermans, D., Williams, J. M. G., & Eelen, P. (2007). A sentence completion procedure as an alternative to the Autobiographical Memory Test for assessing overgeneral memory in non-clinical populations. *Memory*, *15*, 495–507.
- Raes, F., Williams, J. M. G., & Hermans, D. (2009). Reducing cognitive vulnerability to depression: A preliminary investigation of Memory Specificity Training (MEST) in inpatients with depressive symptomatology. *Journal of Behavior Therapy and Experimental Psychiatry*, *40*, 24–38.
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*, 1–75.
- Ricarte, J. J., Hernández-Viadel, J. V., Latorre, J. M., & Ros, L. (2012). Effects of event-specific memory training on autobiographical memory retrieval and depressive symptoms in schizophrenic patients. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*, S12–S20.
- Ricarte, J. J., Hernández-Viadel, J. V., Latorre, J. M., Ros, L., & Serrano, J. P. (2014). Effects of specific positive events training on autobiographical

- memories in people with schizophrenia. *Cognitive Therapy and Research*, 38, 407–415. DOI: 10.1007/s10608-014-9610-3
- Rivard, J. R., Fisher, R. P., Robertson, B., & Hirn Mueller, D. (2014). Testing the Cognitive Interview with professional interviewers: Enhancing recall of specific details of recurring events. *Applied Cognitive Psychology*. DOI: 10.1002/acp.3026
- Roberts, K. P., & Powell, M. B. (2001). Describing individual incidents of sexual abuse: a review of research on the effects of multiple sources of information on children's reports. *Child Abuse & Neglect*, 25, 1643–1659.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 803–814.
- Sartory, G., Cwik, J., Knuppertz, H., Schürholt, B., Lebens, M., Seitz, R. J., & Schulze, R. (2013). In search of the trauma memory: A meta-analysis of functional neuroimaging studies of symptom provocation in posttraumatic stress disorder (PTSD). *PLoS One*, 8, e58150. doi:10.1371/journal.pone.0058150
- Schank, R. C., & Abelson, A. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Lawrence Erlbaum.
- Schneider, L., Price, H. L., Roberts, K. P., & Hedrick, A. M. (2011). Children's episodic and generic reports of alleged abuse. *Applied Cognitive Psychology*, 25, 862–870.
- Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 171–181.
- Scoboria, A., Memon, A., Trang, H., & Frey, M. (2013). Improving responding to questioning using a brief retrieval training. *Journal of Applied Research in Memory and Cognition*, 2, 210–215.
- Serrano, J. P., Latorre, J. M., & Gatz, M. (2007). Autobiographical memory in older adults with and without depressive symptoms. *International Journal of Clinical and Health Psychology*, 7, 41–57.
- Serrano, S. J., Latorre, P. J., Ros, S. L., Navarro, B. B., Aguilar, C. M., Nieto, L. M., ... & Gatz, M. (2012). Life review therapy using autobiographical retrieval practice for older adults with clinical depression. *Psicothema*, 24, 224–229.
- Sharman, S. J., Powell, M. B., & Roberts, K. P. (2011). Children's ability to estimate the frequency of single and repeated events. *International Journal of Police Science & Management*, 13, 234–242.
- Shiffrin, R. M. (1993). Short-term memory: A brief commentary. *Memory & Cognition*, 21, 193–197.

- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366.
- Sjöden, B., Granhag, P. A., Ost, J., & Roos Af Hjelmsäter, E. (2009). Is the truth in the details? Extended narratives help distinguishing false “memories” from false “reports”. *Scandinavian Journal of Psychology*, 50, 203–210.
- Sorenson, S., & Golding, J. M. (1990). Depressive sequelae of recent criminal victimization. *Journal of Traumatic Stress*, 3, 337–350.
- Sporer, S. L. (2012). Making the subjective objective? Computer-assisted quantification of qualitative content cues to deception. In *Proceedings of the Workshop on Computational Approaches to Deception Detection* (pp. 78–85). Association for Computational Linguistics.
- Suengas, A. G., & Johnson, M. K. (1988). Qualitative effects of rehearsal on memories for perceived and imagined complex events. *Journal of Experimental Psychology: General*, 117, 377–389.
- Sumner, J. A., Mineka, S., Zinbarg, R. E., Craske, M. G., Vrshek-Schallhorn, S., & Epstein, A. (2014). Examining the long-term stability of overgeneral autobiographical memory. *Memory*, 22, 163–170.
- Surprenant, A. M., & Neath, I. (2009). *Principles of Memory*. Hove, East Sussex: Psychology Press.
- Strömwall, L.A. (2010). Assessing reliability by analyzing the verbal content: The case of Sweden. In P.A. Granhag (Ed.), *Forensic psychology in context: Nordic and International Approaches* (pp. 264–280). Cullompton: Willan Publishing.
- Strömwall, L. A., & Willén, R. M. (2011). Inside criminal minds: Offenders' strategies when lying. *Journal of Investigative Psychology and Offender Profiling*, 8, 271–281.
- Thompson, C.P., & Mingay, D. (1991). Estimating the frequency of everyday events. *Applied Cognitive Psychology*, 5, 497–510.
- Tulving, E. (1983). *Elements of Episodic Memory*. Oxford: Clarendon.
- Tulving, E. (1974). Cue-dependent forgetting. *American Scientist*, 62, 74–82.
- Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology*, 53, 1–25.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352–373.
- Undeutsch, U. (1982). Statement reality analysis. In A. Trankell (Ed.), *Reconstructing the past: the role of psychologists in criminal trials* (pp. 27–56). Stockholm: Norstedts.
- UNHCR. (2013). *Beyond proof: Credibility assessment in EU asylum systems*. Brussels: United Nations High Commissioner for Refugees.
- van Vreeswijk, M. F., & de Wilde, E. J. (2004). Autobiographical memory specificity, psychopathology, depressed mood and the use of the

- Autobiographical Memory Test: A meta-analysis. *Behaviour Research and Therapy*, 42, 731–743.
- Vrij, A. (2015). Verbal lie detection tools: Statement Validity Analysis, Reality Monitoring and Scientific Content Analysis. In P. A. Granhag, A. Vrij & B. Verschuere (Eds.), *Detection deception: Current challenges and cognitive approaches* (pp. 3–35). Chichester: John Wiley & Sons.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behavior*, 26, 261–283.
- Wasserstein, R. L., & Lazar, N. A. (2016; in press). The ASA's statement on p-values: context, process, and purpose. *The American Statistician*. DOI: 10.1080/00031305.2016.1154108
- Watkins, E. R., Baeyens, C. B., & Read, R. (2009). Concreteness training reduces dysphoria: proof-of-principle for repeated cognitive bias modification in depression. *Journal of Abnormal Psychology*, 118, 55–64.
- Watkins, E. R., & Moberly, N. J. (2009). Concreteness training reduces dysphoria: A pilot proof-of-principle study. *Behaviour Research and Therapy*, 47, 48–53.
- Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, 1, 442–452.
- Willén, R. M., & Strömwall, L. A. (2012). Offenders' lies and truths: an evaluation of the Supreme Court of Sweden's criteria for credibility assessment. *Psychology, Crime & Law*, 18, 745–758.
- Williams, J. M. G., & Broadbent, K. (1986). Autobiographical memory in suicide attempters. *Journal of Abnormal Psychology*, 95, 144–149.
- Williams, J. M. G., Barnhofer, T., Crane, C., Hermans, D., Raes, F., Watkins, E., & Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin*, 133, 122–148.
- Williams, J. M. G., Teasdale, J. D., Segal, Z. V., & Soulsby, J. (2000). Mindfulness-based cognitive therapy reduces overgeneral autobiographical memory in formerly depressed patients. *Journal of Abnormal Psychology*, 109, 150–155.
- Zlomuzica, A., Dere, D., Machulska, A., Adolph, D., Dere, E., & Margraf, J. (2014). Episodic memories in anxiety disorders: Clinical implications. *Frontiers in Behavioral Neuroscience*, 8, 131. DOI: 10.3389/fnbeh.2014.00131

