



HAL
open science

False memories in working memory tasks: a developmental perspective

Manon Rousselle

► **To cite this version:**

Manon Rousselle. False memories in working memory tasks: a developmental perspective. Psychology. Aix-marseille University, 2023. English. NNT: . tel-04147566

HAL Id: tel-04147566

<https://theses.hal.science/tel-04147566v1>

Submitted on 30 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 19 Juin 2023 par

Manon ROUSSELLE--JACQUIER

False memories in working memory tasks: a developmental perspective

Discipline

Psychologie

École doctorale

Cognition, Langage, Éducation. ED 356

Laboratoire de recherche

Laboratoire de Psychologie Cognitive,
UMR 7290

Composition du jury

Dr. Candice Morey	Rapporteuse
Université de Cardiff	
Dr. Gaën Plancher	Rapporteuse
Université Lyon 2	
Pr. Charles Brainerd	Examineur
Université de Cornell	
Pr. Patrick Lemaire	Examineur
Aix-Marseille Université	Président du jury
Pr. Agnès BLAYE	Directrice de thèse
Aix-Marseille Université	
Dr. Marlène ABADIE	Co-directrice de thèse
Aix-Marseille Université	

Affidavit

Je soussignée, Manon Rousselle--Jacquier, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Agnès Blaye et Marlène Abadie, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat. Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Marseille, 14 avril 2023



List of publications and conference participations

1) Liste des publications réalisées dans le cadre du projet de thèse :

1. Rousselle, M., Abadie, M., Blaye, A., & Camos, V. (2023). Children's gist-based false memory in working memory tasks. *Developmental Psychology*, 59(2), 272–284. <https://doi.org/10.1037/dev0001476>
2. Rousselle, M., Blaye, A., & M., Abadie (in prep.). The development of false memory in working memory tasks: implication of maintenance mechanisms at 5- and 8-year-old.
3. Rousselle, M., Blaye, A., & Abadie, M. (in prep.). The role of semantic similarity and working memory maintenance mechanisms in short-term false memories in children and adults.

2) Participation aux conférences au cours de la période de thèse :

1. Rousselle, M., Blaye, A., & Abadie, M. (2022, August). The role of long-term and working memory in false memories: a comparison between children and adults. In European Society of Cognitive Psychology, Lille, France (poster presentation).
2. Rousselle, M., Abadie, M., Bouffier, M., Mariz Elsig, S., & Camos, V. (2022, August). Do maintenance mechanisms modulate frequency effect in working memory? In European Society of Cognitive Psychology Congress, Lille (oral presentation).
3. Rousselle, M., Abadie, M., Camos, V., Blaye, A. (2022, March). False memories in working memory tasks: a developmental perspective. In Swiss Working Memory Meeting, Fribourg, Switzerland (oral communication).
4. Rousselle, M., Abadie, M., Camos, V., Blaye, A. (2021, October) *Les faux souvenirs à court terme chez l'enfant : Implication de la mémoire de travail ?* In Groupe De Recherche Interdisciplinaire Mémoire Congress, Toulouse (poster presentation).
5. Rousselle, M., Abadie, M. (March, 2021). *Les faux souvenirs dans les salles de classes*. In Journée Annuelle des Doctorants Meeting, Aix-en-Provence (oral presentation).
6. Rousselle, M., Abadie, M., Camos, V., Blaye, A. (2020, september). *False memories in children's working memory*. In 10th European Working Memory Symposium, Cardiff (oral presentation).

7. Rousselle, M., Abadie, M., Camos, V., Blaye, A. (2020, Novembre). *The role of gist and verbatim memory on false recognition in children's working memory*. In Journée Scientifique des Jeunes Chercheurs Meeting, Lille (oral presentation).

Abstract

False memories have been the object of numerous studies in the past. These studies seem to indicate that processing of the meaning of items (i.e., gist memory) is responsible for such illusion. The increase of gist memory with age was shown to increase false memories with age in long-term memory tasks. A few recent studies have shown that these semantic false memories could occur in mere seconds in adults in working memory (WM) tasks. Some of them showed that maintenance mechanisms of the WM had a role in their occurrence. Indeed, there was evidence that articulatory rehearsal, a verbal maintenance mechanism, prevents false memories in immediate tests. By contrast, attentional refreshing, a domain-general maintenance mechanism was shown to promote false memories in delayed tests. In this thesis, seven experiments were conducted in which a developmental perspective was adopted to study false memories in WM tasks. WM maintenance mechanisms develop with age, as well as gist memory traces. We therefore compared young (4-5-year-olds) and older children (8-year-olds) as well as older children (9-year-olds) and young adults and tested the impact of WM maintenance mechanisms on their production of false memories in immediate and delayed tests. We also measured the contribution of gist memory to false memories. Results indicated that short-term false memories occur in children as young as 4-year-olds. Moreover, errors in immediate recall became predominantly semantic with age. False recognition in immediate and delayed tests did however not increase with age. False recognitions were underpinned by gist memory traces. These traces increased with age. They also lead to different responses at recognition between age groups, suggesting qualitative differences in memory traces between young and older children. Most importantly there was evidence that maintenance mechanisms in WM decreased false memories in recall at least from age 8. Implication for these findings is discussed.

Keywords: False memories, Working memory, Long-term memory, Development.

Résumé

Les faux souvenirs ont fait l'objet de nombreuses études par le passé. Ces études semblent indiquer que le traitement du sens des items (i.e., les traces gist en mémoire), est responsable de cette illusion. Il a été montré que l'augmentation de la mémoire gist avec l'âge était responsable de l'augmentation des faux souvenirs avec l'âge dans les tâches de mémoire à long-terme. Quelques études récentes ont montré que ces faux souvenirs sémantiques pouvaient survenir en quelques secondes chez l'adulte dans des tâches de mémoire de travail. Certaines de ces études ont montré que les mécanismes de maintien en mémoire de travail jouaient un rôle dans leur occurrence. En effet, la répétition articulatoire, un mécanisme de maintien verbal, semble prévenir les faux souvenirs dans les tests immédiats. Par ailleurs, le rafraîchissement attentionnel, un mécanisme de maintien reposant sur des ressources cognitives générales, favoriserait les faux souvenirs dans les tests différés. Dans cette thèse, sept expériences ont été réalisées en adoptant une perspective développementale pour étudier les faux souvenirs en tâche de mémoire de travail. Les mécanismes de maintien se développent avec l'âge ainsi que les traces gist en mémoire. Nous avons donc comparé des jeunes enfants (4-5 ans) et des enfants plus âgés (8 ans) ainsi que des enfants plus âgés (9 ans) et de jeunes adultes. Nous avons testé l'effet des mécanismes de maintien en mémoire de travail sur la production de faux souvenirs en test immédiat et différé. Nous avons également mesuré la contribution de la mémoire gist dans l'apparition des faux souvenirs. Les résultats indiquent que les faux souvenirs à court-terme surviennent chez des enfants dès l'âge de 4 ans. De plus, les erreurs en test de rappel immédiat deviennent sémantiques de façon prédominante avec l'âge. Les fausses reconnaissances étaient sous-tendues par des traces gist en mémoire. Ces traces augmentaient avec l'âge. Elles ont également conduit à différentes réponses au test de reconnaissance entre les groupes d'âges, suggérant des différences qualitatives des traces en mémoire entre de jeunes enfants et des enfants plus âgés. Enfin, les mécanismes de maintien en mémoire de travail diminuaient les faux souvenirs en rappel dès l'âge de 8 ans. Ces résultats et leurs implications sont discutés.

Mots-clés : Faux souvenirs, Mémoire de travail, Mémoire à long-terme, Développement.

Acknowledgements

Je tiens tout d'abord à remercier mes directrices de thèse sans qui ce projet n'aurait pu avoir lieu. Marlène, merci du fond du cœur de l'encadrement que tu m'as donné depuis que nous avons commencé à travailler ensemble lorsque j'étais en master. Merci pour la confiance que tu m'as accordée pour mener ce projet de thèse à bien. Merci pour le climat bienveillant et rassurant que tu as instauré et dans lequel j'ai pu travailler durant ces années. Merci aussi pour ta disponibilité durant ma thèse. Tout ceci a été très précieux pour moi. Merci pour ton éclairage scientifique, que ce soit au niveau théorique et dans les savoirs-faire. Merci de m'avoir guidée dans mon travail ainsi que pour tes précieux retours. Je garderai un ensemble de souvenirs épisodiques très positifs de cette supervision et je serai heureuse d'avoir encore l'opportunité de travailler avec toi.

Agnès, merci d'avoir accepté dès le départ d'être ma co-directrice. Sans vous ce projet n'aurait pas pu avoir lieu. Merci de vos retours toujours éclairés sur la création de protocoles expérimentaux, sur mes écrits ou bien sur les stratégies les plus pertinentes à adopter dans certaines situations. J'ai été honorée d'avoir pu bénéficier de votre expertise dans ces domaines. Par ailleurs, vous avez toujours su m'apporter un soutien dans les moments nécessaires et je vous en remercie.

I would like to thank Gaën Plancher, Candice Morey and Charles Brainerd for being part of my thesis jury. I am very honored that experts in the field of development, working memory and false memories can review my work and bring their expert opinion on it. I am very much looking forward to the discussions we will get to have during my defense.

Je tiens à remercier Isabelle Régner et Pascale Colé qui ont contribué à ma formation par la recherche en m'encadrant sur mon travail de recherche en première année de master.

Merci à tous mes collègues du labo d'avoir instauré une si bonne ambiance de travail. Merci pour les échanges intra et extra labo ! Merci aussi pour les moments de détente. Merci à Léo, Marco, Elisa, Maud, JP, Guillem, Ladislav, Margaux, Léa, Laure, Jacob, Floor. Eva, merci d'avoir été ma co-bural pendant une grande partie de mon doctorat. Merci pour ton accueil au labo ! Tu as clairement contribué à mon intégration au sein du laboratoire et tu as rendu mes journées plus ensoleillées. Merci pour les échanges scientifiques qu'on a pu avoir. Merci aussi

pour ton amitié et ton soutien. Camille, dès que je suis arrivée au labo tu m'as accueillie. Tu m'as fait me sentir à l'aise. Merci pour ta présence tout au long de cette période de vie à laquelle tu as ajouté de la saveur (notamment avec tes petits plats). Merci pour ton amitié et tes conseils ! Je suis très heureuse d'avoir eu la chance de te rencontrer ! Christelle, je suis si heureuse qu'on fasse partie de la même équipe, sans quoi je ne t'aurais jamais rencontrée ! Merci pour tes super tutos Labvanced, pour tes petits scripts magiques qui m'ont permis de faire mes fameuses petites analyses supplémentaires. Mais par-dessus tout, merci d'être la personne que tu es ! Ta gentillesse, ton écoute, tes retours et surtout ton amitié, sont très précieux pour moi ! JB, merci pour ton aide dans le déblocage de certaines étapes de codage de mes manips ! Merci d'avoir toujours été volontaire pour m'aider si j'avais un souci. Merci pour le climat positif que tu apportes à la vie au labo !

Valérie, travailler avec vous a été une expérience très enrichissante et inspirante pour moi. Merci pour votre accueil chaleureux à Fribourg, pour votre partage d'expérience sur le fonctionnement du monde de la recherche. Merci de m'avoir invitée à faire partie d'une équipe de recherche si cohérente, qui soit dans l'échange et le partage. Merci aussi à toute l'équipe de recherche du département de psychologie du développement de l'université de Fribourg. Marion, j'ai beaucoup apprécié travailler avec toi. Ton sens de la perfection a vraiment été une plus-value à notre travail ensemble ! Nos échanges ont toujours été enrichissants. Je me souviendrai par-dessus tout de tes petites blagues bien à toi qui égayaient mes journées. Christophe, merci de m'avoir accueillie à Fribourg et d'avoir si bien contribué à mon intégration, autant dans l'équipe de recherche que dans les activités de montagne ou au bar à la fin d'une journée de travail. Luisa, merci d'être toi ! Merci d'être toujours ouverte et partante pour de nouvelles aventures. Tu as clairement rendu les moments que j'ai passés à Fribourg ensoleillés ! Enfin, merci à Fiona, Marianne, Philippe pour les moments avec vous à Fribourg !

Je tiens à remercier tous les enfants qui ont participé à mes études, leurs parents qui m'ont accordé leur confiance pour que je m'entretienne avec leurs enfants, ainsi que les enseignants et la direction des écoles qui m'ont gentiment accueillie. Merci pour l'intérêt qu'ils ont eu pour mon travail également. J'ai beaucoup apprécié intervenir dans les écoles, cela m'a permis d'être plus proche des problématiques qui peuvent survenir dans l'enseignement, qui peuvent parfois être loin des sphères de la recherche. Avoir l'opportunité d'échanger avec vous m'a permis de me poser certaines questions que je ne me serais peut-être jamais posées autrement. Merci aux inspecteurs qui ont autorisé mon intervention dans les écoles. Je remercie

également les participants adultes, pour la plupart d'entre eux étudiants qui ont volontairement accepté de participer à l'une de mes études.

Merci aux étudiants de masters : Anaëlle, Michaël et Camille, aux étudiantes de licence : Soline, Mally et Mélanie, que j'ai eu la chance de superviser durant ma thèse et qui ont contribué à la récolte des données.

Fanny, merci d'être toi. Merci d'être là, dans ma vie depuis tant de temps. Merci pour ton soutien. Et merci d'avoir contribué à ma thèse en me partageant ton expertise de la mise en page !

Emily and Kai, thank you for your help in correcting my English spelling mistakes. Your help was very precious to me !

Jordan, mon chéri, merci pour ton soutien au quotidien et pour ton respect de mes choix, de la personne que je suis et de mon travail. Merci d'avoir été là jusqu'à la fin et de m'avoir offert ton écoute dans les moments de besoin. Merci aussi d'avoir contribué à ma thèse en m'aidant à créer les fameuses macros qui m'ont permis d'analyser mes données (beaucoup) plus rapidement. Merci de rendre ma vie plus belle et de partager cette passion de la danse qui nous lie.

Enfin, merci à ma famille, ma mère, mon père, mes frères et sœurs, mes beaux-parents, mes grands-parents, mes oncles et tantes, mon parrain et ma marraine, pour leur amour et leur soutien. Merci à mes parents qui m'ont toujours soutenue dans mes choix de vie, qui me donnent leur amour de façon inconditionnelle et qui ont permis que j'en sois là aujourd'hui. Ma petite maman, merci de m'avoir toujours soutenue dans mes choix de vie et de les avoir respectés. Merci de m'avoir toujours rappelée que tu étais fière de moi et de m'avoir toujours donné autant d'amour. Merci infiniment pour toute l'aide précieuse que tu m'as apporté pour la mise en forme de ma thèse. Sur une fin de thèse, lorsque le temps est compté, cela compte beaucoup justement ! Je t'aime ! Papa, merci pour ton soutien tout au long de ma vie et de mes études. Merci d'avoir toujours cherché à entretenir un lien et des moments de partage. J'ai beaucoup appris de toi mais si je devais retenir une chose qui restera toujours, ce serait de « discuter, respecter et partager ».

Contents

Illustrations	XVII
Figures.....	XVII
Tables.....	XVIII
Acronyms	XXI
Preface	3
Theoretical part	7
Chapter 1. Different cases of false memories in the literature	9
1.1 Suggested false memories.....	9
1.2 Self-generated false memories.....	17
Chapter 2. Theories of false memories	35
2.1 Generalities about memory.....	35
2.2 Spreading activation theories.....	38
2.3 Fuzzy Trace Theory.....	43
2.4 Global matching theory	49
2.5 Theoretical confrontation	52
Chapter 3. False memories in working memory.....	61
3.1 Different memory systems.....	61
3.2 The development of working memory	72
3.3 Short- and long-term false memories in working memory tasks.	79
Chapter 4. Overview of experimental chapters	91
4.1 Study 1	91
4.2 Study 2	94
4.3 Study 3	95
Experimental part	101
Chapter 5. Children’s Gist-based False Memory in Working Memory Tasks.	103
5.1 Introduction	103
5.2 Experiment 1	109
5.3 Experiment 2	117
5.4 General Discussion.....	124
5.5 Conclusion.....	128
Chapter 6. The development of false memory in working memory tasks: implication of maintenance mechanisms for 5- and 8-year-olds.	133
6.1 Introduction	133
6.2 Experiment 1	140
6.3 Experiment 2	153
6.4 General Discussion.....	162
6.5 Conclusion.....	165
Chapter 7. The role of semantic similarity and working memory maintenance mechanisms in short-term false memories in children and adults.	169

7.1	Introduction	169
7.2	Experiment 1	175
7.3	Experiment 2	186
7.4	General discussion	194
7.5	Conclusion.....	197
General Discussion		201
Chapter 8. General discussion and perspectives.....		203
8.1	Synthesis and discussion	196
8.2	Contribution of our findings to general questions	204
8.3	Perspectives.....	209
Conclusion		221
References		217

Illustrations

Figures

Figure 1. Illustration of the process of activation of the critical item in the DRM paradigm according to the spreading activation theories.

Figure 2. Example of the process of global matching in DRM tasks.

Figure 3. Depiction of typical long-term, short-term, working memory Brown-Peterson and complex span tasks.

Figure 4. Illustration of the memorial system in the Embedded-Process model.

Figure 5. Architecture of working memory according to the Time-Based Resource Sharing Model.

Figure 6. Illustration of one trial of the Brown-Peterson task in Experiment 1, in Chapter 5.

Figure 7. True and false recognition accuracy as a function of age group and type of concurrent task in Experiment 1, in Chapter 5.

Figure 8. Multinomial Processing Tree for the simplified conjoint recognition paradigm based on Stahl and Klauer (2008).

Figure 9. Illustration of trials in Experiment 2, in Chapter 5.

Figure 10. True and false recognition accuracy as a function of age group and type of concurrent task for immediate and delayed tests in Experiment 2, in Chapter 5.

Figure 11. Illustration of one trial of the complex span task in Experiment 1, for the 5-year-old group, in Chapter 6.

Figure 12. Correct and false recognition as a function of age group and type of concurrent task, in Experiment 1, in Chapter 6.

Figure 13. Correct and false recognition as a function of age group and type of concurrent task, in Experiment 2, in Chapter 6.

Figure 14. Illustration of Experiment 1's procedure for one trial, in Chapter 7.

Figure 15. True and false recognition proportion as a function of age group and word-relatedness concurrent task pace in Experiment 1, in Chapter 7.

Tables

Table 1. Overview of the studies conducted in the experimental part, with the detail of each experiment per study (i.e., age and number of subjects, task used, list length, stimuli, concurrent task, manipulated variables and collected measures).

Table 2. Estimates for the parameters of the simplified conjoint recognition model as a function of age group and type of concurrent task in Experiment 1, in Chapter 5.

Table 3. Estimates for the parameters of the SCR model as a function of age groups and concurrent tasks in both immediate and delayed tests in Experiment 2, in Chapter 5.

Table 4. Mean percentage of correct recall and recall errors (semantic errors, theme recall, phonological errors, intrusions and other errors) and decomposition of subjective judgments (3, 2 and 1^a) for correct recall, semantic and non-semantic errors as a function of age group and type of concurrent task, in Chapter 6.

Table 5. Mean parameter estimates of the SCR model as a function of age and concurrent task type, in Chapter 6.

Table 6. Mean percentage of correct recall and recall errors (semantic errors, phonological errors, intrusions and other errors) and mean percentage of each subjective judgments (3, 2 and 1^a) associated with correct recall, semantic and non-semantic errors as a function of age group and type of concurrent task, in Chapter 6.

Table 7. Mean parameter estimates of the SCR model as a function of age and concurrent task type, in Chapter 6.

Table 8. Mean estimates of each parameter to the SCR model depending on age, concurrent task pace and word-list relatedness, in Chapter 7.

Table 9. Proportion of correct recall and each error type (semantic, phonological, other, intrusions, omissions) as a function of age group, pace of the concurrent task and word-list relatedness for correct recall and semantic errors, in Chapter 7.

Table 10. Rate of subjective judgments associated to correct recall, semantic and non-semantic errors depending on age group, word-list relatedness and pace of the concurrent task, in Chapter 7.

Acronyms

<i>a</i>	probability of guessing “target”
AMT	Activation-Monitoring Theory
<i>b</i>	probability of guessing that an item is either a target or a related probe
BAS	Backward associative strength
CT	concurrent task
DRM	Deese-Roediger and McDermott
FTT	Fuzzy-Trace Theory
GMM	Global matching model
<i>G_r</i>	probability of retrieving a target’s gist trace given a related probe
<i>G_t</i>	probability of retrieving a target’s gist trace given a target probe
LTM	Long-term memory
LOP	Level of processing
<i>Pr</i>	discriminability index
TBRS	Time-Based-Resource-Sharing model
R	related
STM	Short-term memory
Un	unrelated
<i>V_t</i>	probability of retrieving a target’s verbatim trace given a target probe
<i>V_r</i>	probability of retrieving a target’s verbatim trace given a related probe
SCR	Simplified conjoint recognition model
WM	Working memory

“ Reality is an illusion created entirely within the human mind, but it’s the only place you can get a decent cup of coffee.”

Tom Gauld for The New Scientist.

Preface

'Memories' are defined by most common dictionaries as "some things remembered from the past" (Collins English Dictionary; Oxford Languages dictionary; The Britannica dictionary). Notice that this definition is quite vague. Nonetheless, it involves that some information is retrieved from some point in time. In Psychology, memory is commonly defined as "the faculty of encoding, storing, and retrieving information" (Squire, 2009). This definition is quite pragmatic as it is based on the idea that if we are able to remember things from our past, it is because we kept information in some point and place in time, and got it back when we needed to. Metaphorically, this definition can be compared to the pair of socks that are put in a drawer for the day until we need them again, and when we do, we go back to the drawer and fetch them. Retrieving the pair of socks is like retrieving an event of our past. An event relates to a specific episode which happened in time and is temporally and spatially located. It implies that when experiencing a memory, we retrieve information with its context (Arndt, 2010; Tulving & Thompson, 1973). So, the experience of remembering something can be described as contextualized information that we retrieve at some point in time. However, one mistake would be to believe that because we retrieve contextual memories, our memories contain all details of the truly experienced event. Indeed, we sometimes have the subjective experience that we remember precisely all the details of the lived event. This is not always the case of course. Sometimes, we get the subjective feeling that there are some dark spots in our memory. For instance, I can remember that I was at this party last summer, but I can't remember all the people that were there, I am aware that some information is missing from my memory. But sometimes, we can wrongly believe that we precisely remember an event. One striking and common example is when getting home one night, I put my keys on the counter. The next morning the keys weren't there but I can see myself putting them there, replaying the memory. Because I am so sure of my memory, I start accusing the person I am living with of having moved them. But when going to my car, I see the keys on the seat and only then, I realize, my memory failed me. I had a false memory: I remembered a distorted event from the past.

The subject of false memories has been the interest to both the scientific community and the general public. Moviemakers play with this illusion in their scenarios to make the audience hold their breath. It was also the source of scandals in judiciary cases and in therapy sessions,

with testimonies that were found to be influenced by the questioner of a police investigation or by the therapist. In the scientific literature, false memories have been extensively studied in the past century. Researchers took interest in this topic when they realized that false memories could have severe impacts on important decisions. The most mediatized domain was justice, and for good reason. Indeed, a lot of studies have accumulated showing that testimonies in investigations were biased by false memories (Frenda et al., 2011; Kassin et al., 2010, for reviews). In the field of juvenile cases, there is also research on the reliability of children's testimonies, often comparing participants' accuracy across ages from childhood until adulthood (Ceci & Bruck, 1993, for a review). The authors of these studies were interested in how new suggested information could impact testimonies. They showed that younger children's testimonies tended to be less reliable than older children's or adults. However, more recent studies measured self-generated false memories rather than suggested ones and showed an age increase in false memories during childhood until young adulthood (Brainerd & Reyna, 2012, for a review). These results' differences across studies suggest that false memories are impacted by different factors which are highlighted when mobilized across different paradigms. Recent research has taken interest in false memories in working memory tasks (e.g., Atkins & Reuter-Lorenz, 2008; Flegal et al., 2010; Abadie & Camos, 2019). Working memory (WM) is a system that we use every day for varied numerous tasks going from estimating the cost of our basket at the supermarket, taking notes in a class, to driving. It is a system at the heart of our cognition, functioning as a hub for other cognitive resources such as attention, long-term memory, flexibility and inhibition (Logie et al., 2021). In children, WM was shown to be a predictor of performance in math and reading at school (Peng et al., 2016; Peng et al., 2018, respectively; Swanson & Alloway, 2012). There is therefore a huge deal in assessing the development of false memories in WM tasks during childhood.

This thesis aims for one at reviewing the work conducted on false memories and their development with age in different paradigms (i.e. suggested false memories and self-generated ones), with the presentation of moderator variables, the presentation of explicative theories of self-generated false memories which is the focus of this thesis and to summarize the existing literature on a recent topic: false memories in WM tasks.

Second, it aims at answering whether false memories occur in WM tasks in children of different age groups. The few existing studies examining false memories in WM tasks, were done with young adults and showed that false memories occurred in WM tasks (Abadie & Camos, 2019, Abadie et al., in press.; Atkins & Reuter-Lorenz, 2008; Bixter & Daniel, 2013;

Flegel et al., 2010, 2014; Macé & Caza, 2011; Olszewsha et al., 2015). However, it is not clear whether it would also occur in children. Indeed, false memories are primarily based on the ability to form meaningful associations between stimuli and this ability grows with age. It is therefore likely that levels of false memories could be quite low in young children like they were shown to be in classical long-term memory tasks (Anastasi & Rhodes, 2008; Brainerd et al., 2018; Sugrue & Hayne, 2006; Wimmer & Howe, 2009). We will assess whether false memories increase from childhood until early adulthood in WM like it is the case in LTM tasks.

Third, the activation of memory traces about the meaning of the studied stimuli at retrieval (i.e., gist memory) was shown to cause long-term false memories (Brainerd & Reyna, 2002a; Brainerd et al., 2008a; Chang & Brainerd, 2021; Howe et al., 2009; Roediger & McDermott, 2000). We aim at testing whether the same memory traces impact false memories in WM tasks at different ages by using both objective (modeling) and subjective measures (phenomenological experience judgments).

Finally, WM tasks involve specific processes to temporarily maintain information. A few studies showed that maintenance mechanisms impacted false memories at short and longer delays in young adults (Abadie & Camos, 2019; in press.; Atkins et al., 2011; Macé & Caza, 2011). These mechanisms develop with age and become increasingly efficient (Barrouillet & Camos, 2018, for a review). The third aim of this thesis was to assess whether WM maintenance mechanisms can also impact false memories in children at different ages.

In this manuscript I will start with a theoretical part presenting two types of false memories with their associated controversial developmental effects and the existing moderators (Chapter 1). I will then present the explicative theories of false memories with regards to development (Chapter 2). Finally, I will introduce false memories in WM taking some time to explain generalities about WM and its development throughout childhood (Chapter 3). An overview of the empirical part will be given in Chapter 4. Chapter 5, 6 and 7 are dedicated to present our Experimental studies. In Chapter 5, study 1 compares 4- and 8- year-old children doing a recognition-based WM task. In this study we varied the possibility of stimuli maintenance through manipulating the concurrent task of the WM task to assess the effect of maintenance mechanism' on false memories. We use the simplified conjoint recognition model of the fuzzy trace theory (Brainerd et al., 1999; Stahl & Klauer, 2008), to estimate the type of memory traces on which true and false recognition were based, in the different age groups. In Chapter 6, like in study 1, the effect of WM maintenance mechanisms on false memories at

different ages was tested, but this time in recall. In this study we compared 5- and 8-year-old children, who did a WM task followed by immediate recall and delayed recognition. Age groups comparison allowed to test the effect of the development of memory traces on false memories. Qualitative judgment scale accompanied recall to get an insight on children's subjective experience of their memory. Finally, in Chapter 7, we tested the effect of semantic relatedness between studied words, on short-term false memories in a WM task. We also tested the effect of maintenance mechanisms suppression of false memory in 9-year-olds and adults. This study offers a direct comparison between performances in recognition and recall. Finally, I discuss how our findings integrate to previous findings and offer future research perspective.

Theoretical part

Chapter 1. Different cases of false memories in the literature

In this Chapter, I will introduce a first type of false memory, called suggested-false memory, because it is caused by an external suggestion. Two different paradigms highlighted the existence of such phenomena: the misinformation and the implanted paradigm. Suggested false memories were shown to usually be more frequent in younger than in older children or in adults. I will introduce the different explicative theories of this type of false memories and show that they are caused by both social and memory factors. I will then introduce a second type of false memories: self-generated false memories, which are the result of the organization of memory alone. This type of false memories increases with age. The manipulation of different factors such as list length, the type of associations between studied stimuli, adapted lists, stimuli modality variation, the use of ecological material, warning instructions and their effect on the development of false memories will be presented. I will also introduce false memories in regard of emotion by showing the impact of emotional stimuli and of emotional induction on false memories. I will close this chapter by presenting the effect of some inter-individual differences in children - i.e. reading comprehension difficulties, maltreatment and the association of items to oneself or not on false memories.

1.1 Suggested false memories.

On March 22nd 1984, teachers and head of the Mc Martin preschool, Manhattan Beach, California, were charged with counts of child abuse involving 48 children of the school they worked at. To understand how this fact happened, we would have to rewind somewhat in time. This story started with a single allegation from a mother towards her son's teacher, Ray Buckey, in 1983. According to her, her son's painful bowel movement came from the fact that his teacher sodomized him. The teacher was not prosecuted at first due to lack of evidence. However, Judy Jonhson, the mother, informed other school parents. Soon, school children were in interrogating rooms of a clinic, being asked about potential abuse. Coming out from this, 360 children claimed they had been abused by different teachers of the school. Long story short, the case lasted 7 years. All charges were dropped in the end for all prosecuted individuals. During this period, Ray Buckey served 5 years in prison for crimes he was in the end acquitted of; Mc Martin school was destroyed; Judy Jonhson, was diagnosed with paranoid schizophrenia.

Looking back, how could this snowball effect occur? What caused this gigantic ‘*Titanic*’ situation? Well, this is obviously a complex situation which involves different actors and the role of each of them contributed to it. Obviously, the story started with Judy Jonhson, but the torch was passed to other actors such as the social workers who interrogated the children, the media putting in light what they chose to, and the parents of the children reporting abuse in their turn. All of this creating the perfect explosive cocktail. Now let’s take a moment to talk about the children’s interrogation that led to abuse accusations. Michael P. Maloney, clinical psychologist and professor of psychiatry testified during the trial as an expert witness on the conducted interviews. Based on the videotapes of children’s interviews, he concluded that many of the children’s statements in the interviews were generated by the examiner (Eberle & Eberle, 1993). Indeed, the interviewers’ techniques were highly suggestive and invited children to pretend or speculate about supposed events (Garven et al., 1998). These incidents question the legitimacy of the children’s reports in this case but also in other legal cases. Many interrogations have indeed risen from these situations such as: how reliable is a child’s testimony? Are children’s testimonies as reliable at different ages? Are they as reliable as adults? What causes children to give an erroneous testimony?

1.1.1 The misinformation effect

Scientific research intended to reply to these interrogations by recreating false memories inductions in a laboratory context, which gave birth to numerous studies. One of the princeps study was Loftus and Palmer’s (1974), who showed that the speed estimation of a car in a crash watched on video tape, was determined by the type of verb used in a question about the car speed. Indeed, the verb ‘smashed’ increased the speed rate estimation compared to the verb ‘collided’. Moreover, even though there was no broken glass in the tape, 32% of participants for which the verb ‘smashed’ was used, said one week later that there was broken glass in the scene, against 14% with the verb ‘hit’ and 12% in a condition without question in the first place. This study illustrates well how the words used in interrogations can influence the way we remember or at least report events. This effect is called ‘*the misinformation effect*’. In development, the misinformation effect was evidenced by Ceci et al.’s famous study (1987). In their first experiment, 3- to 12-year-olds were told a story about a girl named Loren who had a stomachache. The day after, another person talked to the children and asked them whether they remembered about the story of Loren ‘having a headache’ in one condition or ‘being sick’ in another condition. In the first condition, the experimenter introduced false information in their speech whereas in the second condition, general information was given which did not contradict

the prior story. Two days after this interview, children were asked to pick two images among four that described the original story. Out of the four images, two were wrongful and two depicted true events from the story. In one of the images, a girl was depicted with a stomachache, which matched the original story, and in another image with a headache, which did not match the true story. The results showed that the rate of correct recognition decreased when children were asked about remembering ‘the headache’ rather than about ‘Loren being sick’, mirroring an increased acceptance rate of the suggested information. It shows that the implementation of false information after the initial exposure to the event can modify the way we report it. Besides, in this study, correct recall decreased by 10% in the 10–12-year-old group dropping from 92% to 82% in the misinformation condition whereas in the 3–4-year-old group, correct recall was reduced by more than half in the false information condition, from 82% to 37%. This performance drop difference between age-groups shows that younger children are more likely to give answers validating the post-event suggestion than older children.

Numerous studies have confirmed the existence of the misinformation effect in developmental studies, corroborating Ceci et al.’s findings (see Bjorklund, 2000; Howe, 2000; Otgaar et al., 2018, for reviews). Indeed, Otgaar et al. (2018) showed that among 30 studies published after 1978, looking at the misinformation effect in children of different ages, 87% of them showed that younger children were more suggestible than older children. These types of studies regroup various scenarios such as watching videos of a thief, a petty crime or an argument (Bjorklund et al., 1998, 2000; Bright-Paul et al., 2008; Cassel & Bjorklund, 1995; Cassel et al., 1996; Cohen & Harnick, 1980; Roebbers & Schneider, 2002; Roebbers et al., 2002), camp experiences (Ackil & Zaragoza, 1995), treasure hunting (Roebbers & Schneider, 2005; Roebbers et al., 2007), or about real staged events going from physical examination (Alexander et al., 2002; Eisen et al., 2002; Goodman et al., 1994; Oates & Shrimpton, 1991; Ornstein et al., 1992; Saywitz, et al., 1991) to witnessing or participating to an interaction with an experimenter or a puppet (Flin et al., 1992; Kulkofsky & Klemfuss, 2008; Marin et al., 1979; Otgaar et al., 2010; Rudy & Goodman, 1991; Stern, 1910). They also cover a wide age range going from very young children starting at 2 years old, to young adults. They usually implement the false information by asking misleading questions. For instance, in Bjorklund et al.’s study (1998), children watched a video about two adolescents arguing about the use of a Gameboy video game. The boy then takes the Gameboy from the girl without permission, staging a thief. Later on, when participants are asked questions about the scene, the misleading question group was asked questions such as “The man owned the Gameboy, didn't he?”. Results of this study

show, like it is usually the case in other studies using a similar protocol, that younger children wrongfully confirm the suggestion more than older children, but also that the misleading questions impact their recall or recognition performance in a delayed test, such that they incorporate the suggested information in their statements, more than older children. In fewer studies, the misinformation is embedded in narratives. For instance, in Ackil and Zaragoza's study (1995), 7-year-olds to college students were shown a short video regrouping several scenes of a Disney movie. An experimenter then resumed the footage and introduced for some participants additional false information (e.g., 'after the snake bit one of the ladies in the leg, a character killed the snake with the boat paddle', in this sentence only the second part was shown in the footage). Results of this particular study show that when misinformation is introduced after the story was told, children have a hard time judging its source and they often report that it came from the original story instead of from the later summary. The younger children were, the bigger this effect was.

1.1.2 False memory implantation.

Other studies have shown that young adults could remember whole false events that allegedly happened during childhood, as true. For instance, 40% of a British sample reported having seen nonexistent footage of a bus exploding in a terrorist attack in London with 35% of them describing details from memory (Ost et al., 2008). Confidence in remembering shaking hands with Bugs bunny, which is not a Disney character at Disneyland increased in 80% of participants, after watching a Disney add with this character in it (Braun et al., 2002, Exp. 2). Twenty-two percent of participants remembered receiving a painful enema during childhood, after receiving information that getting this type of procedure at this period of life is plausible (Hart & Schooler, 2006). Twenty-five percent of a subjects' group wrongfully remembered having been lost in a mall when they were 5 years old (Loftus & Pickrell, 1995). When receiving an interview about a false childhood event (e.g., attending a friends' wedding and spilling over a bowl of punch), even though 0% of the participants could describe the event during interview 1, 17% were able to give details about it at interview 2, when being given the title and some keywords about the event. This rate increased to 25% at interview 3 (Hyman et al., 1995). Similarly, when given encouragement and verbal cues, 26% of participants recalled a stressful childhood false event, that they did not remember when first evoked (i.e., being attacked by an animal, undergoing a serious medical procedure, getting lost, being harmed by another child, getting in an outdoor accident) during one of the three following interviews (Porter et al., 1999). Even though these studies were mostly led with adults, whole false events were also reported

in children. For instance, Strange et al. (2006) have shown that 32% of 6-year-olds reported having had tea with prince Charles when they were 2 years of age during an interview occurring 6-8 days after they were shown a fabricated picture representing this event with them in it (interview 3). This percentage decreased to 20% in 10-year-olds, suggesting that the report of false events is more likely in younger children (see also Otgaar et al., 2009, 2010, for similar results). Overall, these studies investigating admission of a false event as true share a common paradigm, called '*the false memory implantation paradigm*' leading in average 25% of participants to report recalling the false event. The false event is usually presented among true events as being reported by the parents. Experimenters lead several interviews where they encourage participants to elaborate on the event by providing some key information about it, giving instructions to imagine the event. Key information can be pictorial representations of the event, giving a general description of the event and specific verbal cues. It seems that like in the misinformation paradigm, false memory rates are the highest in younger children and decrease with age, although, there have been fewer studies looking at developmental trajectories using this paradigm. These studies as well as studies using the misinformation paradigm, question which mechanisms are responsible for the formation of those suggested false memories.

1.1.3 Mechanisms of suggested false memories.

There have been several interpretations and theories to explain false memories in the misinformation paradigm. The first one, is that the suggested information is written over the original one in memory which creates a distorted memory (Loftus et al., 1978, 1985). Alternatively, another explanation would be that the suggested new information would interfere with the original information and inhibits it in memory. So, both traces would still be stored in memory, but the suggested information would be more accessible (Morton et al., 1985). Ultimately, both alternatives agree on the idea that there is a disruption of the original memory by the suggestion (see Bekerian & Bowers, 1983; Loftus & Loftus, 1980, for more precisions about both alternatives). Additionally, Brainerd & Reyna (1998) theorized that it is because we have fuzzy traces of the general meaning of what was perceived in memory that we have false memories in the first place. Notice indeed that in Ceci et al.'s study, the false information (i.e. headache) shares semantic similarities with the original information (i.e. stomachache). This theory was not developed for suggested false memoires specifically, but contributes to understanding how memory functioning leads to false memories. I will develop it further in Chapter 2. Moreover, according to Johnson et al. (1993), when retrieving the suggested

information, participants would produce false memories when they cannot access the source of the information retrieved in memory. For instance, when participants retrieved that Loren had a headache (false information) instead of stomachache (in Ceci et al.'s study), they could have a false memory only if they cannot remember that the information was provided later on by the experimenter. Several studies have shown that source monitoring increase in efficiency with age (Foley, et al., 1983, 1993; Johnson, et al., 1993). This, added to the interference interpretation (Morton et al., 1985) or to the overwriting interpretation (Loftus et al., 1978, 1985), could explain why younger children produce more suggested false memories in misinformation paradigms. Indeed, they could more often misattribute the source of the misinformation and believe that the information was self-generated, especially if the misinformation is more available in memory than the original information. In consequence, participants, especially young children, could produce bias responses to yes/no recognition tests by responding 'yes' more easily to a probe relating to the false information or instead by responding 'no' to the original item because their confidence in the original memory was altered by the suggested information (Zaragoza & McCloskey, 1989). This process is called 'biased guessing'.

Aside from biased guessing, other studies have shown that children, could comply with the experimenter's suggestion because they represented an authority figure, knowingly, a person with more knowledge than them. Ceci et al. (1987) showed this in their second experiment by replacing the adult telling the misinformation by a child, who does not represent an authority figure compared to adults. They showed that the misinformation effect on recognition was strongly reduced when the child gave the misinformation instead of the adult authority figure. Hence, compliance to authority figures can impact responses to a memory test. This could also explain why in some studies, younger children produce more false memories. Indeed, social compliance likely reduces as participants age because their perception of the experimenter as more knowledgeable changes with age. The older they get, the less they will be prone to compliance effects.

Another position regarding the misinformation effect is the one adopted by McCloskey and Zaragoza (1985). According to them, the misinformation effect is not caused by memory distortion of the original event by the suggested event. To show so, they used a modified misinformation paradigm developed by Zaragoza (1991), in which participants during recognition had to choose between the true original information or new information instead of the usually presented misinformation. Their hypothesis was that if the misinformation really

altered the original trace in memory, the rate of correct recognition of the original true information in this paradigm should be lower when participants were provided with misinformation compared to when they were not. Their results showed that correct recognition rate was similar between groups receiving and not receiving the false information. They concluded that memory was not altered by the misinformation effect and that this effect was caused either by pure social compliance or by the forgetting of the original memory independently of the introduction of a suggestion. Those results were however not replicated by Ceci et al.'s study (Exp. 3 and 4), who found a difference of performance between the two conditions, indicating that the suggestion of new information was involved in the misinformation effect (see also Loftus et al., 1985, for a review). In their study, the misinformation effect was however smaller in the modified paradigm compared to the original conditions in which the true information is opposed to the false information during recognition. It suggested that the misinformation competes with the original trace and depending on which probe is presented at test (the misinformation or new information), the original trace receives more or less interference from the trace created by the misinformation. The presentation of the misinformation probe lead to more interference than the new probe. These findings therefore favor the competing traces memory explanation (Morton et al., 1985; see also Holliday & Hayes, 2001 for evidence of traces co-existence in memory) rather than the rewriting trace explanation (Loftus et al., 1978, 1985). In conclusion, the misinformation effect is facilitated by both memory impairment (interference of a suggested information with the original event, combined with source misattribution) and compliance to the experimenter' suggestion. Younger children produce more false memories in this type of paradigm because they are more prone to social compliance effects, and they are more prone to source misattribution.

Regarding implantation paradigms, interpretations share some similarities with the misinformation paradigm but also some differences. The main conceptual difference between paradigms is that implantation is about the creation of a whole memory instead of the modification of a part of the memory usually found in the misinformation paradigms. Methodologically speaking, both paradigms are also quite different. Indeed, in the misinformation paradigm, false information is suggested usually after the perception of first information. On the contrary, in an implantation paradigm, false information of a whole event is directly suggested without relying on a previous true experience. Hence, there is no interference of new information with a direct existing specific memory. In this situation, participants are usually unsure about how the event occurred and make their 'best guess' about

it depending both on the cues given by the experimenter's information but also on scripted knowledge stored in memory (Newman & Lindsay, 2009). Scripted knowledge or schema relates to every past event that has happened to us and are stored in our memory. They constitute a sort of database to predict and imagine future events sharing a similar context (Bartlett, 1932). In the false memory implantation paradigm, because participants have to build up a whole memory event, they rely on those scripts to create a most likely event which respects the constraints imposed by the experimenter. One pioneer study looking at the effect of scripted knowledge on reports was Bartlett's (1932). He showed that when Cambridge denizens participants had to tell back a story, they changed some details for items they were more familiar with. For instance, the story which took place in an American Indian culture comprised a canoe. This element was sometimes replaced by a rowboat during recall which is more typical of the participant's culture. This study illustrates well the effect of previous knowledge on the way information is recalled. It also led to the conclusion that information retrieval is based on a memory reconstruction from different elements in memory rather than from the retrieval of a single block memory. Since then, many studies have found an effect of scripted knowledge, beliefs and stereotypes on the way we recall information (e.g. Bergman & Roediger, 1999; Bower et al., 1979; Bransford & Johnson, 1972; Brewer & Treyens, 1981 ; Conway & Pleydell-Pearce, 2000; Dodson et al., 2008; Frenda et al., 2013; Johnson et al., 1973; Lindsay & Read, 2006 ; Murphy et al., 2019; Ross, 1989 ; Sulin & Dooling, 1974; Sharman et al., 2008; Wilson & Ross, 2003).

In implantation paradigm, the incitation to produce a memory about the false event, was however not equal across studies. Some studies used more suggestive techniques than others. Indeed, some protocols gave only cues about the supposed event (e.g., Loftus & Pickrell, 1995), while others put in addition a sort of pressure on participants to remember the most about the supposed event with incentives such as 'usually other participants remember the event when they take time' (Porter et al., 1999). In the latter case, compliance to the experimenter's suggestion is even more likely. Participants could either be aware that they are complying on purpose to please the experimenter and commit a false report, also called 'known error'. Or, the experimenter could influence them in producing biased guesses. Hughes and Grieve (1980) gave an excellent illustration of the social compliance effect by showing that 5- and 7-year-olds tended to answer the adult's questions by saying 'yes' or 'no', even if they have no meaning, based on what they perceived was the expected answer (see also Pratt, 1991). Moreover, repeated interviews present in the implantation paradigm can easily lead and impel to answer

changes and to confirming that the false event occurred to comply with the experimenter (Bruck & Ceci, 1999). Therefore, it seems that so called ‘false memories’ could in fact be false reports in some studies and are not caused at all by memory impairment.

Overall, false memories in misinformation paradigms are caused by a combination of memory and social factors, knowingly: an impairment of the original memory trace by the new information, source monitoring deficit, and social compliance effects. False memories in implantation paradigms are based on biased guesses constructed with scripted knowledge in memory. A part of false responses in these paradigms is likely due to pure compliance and results in known errors rather than false memories. Besides, if most studies seem to show that younger children are more prone to false memories in suggestive paradigms, a minority of studies have shown that they were not. Some showed no age effect on the misinformation effect (e.g., Flin et al., 1992; Templeton & Wilcox, 2000). Others showed reversed effects with older children producing more false memories following a suggestion than younger children (e.g., Ceci et al., 2007; Connolly & Price, 2006; Fazio & Marsh, 2008; Fernandez-Dols et al., 2008; Odegard et al., 2009; Otgaar et al., 2016; Principe et al., 2008; Ross et al., 2006). In some of these studies, the suggested information had either more meaning to older children (Ceci et al., 2007) or triggered particularly the processing of the meaning of the implemented false information (Connolly & Price, 2006; Fazio & Marsh, 2008; Odegard et al., 2009; Otgaar et al., 2016). According to Brainerd et al. (2012) who reviewed the mentioned studies, it is precisely because the provided information has more ‘meaning’ to older children that they produced more false memories in the cited studies. Meaning processing was in fact described as the first cause of self-generated false memories. This type of false memory is often opposed in the literature to suggested false memory in that it does not rely on someone else’s suggestion but rather on the way our memory is organized. In the literature, both types of false memories, were shown to be only weakly related (e.g., Otgaar & Candel, 2011; Patihis et al., 2018; Robin et al., 2021).

1.2 Self-generated false memories

1.2.1 Paradigm.

Many studies have looked at false memories arising from internal sources, taking the social influence out of the equation, to understand how memory alone impacts false memories (Brainerd & Reyna, 2005, 2012, Gallo, 2006, for reviews). Different paradigms were used to do so but the most frequently used was the ‘Deese-Roediger and McDermott’ paradigm (DRM). This simple paradigm was shown to be quite convenient to test self-generated false memories

because many parameters can be controlled for. Deese (1959a) created it and Roediger and McDermott (1995) popularized it among the scientific community by replicating Deese's findings. The method consists of showing lists of words associated in meaning to a central word (e.g., navy, soldier, military, infantry, captain, war, uniform, etc., associated to the word 'army') to participants who then have to retrieve the words during a recall test at the end of the list presentation and then a recognition test at the end of all list presentations¹. Results showed that participants falsely recalled and recognized the central word 'army', called usually 'critical item' or 'critical lure' as part of the studied list with a rate rivaling that of correct recall and recognition. Moreover, false recall and recognition of this critical item was much larger than false recognitions of words that were not associated with the studied words. These findings showed that false memories can emerge spontaneously and that they share a common meaning with real experienced events. They are quite persistent as they were shown to be more stable over time than true memories (see Brainerd & Poole, 1997, for a review). They were also shown to occur even under subliminal condition, in which words to learn are presented so fast that they are thought to be processed under the level of consciousness (Cotel et al., 2008; Gallo & Seamon, 2004; McKone & Murphy, 2000; Seamon et al., 1998). One particularity of this error type is that participants report high confidence in their response (Roediger & McDermott, 1995). This confidence is higher for critical or related items than for unrelated items (Ghetti et al., 2008). False recall or recognition of the critical lure were in fact reported to be experienced as a specific detailed episode in memory (e.g., Brainerd et al., 2003; Geraci & McCabe, 2006) with specific sensory inputs (Payne et al., 1996). They produce similar experiences as true memory, even though, true memories were reported to produce in average more perceptive details (Johnson & Raye, 1981; Norman & Schacter 1997). In consequence, it can be difficult to distinguish between a true and a false memory. Moreover, false recognition rate of the critical item or other semantic associates was found to be even greater when preceded by a recall test rather than by a random unrelated task, indicating that the test-retest may enhance the effect (Roediger & McDermott, 1995; see also Gallo, 2006 for a meta-analysis).

1.2.2 Developmental trends and moderators.

¹ In a minority of studies, the recall task was at the end of all list presentations. Recognition tests could also occur at longer delays, from minutes to days after lists perception.

Over the years, the DRM paradigm has shown a great deal of replicability in its findings, even across cultures and languages (Coane et al., 2021). In the developmental field, many studies intended to compare false memories of different age groups from early childhood until adulthood. Brainerd, Reyna and Forrest's study (2002) was one of the pioneers. They presented DRM lists to 5-, 7-, 11-year-old children and to young adults. They showed a developmental trend on false memories production, with younger children (5-7 years) producing fewer semantic false recall and recognition of the critical item and other semantically related items than older children (11 years), who themselves produced fewer than young adults. Errors produced by adults were indeed mainly words related semantically to the studied lists whereas young children's errors included only 50% of semantic errors. To give an illustration, in Experiment 3, for the strongest associated word-lists, 5-year-olds recalled the critical item in 10% of cases, against 27% for 11-year-olds and against 53% for young adults. In the recognition test, the critical item was falsely recognized in 77%, 83% and 92% of cases², respectively for each of the cited age groups. This developmental trend was surprising for some researchers at first, given that studies on suggested false memories using the misinformation paradigm had shown a decreasing pattern with age, most likely because as explained in section 1.1.3, this false memory type is influenced not only by memory factors but also by social factors such as compliance to authority figures in young children. But since then, Brainerd et al.'s findings (2002) have been replicated in more than 100 studies and were supported by imaging data showing that the development of false memories with age was associated to brain changes activations (i.e. the medial temporal lobes and the left ventrolateral prefrontal cortex, Paz-Alonso et al., 2008). A review from Brainerd and Reyna (2012), regrouped many of these studies, showing how developmental performances varied with the manipulation of different factors.

One of those factors was **the type of association** between the studied words. As mentioned earlier, Brainerd et al. 's study (2002) showed that younger children produced only 50% of semantic errors which opens the question regarding the composition of the 50% left over. Dewhurst and Robinson (2004) had the same question and showed that 5-year-olds were more

² The percentages here are A' non-parametric correct scores. See Brainerd et al., (2002) for a full description of the score equation.

likely to falsely recall words that were phonologically similar to studied words (i.e. rhyming words), while 8-year-olds made as many phonological as semantically related errors and 11-year-olds made more semantic than phonological false recall. It confirms that semantic false recall becomes the predominant error with age increase, and it also shows that another error type related to part of the presented words occurs in DRM tasks: phonological errors. This led some researchers to compare two types of list associations. Semantic and phonological associations. To do so, they presented lists of words associated in meaning to participants, like in the classical DRM task (Roediger & McDermott, 1995) or lists of words sharing phonemes (e.g., back, cat, right, and pot) followed by a recall or recognition test. Contrary to classical semantic lists, phonological lists did not create a developmental age trend (e.g., Khanna & Cortese 2009; Dewhurst et al., 2012). In some cases, there were even more phonological errors in younger children (e.g., Holliday & Weekes, 2006). Other research showed that showing lists composed of a mixture of semantically and phonologically related words created an over additive effect of false memories. Indeed, the rate for semantic and phonological false memories increased compared to the presentation of non-hybrid lists in recall and recognition (Finley et al., 2017; Watson et al., 2001, 2003). This effect was found both in children and in adults (Khanna & Cortese, 2009; Sheng et al., 2015).

Other studies have also looked at the **type of semantic associations** between words and their effect on false memories. DRM lists are words considered as associated to a critical item. Word lists were generated by participants who were asked to say the words that they thought of at first when they were told this central item. Association norms allowed to score the strength between each word lists and the critical item. Now if we take a look at the DRM original lists (Roediger & McDermott, 1995) and at the extended lists created in 2001 by Roediger et al., we can notice that some critical items generated rather contextual thematic associations (e.g. hill, valley, climb, summit, top, peak, etc., for the item ‘mountain’) and that other lists generated rather more categorical associates (e.g., apple, vegetable, orange, kiwi, citrus, pear, etc., for the item ‘fruit’). Of course, the lists are not strictly classified as categorical or thematic in the DRM and are usually composed of a mixture of both. But in free associations, some critical items call spontaneously for rather more contextual or categorical associates. Some studies looked at the effect of showing lists of pure categorical associates on false memories at different ages, often showing animal categories to an age range between 5 years and young adults. They showed that the classical DRM false memory increase with age appeared with this type of list as well (Brainerd & Reyna, 2007; Brainerd et al., 2004; Howe, 2006; Howe, Berry, & Knott, 2011;

Knott et al., 2011; Sloutsky & Fisher 2004a; Wilburn & Feeney, 2008). For instance, in Brainerd et al. (2004) and Howe (2006), false memory rates roughly doubled between 5 and 11 years of age. Fisher and Sloutsky (2005) who compared 5-, 7-, 11-year-olds and young adults showed that the biggest increase happened between the ages of 7 and 11, and between 11 and adulthood, which corresponds to important semantic memory expansion periods. This semantic expansion would underly the increase of false recall or recognition of the critical item and semantically related items with age. One point to be raised, is that DRM lists were originally created for adult subjects and the words do not always have the same meaning for adults as they do for children and between children of different ages.

Following this reasoning, if presenting words that activate equal meaning in all compared age groups, then the developmental trend should disappear. The term ‘equal meaning’ here refers to conditions, where the words presented to one age group would trigger similar gist memory activation than the words presented to another age group. Some studies intended to get close to this idea by presenting **lists adapted in meaning** to young children (Anastasi & Rhodes, 2008; Metzger et al., 2008). They showed however that the developmental trend persisted between 3 years of age and adulthood. Carneiro et al. (2007) presented 4-, 7-, 12-year-olds and young adults lists of words generated by peers of their matching age group. In consequence, each age group studied their own version of the lists with the strongest associates, for a given DRM critical item. Total associative strength of lists was similar across age groups. As a result, the developmental trend on false memories was smaller. It seems therefore, that when the when the word-lists’ associative strength is adapted to each age group, the developmental reversal in the DRM can be reduced. Note however that this is not representative of everyday life situations, as we are all exposed to the same stimulations for a given event or situation.

Another parameter that has been tested was **list length**. Coane et al. (2007) have shown that presenting short lists of length 5 or 7 reduced false memory occurrences, with the rate of critical item recall going down to 20% (instead of the usual 50% for studies using 15 wordlists). This rate was nonetheless still above the rate of false recall of unrelated words. Similar results were found in developmental studies across age groups, and the usual false memories increasing pattern was either sometimes preserved in shorter lists (e.g., Carneiro et al. 2007) or suppressed when comparing a smaller age range- i.e. 9-10 year-olds with young adults (e.g. Sugrue & Hayne, 2006; Sugrue et al., 2009).

Moreover, some of the above cited studies used pictures in the presented lists instead of words. Most studies showed that the **use of pictures** to represent a chosen word usually decreased the average rate of false memories in the DRM task in young adults (Dodson & Hege, 2005; Hicks & Starnes, 2006; Israel & Schacter, 1997; Schacter et al., 2001; Smith & Hunt, 2019, see Brainerd & Reyna, 2005 for a review). In developmental studies, the usual age trend in false memories was reduced and sometimes even erased by the presentation of pictures (e.g., Ghetti et al., 2002; Howe, 2006; Howe et al., 2004; Khanna & Corteze, 2009). Howe (2008) showed that the disappearance of this effect came from the processing of distinctive perceptual features in the image. Indeed, when participants between 5 and 11 years were shown items with different specific backgrounds each (e.g., juice depicted in a refrigerator, an apple in a bowl), false memory rates decreased compared to when the background was similar for all studied items (e.g., all fruits being depicted in a refrigerator). Distinctive features and encoding context seem therefore responsible for reducing false memories in picture-lists. In comparison, word-lists would have fewer perceptual distinctive features, leading to fewer correct rejections of related items (Smith & Hunt, 1998).

Beyond DRM tasks, other paradigms used rather more **ecological material**. For instance, sentences were presented to 7- to 10-year-old children (Paris & Carter, 1973). False recognition rate of semantically related sentences was higher compared to unrelated sentences, showing a similar effect to DRM lists. In other studies, pictures depicting scenes were presented to the participants (e.g. the grocery shop). During recognition, the target scene was depicted as well as scenes that were congruent with the studied scene (e.g. a pile of oranges scattered on the floor of the vegetable section of a grocery store) or unrelated scenes. Results show that participants falsely recognized more often the related than the unrelated scene (Hannigan & Reintz, 2001). Like in DRM tasks, there was a developmental trend for false recognition going from near floor levels in 6-year-olds and increasing until stabilizing to 25% at 9-years-old (Lyons et al., 2010). Even though this rate is lower than with auditory words, because it is visual stimuli which allows distinct feature processing, false memories still occur under those more ecological conditions in older children, which is an argument for their applicability to current life situations. Odegard et al. (2009) even proposed a real-life immersion paradigm in which children between 5- and 12-years-old participated in 4 consecutive days of birthday party events. Each party had a specific theme (e.g., sponge bob cartoon character or harry potter movie character). Children participated during the events in activities related to the theme or in unrelated activities. Ten days after the final event, children were interviewed about the events.

False memories about activities related to the birthday's theme (e.g. having to sing happy birthday to the person disguised as Harry Potter, drink Harry Potter's beverage, hear stories about the character, etc.) increased with age and were higher than false memories about unrelated activities (similar activities but without being presented as related to the thematic event). So, it seems that findings using the DRM and using rather more ecological tasks are congruent. However, in other studies comparing the DRM standard lists with the presentation of the words embedded in stories, results between tasks were slightly different. Indeed, Howe and Wilkinson (2011) showed that between 7 and 11 years of age, false memories increased in both conditions but that the increase was smaller in the 'story' condition. Dewhurst et al. (2007) compared a larger age range (5- to 11-year-olds) and showed that the classical age increase with DRM lists was not found when the words were embedded in stories. Indeed, 5-year-olds produced more false recognition on words related to the story than 8- or 11-year-olds under this condition. Similarly, Otgaar et al. (2014) presented visual scenes to 7–8-year-olds and to young adults. A recognition test was presented afterwards, and children had to say whether objects were presented in the scene. Results showed that children produced more false recognition than adults. Otgaar et al. (2013) also compared the presentation of DRM lists with the presentation of a video with missing details (e.g. beach ball for a video happening at the beach). Those missing related details were presented during recognition. The classical false memory increase with age was obtained with DRM lists but not with the videos, where children produced more false recognition than adults. The authors' account is, that presenting semantic information within its context increases the processing of the meaning of stimuli, especially in younger children. In turn, age differences can become smaller or can even be reversed. Additionally, because the provided information is visual, it also boosts distinctive perceptual features in processing and leads to a reduction of age differences as shown earlier in DRM studies using pictures instead of words.

One question raised by some researchers was whether the false memory effect could be prevented when participants were warned about it. Various studies have shown that **warning** the participants about the false memory effect reduced false memories in adults (e.g., Bixter & Daniel, 2013; Peters et al., 2008; Watson et al., 2004) in particular, when using DRM tasks (Neuschatz et al., 2003; Gallo et al., 1997). Carneiro and Fernandez (2010) looked at the effect of warning on false memories using the DRM paradigm in children, comparing a group of 4-5-year-olds with a group of 11–12-year-olds. Half of the subjects in each age group were warned about false memories and given instructions to try avoiding them. The other half of subjects

was not given the warning. Results indicate that false memories for semantically related distractors were decreased by the warning for the older children only. The authors interpreted the absence of effect in younger children as a lack of metacognitive skills at such ages, which prevented them from engaging in strategy mobilization to effectively reduce false memory production. Another study by Del Prete et al. (2014) also assessed the effect of warning on 7- to 13-year-olds and young adults using a similar methodology to Carneiro and Fernandez (2010). Results showed that the warning increased false memories in 7–8-year-olds whereas it had no effect on 10–11-year-olds and it decreased false memories in 12–13-year-olds and young adults. A recent study compared the effect of warning in adolescents (11-12 years of age) and young adults and found that both adolescents and young adults had fewer false memories with warning instructions (Schopen et al., 2022). They also showed that children benefitted more from warning than young adults. Thus, warning seems to be beneficial for false memory reduction only from late childhood. Indeed, it is likely that in younger children, warning instructions encouraged them to process the meaning of the studied words, which they did not do so much spontaneously compared to older children, explaining therefore the increase in false memories in Del Prete et al.'s study.

Another topic of importance is the impact of emotions on false memories. I mentioned earlier the implications of susceptibility to false memories in regard of witness testimonies for justice and police investigations. In this example, the question of emotions is crucial because testimonies are usually about unpleasant events that are likely to trigger negative emotions. The question of whether emotions, and in particular negative emotions, affected false memories has been investigated. There are two ways of thinking about the influence, of emotions on false memories. One can look either at the influence of an emotional stimuli on recall and recognition or at the influence of a person's emotional state, on the way they recall or recognize information. This distinction leads to different paradigms. A review from Bookbinder and Brainerd (2016) summarized the major work done on both topics.

Regarding the impact of **emotional stimuli** on false memories, many studies used the DRM paradigm and created lists with emotional content, basing their definition of emotion as varying on two axes: the valence of emotions, that could go from pleasant to unpleasant experience and the intensity of the triggered emotion, commonly named arousal (Russel, 1991). Arousal is thought of as a physiological response that can go from calm state to excited state. A DRM list could be related rather to a positive, neutral or negative valence and to high or low arousal. An example of a negative valence list with strong arousal would be 'mad', 'rage',

'annoyed', 'furious', etc., related to the critical item 'anger'. Most studies compared negative with neutral valenced lists. Negative emotional stimuli are indeed most relevant to judiciary cases in which targeted events are very likely to come with negative stimuli (e.g. a weapon in a crime scene). There is also evidence in the literature that negative stimuli trigger different brain responses than neutral stimuli (e.g., Anderson & Phelps, 2001; Hariri et al., 2000; Hedger et al., 2015; Öhman, 2005). This evidence suggests that they are processed differently and could lead to differences in false memory production. The first emotional DRM lists were created by Budson et al. (2006) which contained negative and neutral valences. These authors found no difference in false recognition between negative and neutral lists. However, the sample from this study was quite small (19-20 per group), making the results of this study alone not very reliable. Following this, other research used Budson et al.'s lists and found different results. Indeed, Palmer and Dodson (2009), El Sharkawy et al. (2008) found that negatively valenced lists led to higher rates of false recognition. Similar results were obtained in children from age 5 (Howe, 2007, Howe et al., 2010, 2011). The effect was also replicated with lists with an arousal level controlled (Brainerd et al., 2008b; Dehon et al., 2010). Brainerd et al. (2008b), indeed re-created a pool of emotional DRM lists and included valence positive lists. They controlled for arousal level, so that it was similar across the three valence conditions. They showed that false recognition rate of negative lists was higher than of neutral lists which was itself higher than of positive lists. In another study, Brainerd et al. (2010) compared positive and negative valenced lists and varied the arousal of the lists orthogonally. Children aged 7 to 11 years and young adults participated in the study. Results confirmed that false recognition was higher for negative than positive valence lists in all age groups. Moreover, high arousing lists increased false recognition but only on negative lists. This effect was stronger in adults than in children.

The effect of emotional stimuli was also tested using pictures instead of words. A pictorial database called the International Affective Picture Scale (IAPS), containing stimuli varying on both arousal level and emotional valence, was used to do so. Two first studies showed that the presentation of pictures led to similar levels of false recognition of images with negative, neutral or positive valence (Choi et al., 2013; Gallo et al., 2009). A study from Bookbinder & Brainerd (2016) found however an increase of false recognition with negative valence images compared to neutral ones. They showed that this was caused by an increased processing of the meaning of the image rather than of the perceptual details of the image, on negative compared to neutral stimuli. In another study, pictures of scripted events were

presented (e.g., someone taking an orange at the bottom of the pile in a grocery shop) along with a negative-arousing or neutral-non arousing outcomes picture (e.g., someone picks up a fallen orange for the negative outcome). Half of participants had to write a description of the images they were looking at. This manipulation increased meaning processing. Results indicate that only in this condition, they produced more false memories of negative-arousing events than neutral-non arousing ones, in subsequent recognition (Mirandola et al., 2014). Therefore, it seems that negative-arousing stimuli increase false recognition because they lead to increased meaning processing over detailed feature processing. This effect seems to be attenuated or even suppressed when item meaning processing is not favored (e.g., Mirandola et al., 2014; Howe, 2008).

Finally, contrary to recognition, results in recall test using the valence DRM lists are less consistent across studies. Some studies found no differences between negative and neutral lists in false recall (El Sharkawy et al., 2008; Otgaar et al., 2012a, 2012b) while others found more false recall for negative valence lists than neutral ones (Goodman et al., 2011; Otgaar & Smeets, 2010). The most observed pattern however was a higher production of false recall for neutral than for negative lists across age groups (Dehon et al., 2010; Howe, 2007; Howe et al., 2010, 2011; 2011; Palmer & Dodson, 2009), which is the opposite to the findings in recognition tasks. This suggests that the effect of valence on false memories is task-dependent and is less robust in recall than in recognition. It reflects the dissociation of response patterns between recognition and recall, found in numerous past studies (for a review, see Brainerd & Reyna, 2005).

Another set of studies looked at the impact of **emotional induction** on false memory production. To do so, they induced emotions to subjects either via music, pictures or videos and they looked at how DRM lists were recalled or recognized. Most studies showed that contrary to emotional stimuli, negative emotional induction prior to studying DRM lists, usually led to lower false recognition and recall, in comparison with positive or neutral emotions (Emery et al., 2012; Knott et al., 2014; Strobeck, 2013; Strobeck & Clore, 2005, 2011). The explanation to it, is that negative moods trigger perceptive detail stimuli processing, oppositely to positive moods that would trigger more meaning processing. Strobeck (2013) showed this by asking participants to retrieve in addition to the DRM studied words, some contextual detailed features that were presented during the study phase (i.e., spatial position in Exp. 1, font in Exp. 2, a picture accompanying the word in Exp 3.). He showed that when emotional induction via either music or images, was negative, false memory levels dropped compared to a neutral or positive

emotional induction and the memory for detailed features of stimuli was increased. Hence, it seems that negative emotional induction, contrary to emotionally negative stimuli in recognition, decreases false memories because it boosts feature processing. Moreover, arousal level of emotional induction was also found to impact subsequent false memories. Indeed, two studies varied orthogonally emotional valence and arousal induced by music and showed that high arousing music increased false recall and recognition regardless of the valence (Corson & Verrier, 2007; Van Damme, 2013). It seems therefore that when the arousal of a felt emotion is high, we are more likely to produce false memories. This is an important aspect to take under consideration because this is the very type of situation people usually have to testify about in police interrogation rooms or in court rooms.

It seems that emotional induction and emotional stimuli each affect false memories in their own way. Negative stimuli with high arousal produce the highest rates of false recognition when compared to other valences and arousal levels. As for emotional induction, highly arousing stimuli lead to the highest levels of false memories no matter the valence of the emotion. However, for lower levels of arousal, negative emotional induction leads to fewer false memories than neutral or positive, because it triggers feature detailed processing. Some authors looked at the effect of **congruency between emotional induction and emotional stimuli** on false memories. They showed that when the induced emotion (i.e., positive or negative) was consistent with the emotional valence of the DRM list of words (i.e. positive or negative, respectively), false recognition increased (Knott & Thorly, 2014; Ruci et al., 2009). Bland et al. (2016) even conducted a study comparing this congruency effect of qualitatively different emotions. Participants were induced either fear, anger or neutral emotions via videos. They were then exposed to DRM lists relating to fear, anger or neutral emotions. Lists relating to negative emotions had equal arousal. A congruency effect appeared for both anger and fearful emotions on false recognition. Hence, we have more chances of producing false memories when our emotional state is congruent with the emotional valence of the stimuli, whether it is positive or negative.

Individual differences.

We have seen that age influences false memories during childhood, there are however other individual differences that can play a role in false memory production. One of them is disabilities in children. One way of showing that semantic processing is the cause of false memories, besides comparing different age groups or showing lists that are strongly related in

meaning, is to compare children with different abilities to process semantic content of stimuli. If semantic processing causes false memories, we could hypothesize that children with impaired semantic processing would produce fewer false recognitions of critical items. Weekes et al. (2008) had 9- to 11-year-old children with and without **reading comprehension difficulty**, which is associated with a lack of semantic processing, to do a DRM task with semantically and phonologically related lists. They showed that children with reading comprehension difficulties had fewer false memories in recall and recognition than children without difficulties, but only on semantically related lists. This bears support to the hypothesis that, semantic processing is necessary for the production of false memories. These results were extended to learning disabled children (Branch et al., 2007). Brainerd et al. (2006, Exp 2.) also showed that the age increase in false memories between ages 7 and 11 was reduced in learning disabled children.

Another source of individual differences relates to **maltreated** children. It is an important topic for false memories because some children involved in legal cases have a history of maltreatment. One debate was whether maltreatment would lead to an increase of false memories or not (Howe, 2000; Toth & Cicchetti, 1998). There were findings in which the stress generated by a trauma could impair brain structure and proper functioning over an extended period (Bremner & Narayan, 1998; Sapolsky, 1992), in particular in the hippocampus (e.g. Bremner et al., 2003; Carrion et al., 2007, 2010; De Bellis et al., 2001). Memory performance was also found to be impoverished in persons under post-traumatic stress (e.g., Moradi et al., 1999) and in children with abuse experience (Valentino et al., 2009). These brain and memory performance alterations were however not confirmed by all studies' findings (e.g., De Bellis et al., 2001; Beers & De Bellis, 2002; Howe et al., 2006; McWilliams et al., 2014). Moreover, some studies looked at the effect of being maltreated as a child on false memories in the DRM. For instance, Howe et al. (2004) recruited 5- to 12-year-old children, with children coming from families with low-income and coming from families with average income. Among the children with low-income families, half were reported to the local department of human and health services for concerns related to child maltreatment. Results indicate a classical developmental trend of false memories, regardless of the maltreatment status. A limitation of this study is that children with a history of maltreatment only came from low-income families, making the generalization to the other income groups impossible. Other studies (Howe et al., 2011; Baugerud et al., 2016; Otgaar et al., 2017) went further and tested neutral and negative emotional DRM lists. Two of them found that maltreated children produced more false memories on negative

valence lists but not on neutral lists. It suggests that this group might process negative information differently from non-maltreated children. Otgaar et al. (2017) proposed that this population might be more sensitive to negative emotional stimuli. When they are confronted by it, they are more likely to make correct associations in memory between the studied words of the list, but also incorrect associations between the studied and unstudied words, leading in turn to more false memories. Goodman et al. (2010) found that these results were extended to adolescents and adults with a history of child sexual abuse. Moreover, consistently with Howe et al.'s findings (2004), some studies found no differences between maltreated and non-maltreated children in their susceptibility to misinformation (Chae et al., 2011; Eisen et al., 2002, 2007), whereas another showed a decreased level of susceptibility to suggestion in maltreated children (Otgaar et al., 2017). To summarize, maltreated children might be more prone to false memories about negative stimuli when testifying in legal cases. Besides, there is for now more evidence that they would not be more suggestible.

Finally, another factor that has only recently been investigated in children is **self-reference**, as opposed to other reference or neutral reference. Self-reference relates to the association of a stimulus to oneself. For instance, if a child has a specific toy at home depicting an animated movie character, they will likely associate this character to one-self. Other reference is when we associate a stimulus to others. For example, if the child has a friend that has another toy character at his house, this character can be associated to his friend and not to himself. The third option would be that the child sees a toy that he's never seen before, in which case the toy would not be associated to one or the other self. Some studies looked at the effect of self or other reference on false and true memories. Most of them were conducted on adults. They showed that self-reference increased true memories (Klein, 2012; Symons & Johnson, 1997; Wang et al., 2021). Interestingly, it was also found to increase false memories (Ozdes, et al., 2021; Rosa & Gutchess, 2013; Wang et al., 2019; Wang et al., 2021). A recent study showed that adult's false memories were increased by self-reference in DRM tasks (Wang et al., 2019). In children, a study has also shown that self-reference increased true memories, looking at 4- to 6-year-olds (Cunningham et al., 2014). Only one recent study tested the effect of self-reference on false memories in children (Wang et al., 2022). In this study, participants performed a DRM task in which lists of words were either presented with their own name (self-reference condition), someone else's name (other-reference condition) or with a red square (neutral condition). A canonical name was chosen for the other reference so that participants had more chances to know some people with that name, including famous people. Participants

were asked to remember the presented words and the names they were associated with. A recognition test followed the procedure with studied items, critical items and unrelated word items. Results indicate that in the two age groups (children of 8- to 9-year-olds and adults), when lists of words were associated to the participants' name, false recognition of the critical item increased compared to the two other conditions. Therefore, self-reference seems to increase false memories at least from middle childhood. Moreover, during recognition, participants had to specify when they identified the probe as part of the studied list, whether they could remember the event with specific details such as the font, size of the words, etc. (i.e. remember condition) or if they could not recall those details but that they knew they heard the presented word (i.e. familiarity judgment). Results showed an increase in familiarity judgements over remember judgments in the self-reference condition compared to the other-reference condition. So it seems that stimuli associated to oneself trigger a stronger semantic processing rather than detailed processing and therefore lead to both more true and false memories. These findings have everyday life implications and play a role in explaining why testimonies for the same event do not always match between two people. Indeed, depending on the stimuli and to who and what we associate them with, and more generally depending on the meaning they have for us, the way we will remember them and the likelihood of producing false memories about them varies. So, because of inter-individual differences: the self, age, history of maltreatment and semantic processing ability, differences in testimonies about an event can emerge.

In this Chapter I reviewed the topics that have been investigated in false memories, ranging from childhood to early adulthood. In the next chapter, I will focus on explaining the underlying mechanisms of self-generated false memories. I will present the existing different theoretical accounts of false memories.

Summary

In this chapter, I recounted that there are two distinct kinds of false memories: suggested ones and self-generated ones. Suggested false memories, which were evidenced in the misinformation and the implantation paradigm, are usually caused by a combination of memory and social factors, namely: competition of traces between the original memory and the suggested information, semantic processing, source monitoring failure, scripted knowledge, social compliance and reliance on authority figures. On the contrary, self-generated false memories are mainly based on memory causes. I highlighted that suggested false memories were usually higher in younger than older children. Oppositely, self-generated semantic false memories increase with age from childhood until young adulthood. Studies usually used the DRM paradigm to show this effect, where lists of words most associated to a single central word are presented to subjects before recall and recognition. The developmental reversal trend found in the DRM was also found in pure categorical lists of words but not in phonologically similar wordlists. It was also found to be diminished when showing small lists of words or when words-lists presented to each age group were adapted in meaning. Experimental paradigm variants using more ecological material (e.g., pictures, videos, visual scenes) showed similar findings to those obtained with the DRM, with the sole exception that when the meaning of the stimuli was increased because of the material, the developmental age increase of false memories could be reduced, annulled or reversed between some age groups. Moreover, warning subjects about the false memory effect helped to reduce it from late childhood. Negative and arousing emotional stimuli also increased false recognition of related information and produced mixed findings on false recall. The induction of negative emotions on the contrary, decreased false memories compared to neutral or positive emotional induction, but most importantly, high arousal levels in emotional induction were responsible for false memory increases regardless of the valence. Moreover, when the emotion induced prior to stimuli exposure was congruent with the emotional content of the studied stimuli, false memory effects were increased at all ages. Besides, children with difficulties in meaning processing were shown to produce fewer false memories. It underlines the role of semantic processing in false memories. Maltreated children also show a biased sensitivity to negative stimuli, which leads them to produce more false memories on such material. Lastly, the association of stimuli to one-self increases the chances of false memories, because it triggers stronger meaning processing in comparison with other-associated stimuli.

Résumé

Dans ce chapitre, deux types distincts de faux souvenirs ont été présentés : les faux-souvenirs suggérés et auto-générés. Les faux souvenirs suggérés, qui ont été mis en évidence par les paradigmes de désinformation et d'implantation, sont souvent causés par une combinaison de facteurs sociaux et de mémoire, à savoir : la compétition des traces entre le souvenir original et l'information suggérée, le traitement sémantique, l'échec de la récupération de la trace du souvenir en mémoire, les scripts en mémoire, la conformité sociale et la conformisation aux figures d'autorité. Au contraire les faux souvenirs auto-générés sont basés principalement sur le fonctionnement de la mémoire. J'ai souligné que les faux souvenirs suggérés étaient généralement plus nombreux chez les jeunes enfants. A l'inverse, les faux souvenirs auto-générés, augmentent de la jeune enfance à l'âge adulte. Les études, utilisent généralement le paradigme DRM pour montrer cet effet. Dans ce paradigme, des listes de mots associés à un mot central sont présentées à des sujets avant un test de rappel puis de reconnaissance. La tendance développementale inversée observée dans les tâches DRM, a également été montrée avec des listes purement catégorielles mais pas avec des listes de mots reliés phonologiquement. Cette tendance était diminuée, lorsqu'étaient présentées des listes de mots plus courtes ou lorsque les listes étaient adaptées en sens à chaque groupe d'âge. Des paradigmes expérimentaux utilisant du matériel plus écologique (e.g., images, vidéos, scènes visuelles) ont montré des résultats similaires à ceux obtenus avec le DRM, avec la seule exception que lorsque le sens des stimuli était renforcé par le matériel, l'augmentation des faux souvenirs avec l'âge pouvait être réduite, annulée ou même inversée entre certains groupes d'âges. De plus, prévenir les sujets à propos de l'existence des faux souvenirs a permis de réduire leur taux chez des enfants âgés. Par ailleurs, les émotions négatives et intenses augmentent les fausses reconnaissances et produisent des résultats mixtes sur les erreurs sémantiques en rappel. Au contraire, l'induction d'émotions négatives diminue les faux souvenirs en comparaison à l'induction d'émotions neutres ou positives. De façon plus importante, les émotions intenses induites provoquent l'augmentation des faux souvenirs, peu importe leur valence. De plus, lorsque les émotions induites en amont de l'exposition à des stimuli émotionnels, sont congruents avec ces derniers, les faux souvenirs augmentent, chez les enfants comme chez les adultes. Par ailleurs, les enfants ayant des difficultés à traiter le sens de l'information produisent moins de faux souvenirs, ce qui souligne le rôle du traitement sémantique dans la formation de faux souvenirs. Les enfants maltraités ont également un biais de traitement des stimuli négatifs, ce qui les conduit à produire plus de faux souvenirs sur un

tel matériel. Enfin, l'association de stimuli au concept de soi augmente les chances de faux souvenirs, car cela déclencherait une augmentation du traitement du sens de ces stimuli, en comparaison avec des stimuli associés au concept d'autrui.

Chapter 2. Theories of false memories

In this Chapter, I will present the main theories of self-generated false memories. The first type of theories can be regrouped under the name of spreading activation theories, which relates to the idea that information in memory is organized and distributed upon a network. The second one, is a dual process theory, considering two main types of traces in memory : precise, detailed, surface traces called '*verbatim*' and meaning based traces, called '*gist*'. The last theory called global matching, explains false memory in recognition tests, based on the idea that traces in memory are composed of several features, that will more or less match with the recognition probe depending on their similarities. The aforementioned theories bring a different light on false memories and are in my sense, all helpful in understanding it. Before introducing those theories, I will present some general principles regarding memory functioning, and introduce some mechanisms involved at different stages of the memorization process that are useful for understanding false memories.

2.1 Generalities about memory.

Psychologists defined three important stages characterizing the course of a memory in time. The first stage is the moment the memorandum is being 'written' in memory, also classically called 'encoding'. The second stage corresponds to the storage (or maintenance) of the trace previously created in memory. The last stage relates to the moment when the memory is retrieved (i.e. retrieval). During each of these stages, some factors determine the chances of properly retrieving a memory or not. False memories and more generally memory errors, can occur during each of these stages. I will therefore present the factors that contribute to and influence memory retrieval at each stage.

Encoding is a determinant stage for a memory because it defines the information that will be stored. During encoding, the attentional resources are deployed to process and encode environmental stimulations in memory. Attentional resources are however limited, and they cannot be allocated to all environmental stimuli (e.g., Lindsay & Read, 2006). There is then a prioritization of stimuli receiving attentional resources, which depends on various factors, such as whether the stimuli are relevant for the task being done (e.g. learn only the names of the people I'm assigned to work with; i.e.- top-down process; e.g., Carrasco, 2011, 2014), or

whether stimuli are relevant to survival or not (e.g. dangerous stimuli draw attentional resources automatically, i.e. bottom-up process; e.g., Gallo, 2010), etc. The stimuli getting processed with the most attentional resources are the ones that have the best chance of being encoded in memory (e.g., Craik et al., 1996). Moreover, Craik and Tulving (1975) showed that the level of processing of the stimuli (LOP) during encoding impacts their recognition. Indeed, they found that when the same information was processed either by its surface details (shallow LOP, i.e. say whether the word is written in upper or lowercase letters) or by its meaning (deep LOP, i.e. say whether a sentence has a meaning or not when a given word is added to it), there were differences in later recognition. Indeed, correct recognition increased when information was processed by its meaning. They concluded that deeper LOP led to more resistant and robust memory traces in that they were more strongly consolidated (see also Tulving & Thomson, 1973). Consolidation refers to the transfer of information between WM and LTM's storage, occurring within cells during the few hours after encoding, making ultimately memory traces' storage more persistent in time (Davis & Squire, 1984; DeZazzo & Tully, 1995; McGaugh, 2000). Aside from LOP, some environmental factors are also known to facilitate the consolidation of information in LTM. The best known example is sleep (Stickgold, 2005, for a review). Other factors, like alcohol on the contrary, are known to prevent good consolidation (Ryback, 1971; Wilkinson & Poulos, 1987; Browning et al., 1992, for reviews). Hence, consolidation of information is critical for memory trace persistence because it determines whether or not traces are stored in our LTM, which in turn determines their availability or not during retrieval. If the trace was not consolidated in LTM, it won't be available at retrieval, and oppositely if it was. But let's not forget that this is not an 'all or nothing' system, and only parts of the memory can be consolidated, leading to more or less detailed memories at retrieval.

The retrieval of a memory arises from a reconstruction of various elements such as the stimuli's different features, the meaning of the stimuli and its contextual details (Roediger & McDermott, 1995; Surprenant & Neath, 2013). Metaphorically, it is like putting the pieces of a puzzle back together. Schacter and Addis (2007) pointed out that this reconstruction has its perks. Indeed, the process of recombining elements allows the imagination of future scenarios, where traces from different memories can be combined to form a new event. For instance, imagining a pink horse in your grandmother's living room is possible because in the past you were confronted to the color pink, to horses and to your grandmother's living room. All these elements created memory traces and when combining them, we can imagine an event. Reconstruction is therefore a source of mental flexibility. It is however also a source of error.

The same way we can easily create a scenario, we can also create wrongful memories, if putting back the wrong pieces of the puzzle together.

Availability of memory traces is not sufficient for it to be retrieved, it also must be accessible (Tulving & Pearlstone, 1966). Imagine that you are in a jungle, and you leave your pair of socks in a specific place. You come back 6 months later looking for your socks. The chances you'll have to find them will depend on whether you can find your way back to the place you left them. If you never went back on the way, it is likely that you will struggle finding it back because the vegetation would have regrown. On the contrary, you will more likely find your socks if the way is cleared because you returned on it several times since then. In both scenarios, the pair of socks was available, waiting at the same place. It was rather the accessibility that varied. Accessibility depends on a variety of factors such as the time lapse between stimuli perception and retrieval, whether we rehearse the way back to the information and at which frequency. Paradoxically, regularly retrieving information to rehearse it, was shown to increase its correct retrieval (Hintzman, 1976) but it was also shown to lead to error (e.g., Skurnik et al., 2005). Indeed, every time information is activated in memory, a new trace of this information is created, which reinforces the overall strength of the information. The downside is that, because the new trace cannot have the exact same contextual cues, the memory for the primary context decreases, leading to error increase (e.g., Arndt, 2006; Tulving & Thompson, 1971).

Finally, the easiness of accessing a trace in memory depends also on the cues we have around us at the time of retrieval. Indeed, any cue that is assimilated to the memory can enhance the chances of retrieval access (e.g., Tulving & Osler, 1968). This is why studies have shown that being in the same context than the one in which specific information was learned, boosts memory of that information (e.g., Maren & Holt, 2000; Spear, 1973). This is also why the chances of retrieving information in recognition is higher than in recall, like it is the case in the DRM for instance. Indeed, recognition tests usually deliver the intact stimuli or some parts of it, which helps getting access to the stored trace in memory. Those cues can also paradoxically lead to false memories as well. In developmental studies, alongside the development of false memories with age described in Chapter 1, studies usually show a development of correct memories as well (see Brainerd & Reyna, 2012 for a review). To get an illustration, in Brainerd et al.'s study (2002) that we described in Chapter 1, the correct recall rate moved from 20.5% in 5-year-olds to 40.5% in 11-year-olds and to 63% in young adults, in Experiment 3. Correct recognition followed a similar increasing pattern going from 56.5% to 66.75%, and to 69.5%

for each age group respectively. Theories of false memories explain why we find such paradoxical increases in correct and false memories with age.

2.2 Spreading activation theories

2.2.1 Knowledge organization and spreading activation.

Spreading activation theories of false memories such as activation monitoring theory (AMT, Roediger et al., 2001) and associative activation theory (Howe et al., 2009; Otgaar et al., 2019a; see also Anderson, 1983; Anderson & Pirolli, 1984) share the common idea that our knowledge is an ensemble of conceptual nodes, that are interconnected within networks based on their associative strength. Taking two concepts, the more they are associated in memory, the more their connection within the network will be strong. The associative strength between two concepts is determined by life experiences that are common to a culture or individual-dependent. For instance, networks can build on some scripted events, that can be very alike between individuals (e.g., taking breakfast in the morning, the way a doctor appointment happens, going to the movies), leading in turn to similar conceptual associations between-individuals. Similarly, during childhood, some associations form in memory, based on cultural learned categories containing concepts sharing similarities (e.g., the category of animals, the category of words finishing with the sound ‘j’ used in the words cash and flash for instance). Other types of associations depend more on the individual’s personal experience. For instance, if when going to school every morning you walked by a dog named Ralph, it will be associated to the school path in your memory but not in someone else’s, who did not experience these repeated events. This associative process is supported by the long-term potentiation effect which relates to the fact the synapses in the brain between two neurons can produce a long-lasting increase in signal transmission (Cooke & Bliss, 2006, for a review). It can happen when two stimuli are presented together repeatedly (Hebb, 1949). Their association is then strengthened. Resulting from this, several networks constitutes our memory which is underlined by either phonological, contextual (can also be called ‘thematic’) or categorial associations.

The principle of spreading activation theories, as stated in their name, is that when a concept (e.g. ‘soldier’) is activated in memory, the concepts associated to it can also be automatically activated (Collins & Loftus, 1975; Landauer & Dumais, 1997). Indeed, the activation of ‘soldier’ spreads to its closest associates within the network. Because of this, concepts that

have not been experienced, such as the critical item 'army' in the DRM paradigm can be activated in memory and be in consequence mistaken as part of the truly lived event.

Now based on this explanation of the theory only, one could say that they could have a hard time understanding how a whole memory episode is recreated. According to the spreading activation theories, a memory event is reconstructed by the simultaneous activation of the different associative memory networks- i.e. categorical, phonological and contextual. It is the combination of these activations that creates the memory (Roediger et al. 2001). Moreover, the cues present at the moment of retrieval will vary each network's weight activation (Newman & Lindsay, 2009). Indeed, a memory is never retrieved as a whole and the elements that will be retrieved at a specific time will depend on the environmental cues given to the subject.

2.2.2 Source monitoring

When neighboring nodes are activated in memory because of the spreading of the target node's activation, the challenge is to be able to identify that it was not really part of the memory. AMT, one of the most influential spreading activation theories, which is based on a combination of two theories -i.e., the implicit associative response theory (Underwood, 1965) and the source monitoring theory (Johnson et al., 1993), insists on the importance of being able to identify the source of an activation in memory, in order to properly reject the unexperienced words activated. Gallo (2010) describes two kinds of monitoring. The first one is 'criteria-based monitoring'. It relates to making diagnostic decisions in which you judge whether the event or part of it was likely to happen or not, based on the idea that if it happened you would remember it (e.g., 'I did not study the word 'army' because if I did, I would remember it'). The second one is corroboration-based criteria. It refers to disqualifying decisions, in which you discard an item activated in your memory because you know that you experienced another item instead (e.g., 'I did not study the word army because I remember that I saw the word soldier instead'). So, during source monitoring, we judge the credibility and plausibility of the occurrence of an event. When a concept is activated in memory because of its similarity to a perceived stimuli and when people cannot identify the real source of this activation, the production of false memory occurs.

2.2.3 DRM studies as evidence for the spreading activation explanation.

In the DRM paradigm, the word lists are all strongly associated to a central critical item. All words contribute individually and by being activated together, to the activation of this critical item and to other semantically related words (see Figure 1, for an illustration). There is a strong likelihood that the critical item will in consequence be activated in memory. This activation combined with the failure to monitor its source would lead to false memory. It has been shown that the more DRM word lists are judged as associated to the critical item, the higher false memory rates were, both in children and adults (e.g., Brainerd et al., 2002). The score of associative strength between DRM words and the critical item in lists is in fact one of the strongest predictors of false memory occurrence (Roediger et al., 2001; but see Brainerd et al., 2020). In developmental studies, age increase in false memories was amplified for word lists with the highest associative scores (e.g., Howe et al., 2009).

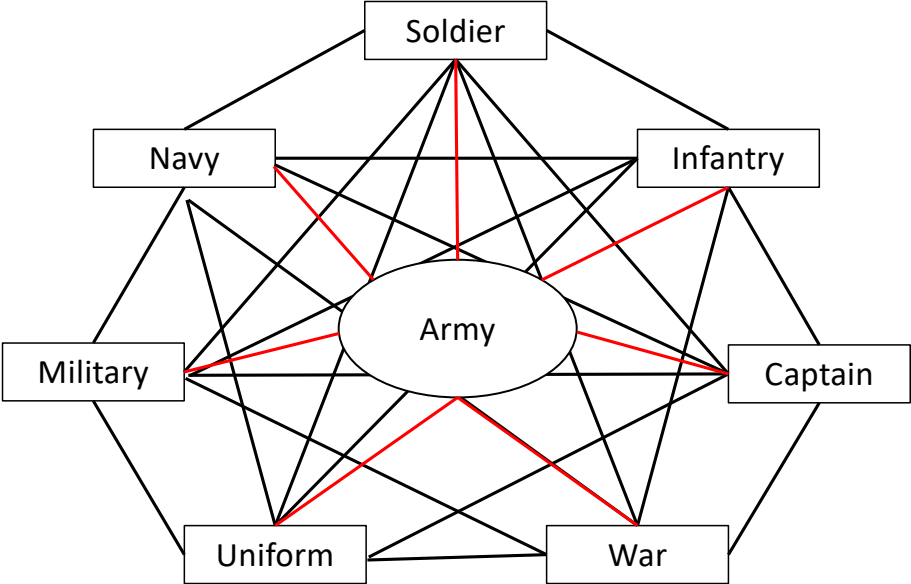


Figure 1. Illustration of the process of activation of the critical item in the DRM paradigm according to the spreading activation theories. Word-lists activate the words strongly associated to them in the network: they activate the other studied words of the list, the critical item (i.e. the strongest associate to the word-list) and also other related words that are not presented in this figure. The word ‘army’ is automatically activated because many strong associates are activated during the study of the DRM lists. Each of these associates contribute to the activation of ‘army’ individually and also by being activated together. Indeed, because those associates are also strongly related, they increase each other’s activation in memory (black lines) and indirectly favor ultimately the activation of the critical item ‘army’ (red lines).

2.2.4 Developmental perspective.

Spreading activation theories explain the false memory increase with age between early childhood and young adulthood, by stating that the associative networks are simply less developed, and concepts are less interconnected in young children relative to older ones and in older children relative to adults. This is because knowledge organization described in a previous section 2.2.1, builds up with age. The older we get, the more experiences we accumulate, in which two given stimuli are encountered together. As a consequence, the strength of associations between stimuli within the network increases with age. Wimmer and Howe (2009) showed indeed that between the ages of 5 and 11, children produced spontaneously more meaningful associates to a given word as they aged and this production became faster with age, indicating that the activation of word associates was more automatic. The term automatic refers here to an “activation of a sequence of nodes that nearly always becomes active in response to a particular configuration and that is activated automatically, without the necessity for active control or attention by the subject” (Schneider & Chein, 2003, p.526-527). In the DRM paradigm, the activation of the critical item or other related stimuli is therefore more automatic with age (Otgaar et al., 2019a, 2019b, for reviews). The development of associative networks with age can account for the developmental reversal effects on false memories described by numerous studies in Chapter 1. Indeed, if stimuli are less connected in meaning on the network, this would result in a smaller production of correct memories but also of false memories. Moreover, there is not just an increase in network connection strength with age, the connections between concepts are also modified with experience. For instance, at some age the strongest associate of ‘army’ could be ‘soldier’ and at another it could become ‘military’. It depends on the frequency that two concepts are presented together, which is likely to vary during the lifetime. This explains why when word-lists are adapted to each age group, with each age group tested on its own strongest associates, the age group differences are reduced (Carneiro et al., 2007). It does not completely abolish them however, because even though the word lists are the most adjusted to the children’s current memory network, it does not remove the fact that connections between concepts are still weaker at youngest ages because they have fewer experiences in which they have encountered the words together.

Regarding source monitoring, studies have shown that even though children as young as 5-years-old were able to monitor the source of some memories (Wimmer & Howe, 2009), and to do recollection-rejection (Brainerd, et al., 2003; Carneiro, et al., 2009; Ghetti, 2008), which is a type of corroboration-based monitoring, source monitoring increased with age. Wimmer and

Howe's study (2009) illustrates this well, by showing that when participants had to do a recognition task in which, they had to judge whether the presented probes were previously generated by themselves or not, correct source attribution increased with age between 5 and 11-years. More generally, children were found to confuse memories of imagined stimuli with memories of actually experienced stimuli (Johnson & Foley, 1984; Lindsay & Johnson, 1987, for reviews) and memories of self- vs. other-produced stimuli (e.g. Hall, 1969; Lindsay et al., 1991; Foley et al., 1991; Foley & Johnson, 1985; Roberts & Blades, 1998). The development of source monitoring is conditioned by the ability to retrieve the details of the source of information (Foley, et al., 1983, 1993; Johnson, et al., 1993) and is closely related to the development of executive functions between early childhood and adulthood (i.e. inhibition, shifting, updating) (Cycowicz et al., 2001; Ruffman et al., 2001; Rybash & Colilla, 1994). It could explain why younger children produce higher rates of suggested false memories. Indeed, younger children likely have more difficulties in distinguishing whether the suggested information came from themselves or someone else, which makes them more prone to errors in recall and recognition in such paradigm. It does not explain however the increase of false memories with age in self-generated false memory tasks. Indeed, we could expect that the development of source monitoring with age would help decrease this false memory type as well. But it is likely that source monitoring development might not play such an important role in the development of this error type and that the main determinant of the developmental trend in this case would be the strength of association between studied items, which is responsible for the activation of semantically related associates, independently of the variability in source monitoring with age. Indeed, even though Wimmer and Howe (2009) showed an increase in source monitoring with age, they defended that youngest children were already good at discriminating between externally versus internally generated information and actions (e.g., Foley, et al., 1983; Johnson, et al., 1993). So, even though it is harder for young children to monitor the source of a memory, it does not mean that they cannot monitor the source of information at all. Studies reported in fact, that children could properly reject a distractor using corroboration-based monitoring (also called recollection-rejection, in the cited studies) (Brainerd, et al., 2003; Carneiro, et al., 2009; Ghetti, 2008). This is why the associative activation theory focuses mainly on the spreading activation principle to explain self-generated false memories (Otgaar et al., 2019a, for a review). Nonetheless, intrusion of items from previous episodes, which is greater in younger children (Kail, 2002, for a review) is an error type that seems to be impacted by the development of source monitoring in DRM tasks (e.g., Metzger et al., 2008). The decrease of this error type with age is in fact closely related to the

maturation of executive system (see McCormack et al., 2000, for a review). Note that suggested false memories are in some aspect similar to intrusion errors in DRM tasks, as the object of the error came from a distinct episode. Hence, it seems that the development of source monitoring with age prevents intrusions from previous episodes, leading to a decrease of this error type with age but not provoking a decrease of self-generated meaning based false memories with age.

2.2.5 Limits.

Spreading activation theories have been criticized because they cannot account for all findings regarding false and true memories. Indeed, it was shown that true and false memories are not always related and can therefore not depend only a common process. The manipulation of some variables was shown to affect indeed only one of the two (i.e., single dissociation) while other variables affected both (i.e., double dissociation), suggesting the involvement of dual processes (Bookbinder & Brainerd, 2016, for a review). Besides, this model does not give precisions about the nature of the retrieved traces. Indeed, it seems that the activated conceptual nodes of the network are like abstract units. This model lacks precision about the integration of potential detailed surface memory traces represented in their sensory modality (visual, auditory, etc.), which could play an important role in the prevention of false memories when they are retrieved. This lack of precision is likely why it is not so obvious to understand how a detailed memory of an episode is recreated. I will present now a theory that brings precisions on both surface feature traces and meaning based traces in memory.

2.3 Fuzzy Trace Theory

2.3.1 Principles of the theory.

Fuzzy trace theory (FTT) is one of the most influential theories in explaining false memories and their development (Brainerd & Reyna, 2002a; Brainerd et al., 2008a). This dual process theory proposes that we have two types of memory traces. The first type correspond to meaning-based traces, called ‘gist’ memory because it relates to the general concept of studied stimuli. For instance, if I heard the word ‘soldier’, my gist memory of it would be that I studied an item about a profession related to the army. The second type of memory trace relates to the detailed surface feature traces of a stimulus, called ‘verbatim’. The way ‘soldier’ sounded, its phonological features will form verbatim memory traces. Contrary to verbatim, gist memory traces form not only on the basis of each stimulus but also on the combination of stimuli presented together. Indeed, if the word ‘soldier’ was presented with other related words, such

as ‘military, infantry, captain, war, uniform’, a gist trace composed of the shared meaning between those words would be created (e.g., words related to the army) in addition to the gist triggered by each individual words (e.g. ‘uniform’ is a type of clothing, a semantic characteristic not shared with the other words). Both gist and verbatim traces are stored in parallel in memory. At retrieval, the type of traces recovered determines whether we produce true, false memories, or other errors (Brainerd & Reyna, 2005). According to the FTT, verbatim memory allows to properly retrieve information and produce correct responses only. Indeed, when we have access to the exact form of the perceived stimuli, we can either correctly recall or recognize it, or correctly reject any distractor that would not match the trace (e.g., Brainerd et al., 2003). Gist memory however produces both correct and false memories. Imagine that you only retrieve that you heard a word about the military, you could easily correctly recall that you studied ‘soldier’, but you could also falsely recall that you studied ‘army’. Hence, gist memory is useful for recalling an event, but it is error prone. Finally, if neither gist nor verbatim traces are retrieved in memory, participants can guess about what was experienced if necessary, because they have no memory of it. One particularity of verbatim traces, relative to gist traces, is that they are not very persistent in time (Seamon et al., 2002) or to new stimuli that could easily interfere with them (Abadie & Waroquier, 2020; Abadie et al., 2013; 2017). In consequence, when retrieving information after a few minutes, we mainly rely on gist memory. This explains why under some conditions, false memory rate is low and in others it is high. Indeed, if we have enough verbatim traces available in memory, like it is the case in short term memory tasks, it could counter gist memory and allow to correctly reject a certain number of distractors. On the contrary, if gist memory is very strong, like it is the case in DRM-like tasks, and that verbatim memory is poor because it faded with time, false memories will more likely occur. Indeed, with verbatim memory being less strong than gist, it won’t allow correct rejection of related distractors, hence, it won’t counter gist activation, and the chances of false memories will increase. In DRM-like tasks, it is the presentation of meaning-related items that provokes strong gist memory activation, relative to the presentation of unrelated stimuli. This is why participants produce many correct but also false memories in this type of task.

2.3.2 Conjoint recognition paradigm.

One of the FTT’s strong suit, which made it more advanced than other models, is that a multinomial processing tree model called ‘conjoint recognition’ was developed quite early to estimate gist and verbatim processes underlying false memories in recognition (Brainerd et al., 1999). A simplified procedure was then created by Stahl and Klauer in 2008. In this model, the

estimation of gist, memory and guessing is based on the responses given to the subjects following the presentation of a recognition probe. Following Stahl and Klauer's procedure (2008), three types of probes can be presented during a recognition test following stimuli exposure, like DRM lists: target probes, which are part of the studied material, related probes, which are not part of the studied material but related to it, and unrelated probes, which are non-studied stimuli without link to the studied items. During recognition, participants are asked to identify the nature of each probe type. They therefore can classify them as 'targets', 'related' or 'unrelated' probes. Hence, for each probe type, there can be three responses. Only one of them is correct (i.e., target response to target probe, related responses to related probes or unrelated response to unrelated probes). The rest of the responses are errors. When participants identify related or unrelated probes as 'targets', they produce false recognition of such probes. False memories are considered here to correspond to false recognition of related probes, because the probe shares a semantic relationship with the studied stimuli. According to the multinomial processing tree, which is based on the FTT, correct identification of targets and related probes can be based either on gist, verbatim memory, or on guessing parameters. False recognition of related probes can be based either on gist memory or on guessing parameters. The equation of those parameters can be found in Sthal and Klauer's (2008) article. All 'unrelated' responses to target, related and unrelated probes, as well as 'target' and 'related' responses to unrelated probes, are thought to be based on guessing parameters because this model makes the simplified assumption that unrelated items or 'unrelated' responses do not trigger either 'gist' nor 'verbatim' traces in memory. A figure of this model can be found in Chapter 4. Brainerd et al.'s model (1999) was used in several studies and showed that false memories were associated to higher levels of gist memory than guessing (see Brainerd et al., 2021 for a review).

2.3.3 Development of false memories explained by the FTT.

The FTT explains the increase of false memories with age by showing that both gist and verbatim memory develops with age (see Brainerd & Reyna, 2002a; Brainerd & Reyna, 2015, for reviews). Between early childhood and adulthood, our semantic knowledge about items builds up, as well as our ability to form connections between stimuli, which becomes more automatic (Bjorklund & Hock, 1982; Bjorklund & Jacobs, 1985; Bjorklund & Muir, 1988). As for verbatim, our ability to retrieve traces of specific surface features also improves with age (Brainerd et al., 2002c; Reyna & Kiernan, 1994). In consequence to both gist and verbatim increases, true and false memories increase during development, especially in tasks triggering

a strong meaning processing like DRM tasks (see Chapter 1), gist development being responsible for the increase of false memories.

Gist manipulation.

Many studies support the idea that gist memory is responsible for false memories. In fact, a recent study from Brainerd and colleagues (2020) showed that gist scores attributed to DRM lists, were predictive of false recognitions of the critical item. In this study, participants were asked to rate how strongly the words within DRM lists were related using a 6-point scale, going from ‘highly unrelated’ to ‘highly related’. Results indicate that the higher the scores were, the more false memories were produced. Besides, when experimental designs managed to adjust gist memory to different age group (5-year-olds to adults), by presenting lists adapted in meaning to each group (i.e., Carneiro et al., 2007), the usual developmental trend was reduced, supporting an explanation of the development of false memories with age by gist trace differences across age groups. The presentation of contextualized paradigms, in which related words are embedded in sentences or in scripted visual events, which favors the processing of gist, also reduced the developmental trend on false memories (see Chapter 1, for more details). Moreover, gist cuing manipulation aims at favoring the gist extraction and processing from the studied material. In some studies, it was done by instructions prior to the study phase (e.g., Brainerd et al., 2006; Brainerd et al., 2008c; Holliday et al., 2008; Lampinen et al., 2006) or at retrieval (e.g., Odegard et al., 2008), where participants could be asked to try to extract the common meaning between words or where they were given the category label relating the words of a list. In one study, gist cuing was favored by the presentation of lists that were found to more automatically trigger the retrieval of the critical item, thanks to a pre-test where participants had to find the critical item related to the presented list (Carneiro et al., 2009). Results of studies on gist cuing indicate that the overall rate of false memories increased and age differences in false memory production were reduced under gist cuing conditions. This reduction in age difference is likely due to the fact that young children, who do not usually spontaneously process the gist of studied words under non-cuing conditions, benefitted more from gist cuing than older children or young adults. Altogether, it supports the idea that gist memory is responsible for false memories.

Verbatim manipulation.

Other studies manipulated verbatim processing during encoding and showed that it decreased the chances of false memory. Indeed, in a study where 7- and 11-year-old children were asked to complete the last letter of DRM-lists words, which favors the processing of surface details, false memory rates decreased in both age groups in comparison with a classical encoding condition (Holliday et al., 2011). A meta-analysis from Huff et al. (2015), on 9 studies led to a similar conclusion (see also Brainerd et al., 2002d). One remark though regarding those studies, is that in the process of making a more featured distinctive encoding of stimuli, it might also have potentiated the treatment of the item's meaning. Howe (2008) solved this confounding issue by using a protocol in which he orthogonally manipulated gist and verbatim processing. To do so, he varied the background of the studied image presented to children of ages 5, 7 and 11, on two axes. The first one was congruency (i.e., gist manipulation). The background could be either meaning-congruent or incongruent with the primary content of the image depicting a DRM word (e.g., the background could be a refrigerator when congruent with the word apple or an aquarium when incongruent). This manipulation tested the relevance of the background meaning. The second axis was specificity (i.e., verbatim manipulation). The background could be either the same for all the presented words (e.g., a refrigerator for the fruits list) or different for each presented words (e.g., refrigerator, bowl, kitchen table respectively of the different words of the list). It tested whether associating specific features to an item would affect false memories. Item specific backgrounds reduced false memories and eliminated age group differences compared to similar backgrounds. On the contrary, congruency of the background did not impact false memories. It confirms that feature and contextual specificity but not meaning processing reduces false memories. An experiment done with bilinguals also fosters our point (Howe et al., 2008). In this study, participants of different ages (6-year-olds and adults) were presented with words in a recognition test, either in the same language as the words from the study phase or in their other spoken language. When the language did not match between study and recognition phase, participants could not rely on verbatim memory because the surface features of the words were changed. Only the meaning stayed the same. Results showed that false memory rates increased, under language mismatching condition, which corroborates FTT's predictions that verbatim prevents false memories.

2.3.4 Phenomenological experience of false memory.

One of the reasons FTT is influential is because it is a robust theory in that it accounts for many of the literature's results on self-generated false memories. Besides, the phenomenological experience associated to correct and false memories was explained by the

FTT. Indeed, our memories do not all have the same detailed precision. In the literature, there is a distinction between souvenirs that are recollected with their contextual details (i.e., recollection) and memories of stimuli that seem familiar but are not recollected with their encoding context (i.e., familiarity) (Mandler, 1980). FTT is not a synonym of the familiarity-recollection distinction, however, it also helps understand why we have more or less strong and precise memories. Indeed, strong and precise memories such as recollection, are memories of an event composed of both detailed surface and meaning-based traces. On the contrary, a weaker and fuzzier memory, tapping into more familiarity-based memory, will generally be based on gist traces. In the domain of false memories, many studies have used confidence scales to ask participants to specify their phenomenological experience of their memories (e.g., Roediger & McDermott, 1995). Even though false memories are usually reported with higher confidence rates than other errors, both in children and adults (Lyons, et al., 2010), studies show that false memories are not all associated to the same confidence levels (e.g., Hicks & Starns, 2006; Brainerd et al., 2001; Gallo & Roediger, 2003). Some false memories come with strong confidence and are associated to strong remembering (e.g., Brainerd et al., 2003; Roediger & McDermott, 1995), in which participants report details related to the context in which stimuli were allegedly presented (e.g., the sound of the supposedly pronounced word) (Ghetti, et al., 2002; Mather, et al., 1997). Other false memories come with lower confidence levels, are associated to fuzzier, less detailed reports and rather more familiarity-based judgments (e.g., Ghetti, 2008).

According to the FTT, all false memories are caused by gist. However, strong detailed false memories, also called ‘phantom recollection’, are thought to result from more than gist memory (Brainerd et al., 2001). Phantom recollection is described as a gist-based false memory accompanied with a recollective experience (Brainerd et al., 2001). This recollective experience would be caused by the activation of surface details coming from the context of truly lived events (e.g., characteristics of the voice pronouncing the DRM word lists, such as tone, pitch, timbre, etc.) rather than from details specific to the studied stimuli (e.g., the phonemes composing the words of the lists) (Brainerd et al., 2021). A model alternative to conjoint recognition was developed in 2001 by Brainerd et al. (see Stahl & Klauer, 2009 for a simplified version of the model), where they included the estimation of phantom recollection by contrast with fuzzy gist based false memories. Brainerd et al. (2004) used this model and showed that phantom recollection almost tripled between 7- and 11-year-olds (going from 13% to 31%), underpinning an increase of false alarm rates. This rate stabilized between 11- and 14-year-olds

along with the rate of false alarms. In comparison, fuzzy gist-based traces did not vary between 7- and 14-year-olds. In line with those results, Odegard et al. (2008) showed no increase of phantom recollection between 11-year-olds and adults, which suggests that the increase happens mainly between early childhood and adolescence.

2.4 Global matching theory

Spreading activation and the fuzzy trace theory are the most elaborated theories about false memories because their explanation of self-generated false memories is targeted for different age groups, different tasks and modality of test (i.e., recall and recognition). There is however another model of false recognitions called Global Matching (Arndt & Hirshman, 1998; Hintzman, 1988) which is based on another model, called MINERVA II (Hintzman, 1984). Global Matching models (GMM) applies to recognition only, and the development of false memories with age was not their focus, but it stays interesting because it provides a good perspective of how different qualitative features can sum up to create a false memory.

2.4.1 Principles of the theory.

Stimuli comprise a set of features (e.g., the shape, color, size, meaning, etc., for a visual stimuli). According to the GMM, these features are encoded into their own vector in memory. A memory trace is composed of an ensemble of distinct features. During recognition, a probe is presented to participants. The similarities between the probe's features and the features of the traces stored in memory are compared. The more similar features there are between the probe and memory traces, the more the probe is considered as matching to the memory and is therefore likely to be identified as a studied stimulus. This global matching procedure corresponds then to an addition of parallel featured activations in memory, which are based on the similarity between the features of one or several traces and the presented probe. False memories are likely to occur when an unstudied probe shares too many similarities with one or several truly studied item(s) and that in consequence many trace activations occur in memory. In the DRM, false memories are caused by activations of the shared features between each word list and the critical item. Shared activations with all studied words accumulate, leading to a strong matching activation in memory (Zhu et al., 2019). In fact, a study from Arndt and Hirshman (1998) has shown that correct recognition was usually related to the strong match of the probe with a single memory trace, while false recognition resulted from the sum of several small-match trace activations. Indeed, when the number of words in DRM lists were decreased, false recognition rates decreased more than true recognition. Besides, this model explains the effect of knowledge

and beliefs on false memories described first by Bartlett (1932) (see Chapter 1, section 1.1.3). Indeed, according to the GMM, knowledge forms in memory by an accumulation of featured traces. In consequence, events that are repeated and/or share many similarities (e.g., taking breakfast in the morning, seeing dogs in different contexts), form an accumulation of traces in each specific feature vector. They create prototypes. These prototypes create an increased likelihood of accepting a distractor when it shares similarities with it and therefore, have false memories.

2.4.2 The role of contextual features.

In the GMM, there is a distinction between features that relate directly to the stimuli with features relating to the context of the stimuli (Arndt, 2010). For instance, if the list of words ‘navy, soldier, military, infantry, captain, war, uniform’ is pronounced by a computerized female voice, the semantics and the phonological sounds specific to each word will be encoded (e.g., the first word started by the sound ‘na’). The contextualized features, common to all the words such as the timbre, the intensity, the speech rate, the sound pitch, the tone and the intonation of the specific female voice, will also be encoded. When contextual features of a true memory match that of an unstudied probe, it increases the chances of the probe being identified as part of the original true event. In the example I gave, when the word ‘army’ is presented as a recognition probe, like in DRM tasks, if the original female voice pronounces it, the contextual features will be common with the studied words. Hence, there will be an activation of the contextual traces in memory, because they match with the truly experienced event. Additionally, item specific features will be activated (semantic features in the example I gave). This sum of activation between contextual and item features, will make it more likely for the word ‘army’ to be identified as a false memory than if it was pronounced by another voice, because in the latter case, it would not match as well the contextual feature activations in memory (see Figure 2).

Arndt, one of the authors that wrote the most about GMM and false memories, showed this contextual feature similarity effect. Indeed, in several studies, he found that when presenting a probe during recognition in the same font than the one used for the studied DRM words, true but also false recognition rate (on studied and non-studied probes, respectively) increased compared to a condition in which different fonts characterized the studied words and the recognition probe (Arndt, 2006, 2010, 2015; Arndt & Reder, 2003). According to the GMM, when the context of the recognition probe is similar to the one of studied items, there is an

activation of a single trace in memory, composed of an addition of featured activations. By contrast, context differences between the recognition probe and the experienced event, creates an activation of dissociated traces for item and contextual features. The activation of a single trace for both item and contextual features causes a larger signal activation, explaining hence an acute chance of accepting a stimulus as studied. This effect is called ‘interactive cuing’ (Clark & Gronlund, 1996; Hicks & Starns, 2006a). A recent study from Zhu et al. (2019) supports this effect, by showing greater matching activations in the visual cortex and the left hippocampus when the recognition probe is presented in the same modality as during encoding (i.e. visual modality) instead of different modalities (i.e. auditory for encoding and visual for recognition).

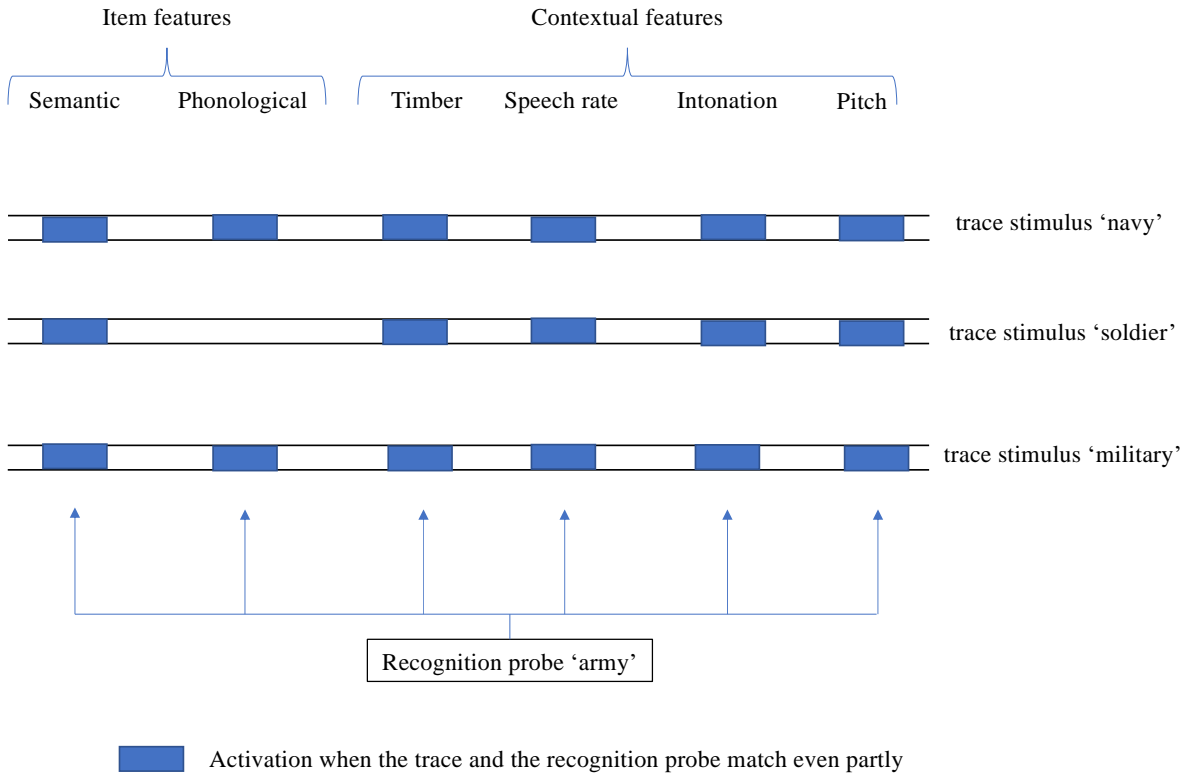


Figure 2. Example of the process of global matching in DRM tasks. When the word ‘army’ is presented as a probe during recognition, the item specific and item contextual features shared between the probe and the memory traces created during the study of each word list (navy, soldier, military) activate in memory. The more features are activated, the stronger a trace for a word is. For instance, trace strength is stronger for ‘military’ than ‘soldier’ in this figure. The activation strength of all traces sums up. The bigger this sum is, the more likely the probe will

be identified as part of the original list. The presented features are not exhaustive. They are only examples to illustrate the global matching recognition system.

2.4.3 Advantages and limits.

One strong point about the GMM is that it accounts for all sorts of false recognitions, such as false recognition of semantically related probes but also of more abstract material, such as non-words (Zeelenberg et al., 2005) or abstract images (Koutstaal et al., 2003) and more generally to a lot of diverse contextual recognition conditions (e.g., memory for odors, for sounds, for language, for entire scenes or specific stimulations). Moreover, it reminds the importance of context in memories. Note however that it does not focus on explaining the development of false memories with age and it does not account for false recall. Moreover, this multi-component model is qualitatively detailed, but it seems too complex to find a way to model it computationally, to measure the specific traces and feature activations on which recognition relies. Besides, literature on GMM lacks precisions about the model. For instance, it lacks examples and definition of what features refer to. In this manuscript, I gave an example of visual or auditory features, based on my understanding of the model, but in reality, the model usually represents the different features as abstract, naming them with numbers rather than qualities. Some clarifications in the future are needed. Testing the model could help to do so.

2.5 Theoretical confrontation

2.5.1 Similarity.

Categorical versus associative similarity.

Spreading activation, Fuzzy-trace and Global-matching Theories share the central idea that similarities between concepts is the primary cause of false memories. In spreading activation theories, similarity is determined by the strength connections between conceptual nodes. In FTT, it comes from the activation of meaning-based gist traces. In GMM, it comes from shared features between a probe and memory traces.

Some studies introduced the idea that the concept of similarity explained by gist traces was different than when explained by spreading activation. Indeed, gist memory would reflect categorical relationships (i.e., category-based co-occurrence or membership, that are based on shared primitive features and share a similar meaning, e.g., farm animals, have skin, have legs, breathes, etc.). Spreading activation, in particular the AMT, would be based rather on

associative relationship (here, based on lexical co-occurrences, e.g., cereals and bowl are words that often co-occur because they are used in a similar context). GMM would not make such distinction (Coane et al., 2021, for a review). Some studies have showed that associative relationship caused more false memories than categorical relationships, by comparing DRM lists with purely categorical lists (Howe, 2006; Howe et al., 2009; Hutchison & Balota, 2005). The problem with this comparison is that, as mentioned in Chapter 1, DRM lists are composed of both categorical and associative relationships. In fact, in reality, there seems to be an overlap between the definition of categorical and associative relationship in that some lexical co-occurrences sometimes correspond to categorical relationships (e.g. at the farm today I saw cows and goats). This confound does not help to understand which type of relation to the critical item (categorical or associative) favors the occurrence of false memories. Other studies compared the production of false memories in controlled mixed lists and in pure categorical or associative lists, in which for the latter, the absence of categorical associations was controlled. Those studies confirmed that false memory rates increased in mixed lists compared to categorical or associative-non-categorical only lists (Coane et al., 2016, 2020; Hutchison, 2003; Montefinese et al., 2015). Hence, it seems that both categorical and associative-non-categorical relationships can lead to false memories and that when both types are mixed in a list, chances of false memories are boosted. This effect was called the associative boost effect and was found in children and adults (see Chapter 1.2, for studies showing a developmental increase in false memories with age on DRM mixed-lists and categorical lists).

The first limit in my sense with comparing categorical with associative relationships as I already mentioned, is that associative relationship can comprise some categorical relationships, making their distinction not very clear. Secondly, I do not think that the FTT's definition of gist memory is limited to categorical associations, contrary to the studies I mentioned above. Indeed, there can also be a meaning-based relationship between two contextual items (e.g. the common gist between 'cow' and 'tractor' is that they both belong to the farm). Moreover, as described in section 2.2, spreading activation theories acknowledge the development of categorical networks based on semantic similarities, which starts to be built during childhood through cultural learning. Therefore, even though it is informing to look at whether categorical versus associative relationships are the primary cause of false memories, which informs in a larger way on how knowledge is organized in memory, I am not sure that it would really allow us to decide which theory better explains false memories. Besides, note that both theories use similar

material in their experiments (i.e., DRM lists) and ultimately, make similar predictions regarding semantic false memories (see Chang & Brainerd, 2021).

Regarding GMM, some would say that because it does not make the specific distinction between categorical and association relationships, it could account for both types, making it a good predictor of false memory occurrences. However, the fact that this model is broadly defined could also be considered as a limitation, in that it could explain all accounts of false memories, but at the same time crucially lack some precisions to allow any future predictions.

Similarity indicator.

In the AMT, one of the most used indicators of similarity is the ‘backward associative strength’ (BAS) score. This score represents the probability that an individual would produce the critical item as a first response when presented one of the words of the list. Each word of the list has a distinct probability, and all BAS scores are averaged to produce a single mean BAS score. According to some authors, this score reflects lexical level associations such as linguistic co-occurrence (Coane et al., 2020), which rejoin the definition of associative relatedness. BAS scores were shown to be a great predictor of false memory occurrence. Indeed, the higher BAS scores are, the more likely the critical item associated to a list will be accepted as studied (e.g., Deese, 1959a; Roediger et al., 2001; but see Brainerd et al., 2008d). A recent study from Brainerd et al. (2020), already mentioned in section (2.3.3), proposed another predictor of false memories of DRM lists- i.e., ‘gist strength score’. Gist strength scores refers to judgment of the relatedness between the words of a DRM list. By contrast with the BAS score, it would reflect an indicator of the global semantic meaning created by the ensemble of the words of a list. Somehow, comparing BAS scores with gist strength scores seems to be another way of looking at whether associative versus categorical associations are better predictors of false memory occurrences in the DRM. Brainerd et al. (2020) showed that this gist strength score was a better predictor of false memories than the ‘BAS’ score. Indeed, BAS predicted false memories only of lists having low gist strength scores. These results could be the argument that similarity is after all not the same for the FTT and spreading activation theories. Nonetheless, it is worthwhile noticing that whilst BAS and gist scores measure different things, they both exist as concepts in each theory.

Indeed, in the FTT, gist scores as described by Brainerd et al. correspond to the extraction of the global gist of a list, common to all the words of a list, while BAS scores could be

compared in the FTT to item specific gist, individual to each word of the DRM list. In spreading activation theories, BAS scores would be the reflection of the individual activation of an isolated word with the critical item, while gist scores would be the result of simultaneous word activations on a memory network that would influence the activation of a critical lure altogether. In fact, gist scores can be assimilated to a similar calculation in the spreading activation theories, called ‘interim associative strength’ (Deese 1959b) or ‘connectivity’ (McEvoy et al., 1999; Roediger et al., 2001). This type of score is however not usually used as a reference of associative strength contrary to BAS scores. Hence, BAS and gist scores do not measure the same things but seem important and complementary to understanding false memory chances. Besides, congruently with both theories which agree on the idea that both item specific and between item meaning activation contribute to false memories, Brainerd et al. (2020) have shown additive effects of BAS and gist scores on false memories. Indeed, those errors were boosted when gist and BAS scores were both high, showing a contribution of both indicators in false memories. Note that this finding is in line with the associative boost effect found in some studies (e.g., Coane et al., 2016, 2020). It therefore reinforces the idea that both associative and categorical relationships, which are closely interrelated, contribute to false memories.

2.5.2 The limitations of contextual features matching effect on false memories in global matching theory.

Arndt (2010) insists that GMM clearly explains the context congruency effect on false recollection experiences (Brainerd et al., 2014). Even though Arndt showed that false memory increased when the contextual features of studied items and recognition probes matched, other studies showed opposite results. For instance, in Howe et al.’s study (2008), that I presented in section 2.3.3, bilingual participants produced more false memories when the recognition probe was in a different language than the studied items. Similarly, in Zhu et al.’s study (2019), participants produced more false memories when they were presented auditory stimuli to study and visual stimuli recognition probes, rather than when they were presented visual stimuli at both stages. These result discrepancies between studies suggest that there might be at least another factor to take into account in the equation. In Howe et al. (2008) and Zhu et al.’s (2019) studies, note indeed that not only contextual features varied between mismatching conditions, but item specific features were also different between study and recognition. As described earlier, it was when participants could not rely on those item specific details, that they failed to properly reject a distractor or to accept a target. Hence, it seems that similar contextual cues

between study and recognition could increase false memories as shown by Arndt, but only when item specific features are also invariant (e.g., same modality than from encoding). As soon as those features vary between study and recognition, it increases the likelihood of familiarity retrieval and increases in turn the chances of false memories (Yonelinas, 2002, for a review).

2.5.3 Conclusion.

To conclude, the aim of this Chapter was to present the existing theories to explain self-generated false memories. The different perspective of each theory is interesting and rather complementary. In the experimental part of the thesis, we based our studies on the FTT. It is indeed a robust model accounting for the majority of the findings on self-generated false memories, in particular in developmental tasks. It offers qualitative precisions about memory traces by emphasizing the existence of specific detailed surface traces and of meaning-based traces, which explains why we have more or less precise memories. Moreover, the existence of the conjoint recognition model, allows to make the parallel between responses in recognition and underlying memory traces (i.e., gist or verbatim), allowing a more direct access to the explicative processes of false memories. In the next Chapter, I will talk about false memories in WM tasks, an emerging topic in the literature.

Summary

In this Chapter I first presented the principal factors determining the correct retrieval of an episode in memory, at each step of the 'life' of a memory- i.e., encoding, consolidation, storage and retrieval. False memories are built on the success and failures of each step. I then presented explicative theories of self-generated false memories. According to spreading activation theories, false memories are the consequence of the activation of a network composed of conceptual nodes. During retrieval, the activation of conceptual nodes in memory relating to true experience, spreads to its closest associates, leading in turn to false memories of related nodes. The activation monitoring theory proposes that when we are not able to identify the source of activation of the associates (i.e., the self), false memories are the most likely to occur. Spreading activation theories explain the development of false memories by the development of conceptual networks with age, which increases the automaticity of activation of both studied and their related nodes. According to the FTT, false memories are caused by traces in memory called 'gist', which relates to the processing of the general meaning of stimuli. Opposed to gist, verbatim memory prevents false memories by allowing a rejection of the distractors because of its specific details. One advantage of the FTT is that a multinomial processing tree model was developed to make a direct link between false recognition and the underlying processes responsible for it. The development of both gist and verbatim memory between early childhood and adulthood explains an increase of correct responses with age in memory tasks, but also the increase in false memory with age. According to Global matching models, memory traces are composed of several sensory and semantic features. The presentation of a recognition probe activates memory traces based on its similarity to each feature. Those activations sum on one or several traces and the chances of false memories are proportionate to this sum. This model brings insightful evidence that contextual features can contribute to correct recall but also to false memories. However, it accounts only for false recognition and it was not applied to varied age groups. There was a debate between spreading activation theories and the FTT regarding the definition of similarity between items, as the cause of false memories. The former theory defends associative relationship, while the latter represents categorical relationship. Even though there is an overlap between both definitions, there was evidence that both types of relations contributed to false memories and even created an additive effect on it. This was also confirmed by the comparison between BAS and gist strength scores, two indicators of each theory respectively. Finally, I discuss the limit of the effect of contextual matching between study and recognition found by global matching model.

Résumé

Dans ce chapitre, j'ai présenté les déterminants principaux de la récupération d'un épisode en mémoire à chaque étape de la 'vie' d'un souvenir- i.e., l'encodage, le stockage, la récupération. Les faux souvenirs sont construits sur le succès et les échecs de chaque étape. J'ai ensuite présenté les théories explicatives des faux souvenirs auto-générés. D'après les théories de l'activation par diffusion, les faux souvenirs sont la conséquence de l'activation d'un réseau composé de 'nœuds conceptuels'. Durant la récupération, l'activation de ces nœuds conceptuels en mémoire qui se réfèrent à une expérience vécue, se propage aux nœuds associés les plus proches. Ceci conduit à des faux souvenirs sur les concepts représentés par ces nœuds associés. La théorie de l'« activation monitoring » précise que lorsque nous ne sommes pas capables d'identifier la source d'activation de ces nœuds associés (i.e., soi), les faux souvenirs ont le plus de chances de survenir. Les théories de l'activation par diffusion expliquent le développement des faux souvenirs par le développement des réseaux conceptuels avec l'âge, ce qui augmente l'activation automatique d'items étudiés ainsi que les nœuds qui leur sont reliés. D'après la théorie des traces floues, les faux souvenirs sont causés par des traces 'gist ' en mémoire, qui font référence au traitement du sens général des stimuli. A l'inverse, la mémoire verbatim empêche les faux souvenirs, en permettant le rejet des distracteurs grâce à ses détails spécifiques. L'avantage de la théorie des traces floues, est qu'un modèle multinomial a été développé pour faire le lien direct entre fausse reconnaissance et les processus sous-jacents associés. Le développement de la mémoire gist et verbatim entre la jeune enfance et l'âge adulte explique l'augmentation des réponses correctes mais également des faux souvenirs avec l'âge. D'après les modèles de 'global matching', les traces en mémoire sont composées de plusieurs traits sensoriels et sémantiques. La présentation d'un item à reconnaître active des traces en mémoire basées sur sa similarité avec chaque type de trait. Ces activations s'additionnent pour une même trace ou un ensemble de traces et les chances de faux souvenirs sont proportionnels à cette somme. Ce modèle apporte des éléments en faveur de l'implication des traits contextuels dans la reconnaissance correcte ainsi que dans les faux souvenirs. Cependant, ce modèle ne s'applique qu'à la reconnaissance et n'a pas été appliqué à différents groupes d'âges. Il y a un débat entre les théories d'activation et la théorie des traces floues à propos de la définition de la similarité entre items en tant que cause des faux souvenirs. La première défend une relation associative alors que la dernière représente une relation catégorielle. Bien qu'il y ait un

recouvrement entre ces deux définitions, il a été montré que les deux types de relations contribuent aux faux souvenirs et créent un effet additif sur leur chance d'apparition. Ceci a été également confirmé par la comparaison de deux indicateurs de chaque théorie respective : les scores BAS et gist. Enfin, je discute des limites de l'effet de 'matching contextuel' entre phase d'étude et de reconnaissance montré par les modèles de 'global matching'.

Chapter 3. False memories in working memory.

Most of the work done on false memories used memory tasks that could be qualified as ‘long-term memory’ tasks (see Chapter 1 and 2). In this Chapter, after specifying different memory systems, I will describe two of the most influential models of working memory, namely the Embedded-process model (Cowan et al., 2021) and the Time-Based Resource sharing model (Barrouillet & Camos, 2021). The difference between long-term and working-memory tasks will then be detailed. The sources of WM development during childhood will then be presented. In the last part, I will introduce the few studies that assessed false memories in WM tasks in young adults. Some studies particularly tested the effect of WM maintenance mechanisms on false memories in immediate and delayed tests. Other studies have looked at whether there was a dissociation between false memories in immediate and delayed tests. Finally, I will present evidence showing that false memories in WM tasks are based on gist memory and review the two studies that assessed phenomenological experience associated to short-term false memories.

3.1 Different memory systems

Atkinson and Shiffrin (1968), distinguished between three distinct stores in memory, which are mainly ruled by temporality and storage capacity. According to their view, all the sensory detailed information captured by our senses during encoding goes through sensory memory during a few milliseconds. A very small amount of that information then transfer to short-term memory (i.e., 7 ± 2 , Miller, 1956) where it can be held for a few seconds for the purpose of the ongoing task. After a few seconds, some of that information then transfers into long-term memory, a storage system containing episodes of one’s life composed of detailed to more abstract memory traces. Hence, those systems are used successively in time, serving different purposes and can hold different amounts of information.

Short-term memory was first described by Atkinson and Shiffrin (1968) as a passive automatic system transferring information into LTM, researchers then rapidly discovered that there was a more sophisticated system called ‘working memory’. This system is consensually acknowledged as allowing to voluntarily maintain and process information for a few seconds (e.g., Baddeley, 2000; Barrouillet & Camos, 2022; Cowan, 2017; Daneman & Carpenter, 1980).

Most daily life activities require WM (e.g., calculating the price of a basket at the grocery shop, taking a class, making a presentation, cooking, having a conversation, etc.). Currently, the term short-term memory still exists and is used as its original definition (i.e., a short-term passive store). It is differentiated to WM, which is thought to rely on attentional resources to allow active maintenance and processing of information (Cowan, 2017, see Figure 3 in section 3.1.1 for an illustration of WM and short-term memory tasks). Besides, some views consider that WM encompasses short-term memory (Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974).

Moreover, WM and LTM, work bidirectionally. Indeed, information going through WM will define the one going into LTM, and conversely, information already stored in LTM can impact the maintenance and interpretation of information in working-term memory. For instance, if a cow is part of the movie scene I just watched, this stimulus will go to my short-term memory to then be stored in my LTM as part of a lived episode. At the same time, the semantic and associative information related to the cow from past lived episodes, stored in my LTM, are activated and impacts the way I will process the actual ‘cow’ stimuli in my short-term memory. Several effects of LTM’s influence on short-term memory have been examined and shown in the literature. For example, words were shown to be better recalled at short-term than non-words (i.e., lexicality effect, Roodenrys et al., 1993) or than words in a foreign language (e.g., Hulme et al., 1991). Indeed, when the studied material has a known meaning, participants can rely on LTM representations, which facilitates retrieval. Another example is the frequency effect (Hulme et al., 1997; Kowialiewski & Majerus, 2018a; Poirier & Saint- Aubin, 1996; Saint-Aubin & Leblanc, 2005; Stuart & Hulme, 2000; Watkins & Watkins, 1977), in which words encountered frequently in our current life are better recalled after delays of a few seconds than words encountered less frequently, as their LTM representation is stronger. There is also the imageability effect (also called concreteness effect), in which words that can be more easily represented by a mental image are also more easily recalled (e.g., Acheson et al., 2010; Bourassa & Besner, 1994; Campoy et al., 2015; Miller & Roodenrys, 2009; Romani et al., 2008; Walker & Hulme, 1999). A final example, that is of particular interest for this thesis, is the semantic similarity effect. Indeed, when words are semantically related either categorically (e.g., eye-toe) or associatively (e.g., black-eye), the rate of immediate correct recall was found to be higher than when studied words were unrelated (e.g., Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier, 1999b; Poirier et al., 2011). Those effects illustrate well that short-term memory relies on knowledge stored in LTM for retrieval cues during recall.

3.1.1 Different tasks for different systems

Some tasks are qualified as ‘WM tasks’ in opposition to LTM tasks or even to short-term memory tasks. A WM task captures the effect of maintenance and processing done within WM on recall or recognition. It does not imply that only WM is involved in WM tasks however (see the previous section for examples of the influence of LTM on recall in WM tasks).

To discriminate between LTM and WM tasks, two notions are important. The first one is the number of stimuli to be memorized in the task. In WM tasks, because maintenance capacity is limited, not more than 8 stimuli are usually presented. This number is nonetheless modulated by the nature of stimuli to be processed (e.g., their length, their meaning, etc.) and the possibility of chunking. In LTM, the number of stimuli to be processed is not limited, because it is not bound by the quantity of information that can be maintained and usually goes beyond 8 stimuli. Classical DRM tasks are a good example of LTM tasks where the quantity of information to be learned exceeds the WM maintenance capacity. All items cannot be maintained in WM by attentional resources or by the use of the phonological loop.

A second aspect to take under consideration when comparing LTM and WM tasks is the notion of temporality. Traces decline with time in memory (Unsworth & Engle, 2007). Usually, maintenance mechanisms prevent this decline, but only for a few seconds (Cowan 1992). In a situation where articulatory rehearsal could maintain all the relevant stimuli for the task, as long as it is not interrupted, it could theoretically keep intact the traces of interest. But in real-life, new information usually comes to disrupt this maintenance process, which is why information can be maintained only for a few seconds. Hence, in WM tasks, the memory test is usually presented a few seconds after the end of the presentation of the last stimuli to-be-learned (i.e., immediate memory tests, a term used interchangeably with ‘short-term memory’ measure). Note that some recent studies have also taken interest in the effect of maintenance in WM at longer delays. They have therefore implemented a delayed memory test (recall or recognition) in their protocol, which occurred minutes after item encoding (e.g., Abadie & Camos, 2018; Abadie & Camos, 2019; Camos & Portrat, 2015; Loaiza & McCabe, 2013; Loaiza & Camos, 2018; Olszewsha et al., 2015). These tests are referred to as ‘delayed’ memory tests. In some studies, they also have been described as a measure of ‘long-term memory’. It is important to note however that even though delayed tests might measure ‘long-term memories’ when they are used in LTM and in WM tasks, both measures cannot be compared so easily, because they are associated to different tasks to begin with.

Finally, as already mentioned in the previous section, the term ‘WM task’ is usually used in the literature for tasks that involve both maintenance and processing of information (e.g., Daneman & Carpenter, 1980). They are usually designed as dual tasks where a concurrent task requiring processing comes during the retention delay of a few seconds between or after the presentation of the stimuli to memorize. This configuration allows the reproduction of real life situations where dual tasks are imposed (e.g., taking notes during classes requires switching between listening to the teacher, and taking notes; estimating the cost of your basket at the supermarket requires to switch between maintaining the items to buy in memory and calculating the total price of these items). The Brown-Peterson paradigm (Brown, 1958; Peterson & Peterson, 1959) and the complex span task (Case et al., 1982; Daneman & Carpenter, 1980; Turner & Engle, 1989) are two most commonly used paradigms to measure WM. In the Brown-Peterson paradigm, stimuli to be learned are presented sequentially in time (or sometimes simultaneously). Following this occurs a retention interval of a few seconds, during which a concurrent task is implemented before the introduction of a recall or recognition test. This concurrent task introduces concurrent processing and depending on the nature of the task, can impair maintenance mechanisms of WM (see the next section for more details). The complex span paradigm differs from the Brown-Peterson paradigm in that the concurrent task lasting a few seconds is introduced in between the presentation of each stimulus (Figure 3). The main distinction between WM and short-term memory tasks, is that in the latter, no concurrent task is implemented during retention intervals. Moreover, in classical short-term memory tasks, recall or recognition follows the presentation of the last to be studied item directly. Hence, there is usually no retention interval. In some protocols, however, recall or recognition can occur a few seconds after the presentation of the last stimulus. It could be done for instance, to make a direct comparison with Brown-Peterson or Complex span tasks.

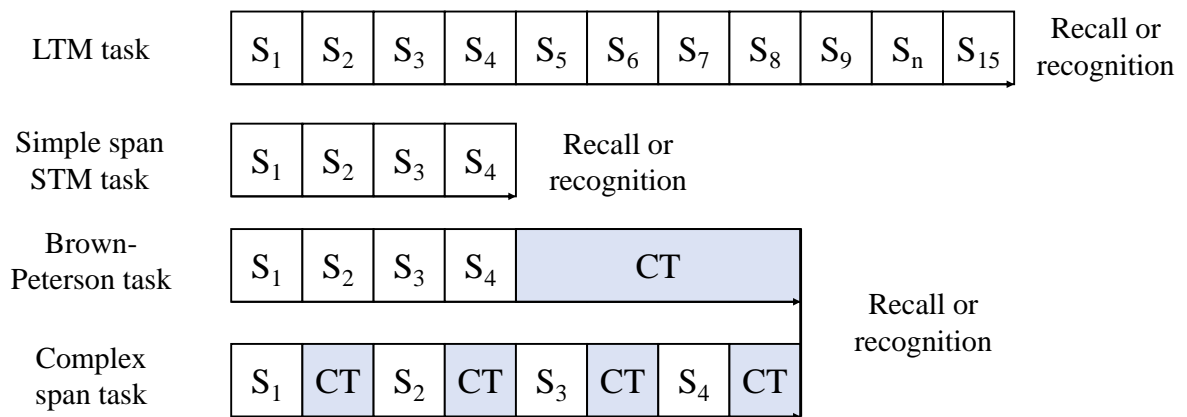


Figure 3. Depiction of typical long-term, short-term, working memory Brown-Peterson and complex span tasks. S = stimulus, CT = concurrent task, in STM = short-term memory. Arrows represent timeline.

3.1.2 Models of working memory

There is variety of existing models to explain and define WM (see Cowan, 2017; Logie et al., 2021, for reviews). I will introduce two of the most influential actual models of WM in cognitive psychology. The first one is the ‘Embedded-Process’ model of WM (Cowan et al., 2021), which adopts a single view of memory, where WM would be an activated part of LTM. By contrast, the second model called ‘Time-Based Resource Sharing’ model of WM (Barrouillet & Camos, 2021), supports a dual view of memory, in which WM is thought as a system distinct from LTM, even though both systems communicate.

Embedded-Process-Model of working memory.

In this view, WM is defined as “the ensemble of components of the mind that hold a limited amount of information temporarily in a heightened state of availability for use in ongoing information processing” (Cowan et al. 2021, p.45). In this model, WM is described as functionally distinct from LTM but structurally embedded in it (Cowan, 1995). Indeed, according to this view, stimuli enter first in sensory memory for a brief delay of a few milliseconds. There is then an activation of the corresponding information in LTM. A part of this activated information is processed and maintained by the focus of attention. The focus of attention has indeed a very limited capacity and studies indicate that it can maintain and process three to four pieces of information at once (Cowan, 2001, for a review; Sperling, 1960). Hence,

not all the activated part of LTM is kept in the focus of attention. All information that is not in the focus of attention declines within a few seconds in memory. Hence the information that will be recalled is the one maintained in the focus of attention (Cowan et al., 1999). Nonetheless, the information that declines outside of the focus of attention can be reactivated and refreshed if it is brought back into the focus of attention, as long as they did not completely decline. Information within the focus of attention is associated to a new LTM trace (Cowan, 2019). The focus of attention facilitates association of elements in the activated part of LTM. For instance, a serial position is associated to each element maintained within the focus of attention. It helps consolidating the trace, which will stay accessible in LTM for a longer time. Moreover, the information that will be maintained and processed by the focus of attention will depend on the demands of the task as well as the nature of the stimuli. Indeed, some stimuli will capture the focus of attention spontaneously, whereas in other cases, the focus of attention will be voluntarily directed towards some specific stimuli. In the second case, the central executive, which relies on executive functions, controls and directs the attentional resources on the stimuli to be kept in the focus of attention (Figure 4). Moreover, the central executive allows the focus of attention to be directed either on external environmental stimuli or on internal LTM traces.

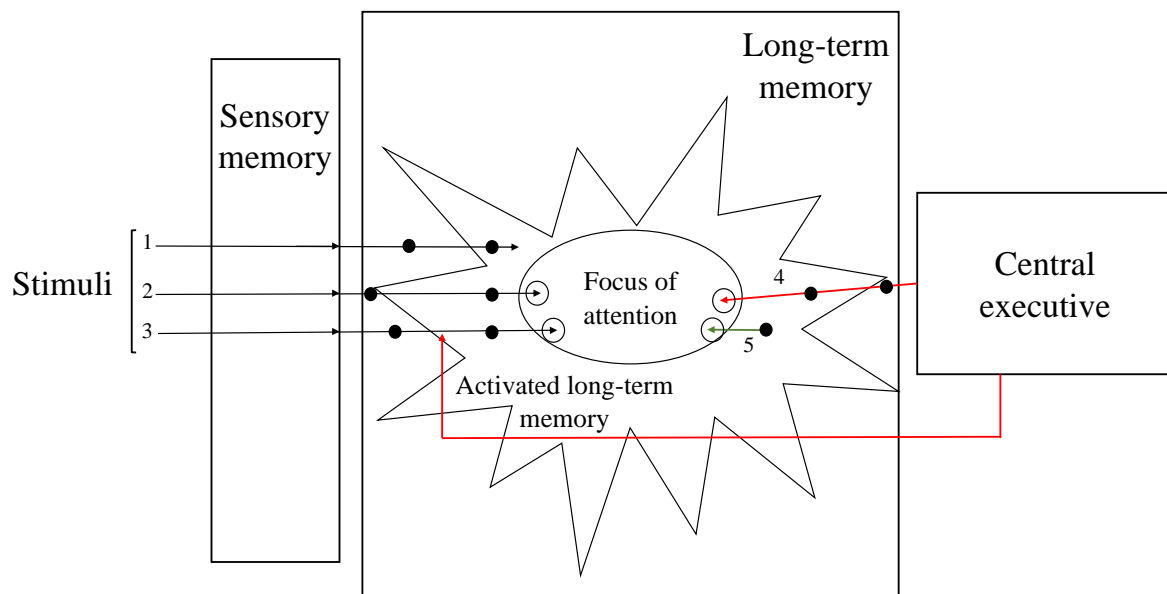


Figure 4. Illustration of the memorial system in the Embedded-Process model. 1: Stimulus left out of the focus of attention. Its activation decays over time as it does not receive attention. 2 and 3: Stimuli entering the focus of attention. Can be maintained and processed. 3: deliberately attended stimulus. Supervised by the central executive. 4: information deliberately retrieved

from LTM. This activation is controlled by the central executive, which directs attention on LTM information. This information then enters the focus of attention. 5: Automatic association that attracts attention (e.g. reading the words ‘cow’ and ‘cheese’ could automatically activate the associate ‘milk’, which would then enter the focus of attention). Adapted from Cowan (1988, 1995, 2001).

Besides, Cowan has shown that the three to four pieces of information that could be maintained in the focus of attention were not necessarily items, but rather chunks of items. Chunks represent the grouping of items that can be made based on previous knowledge (e.g., acronyms such as LTM, WM; words that share a common meaning such as boat and sail). Studies have indeed manipulated the size of chunks by varying the number of words presented per chunks (i.e. single word or pairs of words) in lists of words to learn. They have shown that the number of words per chunks in lists did not determine the number of chunks recalled (Chen & Cowan, 2009; Cowan et al., 2012). Hence, individuals are able to keep in the focus of attention three to four chunks, even if the size of a chunk varies, as long as they have previous knowledge to form those group of items. Moreover, even though it was shown that 3 to 4 chunks can be maintained by the focus of attention, the capacity of recall in WM tasks sometimes exceeds 4 units (e.g., Miller, 1956). The embedded-process model explains this by the existence of another maintenance mechanism than the focus of attention, namely: articulatory rehearsal. This mechanism was first described by Baddeley in its multicomponent model of WM³ (1986) and acknowledged in most WM models (see Logie et al., 2021). Articulatory rehearsal consists of a recirculation of memory traces in an articulatory loop to preserve its phonological form (Barrouillet & Camos, 2022). In Chen and Cowan study (2009), the role of articulatory rehearsal in recall was in fact assessed in addition to the role of the focus of attention. Indeed, span performance of four chunks in this study was found in a condition where articulatory rehearsal was prevented by the introduction of a concurrent articulation. In this condition, participants had to repeat continuously the word ‘the’ while trying to maintain the words to recall, which prevented them from rehearsing the words. This condition did not however prevent them from using the focus of attention for maintenance. In a condition where

³ The multicomponent model of WM was one of the pioneering models of WM. Many models that emerged after this model were based on it, in particularly the Time-Based-Resource-Sharing Model (Barrouillet & Camos, 2021). In the multicomponent model of WM, Baddeley describes WM as composed of multiple components which allow maintenance and processing of information to be orchestrated for a brief period of time. In this conception WM is thought as independent from LTM. For a more exhaustive description of the model, you can refer to Logie et al. (2021, Chapter 2)

articulatory rehearsal could be used in addition to the focus of attention, because no concurrent articulation was introduced, participants recalled up to 8 words (see also Cowan et al., 2004). It confirms the existence of two maintenance mechanisms of different nature: maintenance via the focus of attention and via articulatory rehearsal. Note that such findings and interpretation apply to verbal material specifically, given that articulatory rehearsal maintains verbal material only or material that can be labeled with words.

Overall, the Embedded-Process model gives a central role to attention in WM for maintenance and processing. Some would say that one limitation of this model is to consider WM as embedded in LTM, rather than emphasizing the existence of some independent transient store for information. Barrouillet and Camos (2022, Chapter 5), question indeed the capacity of LTM in forming and updating constantly new representations. An independent temporary system for those operations to occur in before transferring into LTM, seems more appropriate to them and several other authors (e.g., Baddeley, 1986; Norris, 2019). Moreover, they discuss the idea that the Embedded-Process model is close to a dual process model of WM, without acknowledging the existence of separate entities of WM and LTM. In fact, the Embedded-Process model shares many principles with the Time-Based-Resource-Sharing model of WM, proposed by Barrouillet and Camos (2021).

Time-based-resource-sharing model of working memory

This model defines WM as “a structure where mental representations are built, maintained, and modified according to our goals” (Barrouillet & Camos, 2021, p.85).

The basis of this model is that domain-general attentional resources, are shared in time between processing and maintenance of information. The time spent on information processing is the time during which other information cannot be maintained, because attentional resources cannot be deployed on more than one item at a given t time (Barrouillet et al., 2011; Barrouillet & Camos, 2015; see also Case, 1985; Daneman & Carpenter, 1980; Just & Carpenter, 1992 for a similar view). During processing, traces that cannot be maintained, decay in memory. Therefore, recall depends on the time spent on maintenance and processing. The shorter processing takes, the more time is left for maintenance. Chances of correct recall increase with the increase in time for maintenance. In the literature, this resource sharing has been evidenced by studies in which the duration of a concurrent task, occurring during the retention delays between the presentation of the stimuli to memorize and recall, was varied. Studies confirm that

when processing takes more time, because it captures attentional resources longer, recall performance decreases (Barrouillet et al., 2004; Lépine et al., 2005). More specifically, these studies have shown that when the concurrent task's pace was faster (e.g. read 10 digits compared to four, in the same time lapse), recall of letters (i.e. items to be learned), decreased. Symmetrically, when the number of stimuli to be processed in the concurrent task was equalized and that the duration of the concurrent task varied, performance decreased under shorter duration, as more stimuli had to be processed under a smaller interval of time. It supports the idea that when processing takes shorter time, participants can redirect their attention on maintenance (see also Barrouillet et al., 2007, 2012, 2018). Hence, the ratio between the time needed to process a distracting task and the remaining time allowing the maintenance of memory traces via attentional resources (i.e., cognitive load), is a determinant of WM span performance.

In this model, attentional resources are essential as in the Embedded-Process model. Indeed, for one the executive loop, which could be compared to the central executive in Baddeley's multicomponent model (2000) or Cowan's model, relies on attentional resources and other executive functions (i.e., inhibition, flexibility) to direct the shifting between maintenance and processing. Attentional resources also allow information to be processed and maintained by selecting and manipulating information. For maintenance, a specific mechanism, called 'attentional refreshing' was found to reactivate and refresh memory traces, via the focus of attention (Barrouillet, Bernardin, et al., 2004; Barrouillet et al., 2007; Barrouillet & Camos, 2015, 2021; Barrouillet et al., 2011; Camos et al., 2009). More specifically, it was described as "refocusing attentional resources on temporarily stored information in WM, to restore and preserve their decline with time, with a sufficient activation level to avoid their forgetting" (Barrouillet & Camos, 2021). The existence of this mechanism was also acknowledged by other authors (Baddeley et al., 2021; Cowan et al., 2021, but see Oberauer, 2021). It was evidenced in WM tasks, in which the attentional demand of the concurrent task during the retention intervals was manipulated. Indeed, as described above, when attentional demand of the concurrent task was increased, by introducing a bigger cognitive load (Barrouillet et al., 2004, 2007, 2012, 2018), performance in recall was shown to drop. One particularity of attentional refreshing is that only one item can be refreshed at a time (see also Cowan, 1995; Oberauer, 2003). Hence, the focus of attention moves rapidly and successively from one item to another in time to refresh them. Moreover, the TBRS described articulatory rehearsal as another WM maintenance mechanism (see the previous section for a definition). In this model, the

recirculation of memory traces in the articulatory loop is thought to rely on motor programs. Like for attentional resources, this loop has a limited capacity, which explains why we can maintain only a very small amount of information in WM (Baddeley, 1986). As described in the section on the Embedded-Process model, a simple way of evidencing the role of rehearsal in maintenance is to introduce a concurrent articulation during retention intervals of a WM task. When doing so, participants' performances at the task were found to decrease (e.g. Camos, Lagner & Barrouillet, 2009). Besides, articulatory rehearsal was shown to be independent from attentional refreshing. Indeed, in Camos et al.'s study (2009), attentional demand and articulatory rehearsal were varied orthogonally. Correct recall decreased when attentional demand of the concurrent task was higher as well as when a concurrent articulation was introduced. There were additive effects of the suppression of both mechanisms on performance, where recall rate was the lowest when both mechanisms were suppressed (i.e. strong attentionally demanding aloud task compared to low-attention with concurrent articulation or high attention-without concurrent articulation). There was however no interaction between the manipulation of the two mechanisms. Besides, imaging studies have evidenced that specific cortical activations were associated to articulatory rehearsal (i.e., Borca's area, left premotor cortex, left intraparietal sulcus and right cerebellum). A different activation pattern was found under concurrent articulation (i.e., dorsolateral prefrontal cortex; Gruber, 2001; Raye et al., 2007; Smith & Jonides, 1999).

Architecturally speaking, in the TBRS, representations in WM are thought to be stored in an episodic buffer, which can be considered as a sort of hub, integrating inputs from passive sensory buffers (e.g. phonological episodic buffers, motor buffer, etc.), from LTM, and from representations that are maintained in the phonological loop (see Figure 5). The communication and coordination between these different systems and the episodic buffer is done via the executive loop which allows *in fine* to maintain, process and bind elements in the episodic buffer. More precisely, executive instructions for maintenance and processing are given by a 'production system' and are executed by the executive loop. Note that consensually with Cowan's findings (2001; Chen & Cowan, 2009), evidence indicates that up to four items can be held at once in the executive loop (Langerock et al., 2014; Vergauwe et al., 2014). In this model, contrary to Cowan's Embedded-Process model, LTM is a peripheral system to WM, which can support maintenance and processing of temporary representations in WM (see examples of the influence of LTM on WM's recall in the previous section).

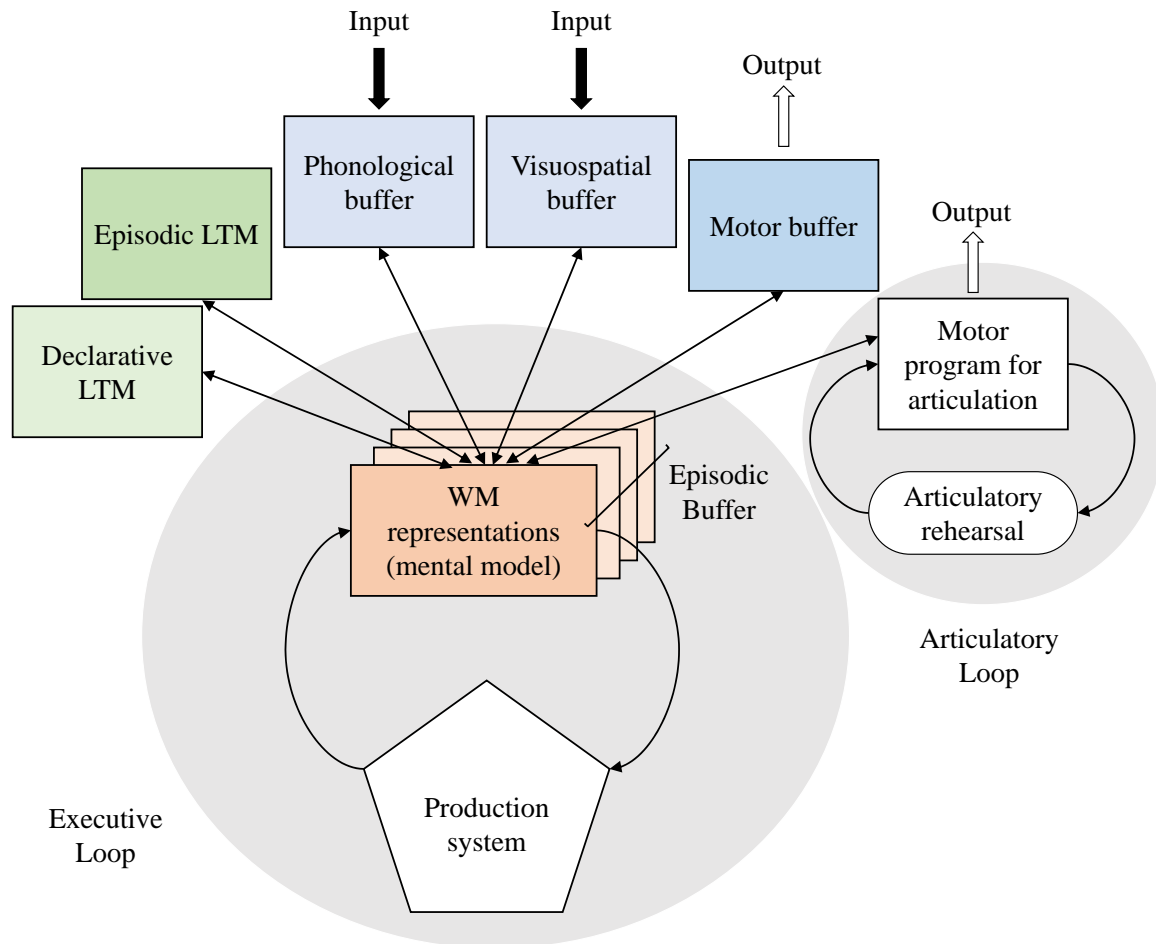


Figure 5. Architecture of working memory according to the Time-Based Resource Sharing Model. Adapted from Barrouillet and Camos (2022). The authors specify that there are probably more peripheral buffers (e.g., auditory, musical, kinesthetic, haptic) than the ones represented in the figure.

Both the Embedded-Process and the TBRS model of WM share some major common principles even if they do not agree on whether WM is an activated part of LTM or a distinct system. First, they share the idea that attentional resources are central to WM. They also share the idea that traces stored in LTM contribute to WM activities. Finally, they both agree that information is maintained in WM via both attentional resources and an independent articulatory system. In this thesis, we do not aim at testing which model is better. Indeed, the common ground between the two models are sufficient for our questions of interest as we are mainly interested in the role of maintenance mechanisms, which are described in both models. Hence, even though the theoretical framework of the studies conducted in this thesis is based on the

TBRS, in particular because it defends the idea that maintenance and processing share in time on a common attentional resource, a premise on which we base our hypothesis and result's interpretation, it does not mean that the Embedded approach could not account for the findings presented in the experimental part of this thesis.

3.2 The development of working memory

The main indicator of WM's growth with age is the increase in the number of items children can immediately recall (i.e., span capacity). Many studies have indeed shown that the span capacity increased with age during childhood. Dempster (1981), have in fact shown, by making a review based on several studies, that children could remember around 2 digits at 2 years of age, 4 digits at 5-years, 5 digits at 7-years, 6 digits at 9-years, 6.5 at 12-years and 7 in adults, when using a short-term memory digit span task, a task in which participants are presented lists of an increasing number of digits and have to immediately recall them. Similar findings were obtained with lists of words, although the increase was milder, ranging from 3 words at 2-years to 5 words in adults (see also, Dempster, 1985; Hitch, Halliday, Dodd & Littler, 1989, for similar findings). A genuine question, here would be to ask what explains the increase in WM capacity during children's development? Camos and Barrouillet' chapters on WM development (2018, see also Barrouillet & Camos, 2022), offer a quite complete view of the factors involved in WM development. I will rely on it, to summarize the different sources of WM's development.

First, one important factor of WM's growth is the increase of processing efficacy with age. Indeed, it was shown that adults were faster than 6-year-olds at enumerating digits and words, suggesting that they processed information more efficiently (Case et al., 1982). This efficiency increase suggests that the number of cognitive resources necessary for a task, decreases with age. In consequence, WM performance increases. In fact, to show additional evidence toward this hypothesis, Case et al. (1982) showed that when word repetition is equated on speed between children and adults by manipulating word familiarity, word span performance was equalized across age groups. There were also similar findings when using a counting task in which participants had to count the number of dots on cards presented sequentially and then recall the number of dots per card in the correct order. When adults and children were equated on counting speed by asking adults to count in an unfamiliar language, span performance difference between age groups vanished. This pleads in favor of an explanation of WM growth by processing speed increases with age. More precisely, Case and colleagues, defend the idea that the global capacity of WM resources would remain similar with age, but the processing

efficiency would increase with age. In their study, there is however one confound that has to be considered. Indeed, if older participants are faster at processing stimuli, it means that they take less time to process information than younger children. If we postulate like the TBRS does (Barrouillet & Camos, 2015), that processing and maintenance of information are two successive tasks in time that share a common resource, and that while doing processing of a concurrent task, maintenance of items cannot occur, then shorter time processing would leave more time for maintenance. Towse and Hitch (1995), reinterpreted Case et al.'s results and put the emphasis on this last point. To them, it was not so much processing efficiency increases with age that was important to explain WM's performance increase with age, but rather the duration of the processing and maintenance. Indeed, all the time not spent on memory traces maintenance, leads to memory decay. Hence, spending less time on processing and more time on maintenance, would explain the age-related span increase. To show this, they asked 6- to 11-year-olds to do a counting span task, in which processing difficulty varied (i.e., either the items to count were mixed with some distractors or not) and the number of items to count varied which ultimately varied the time of counting. Results show that children's span was higher when the task took less time and was less difficult, at all ages. However, for the same duration, the presence or not of distractors did not vary performance at the task, suggesting that it is the time spent on the concurrent task rather than its difficulty that is responsible for span performance increase with age (see also Towse, et al., 1998, for similar results with reading or operation span). One criticism made by Barrouillet and Camos (2001), to Towse and Hitch's protocol (1995) is that they did not consider that increasing the number of items to be counted, might not only impact the duration of the concurrent task but might also increase the processing difficulty of the task (see also Camos, et al., 2001). Hence, there might be a confound in their study. To solve this issue, Barrouillet and Camos (2001), equalized the duration of processing in two complex span tasks, where 9 and 11-year-olds were shown letters to memorize, followed by the introduction of a concurrent task in between each of them. The difficulty of the concurrent task was varied so that processing was more or less demanding in terms of attentional resources. Results indicate that when the concurrent task asked for more attentional resources, the span performance decreases for the same task duration. Hence, based on Case (1985) and Towse and Hitch's findings (1995), the TBRS model (Barrouillet & Camos, 2015), proposed that both task difficulty and duration of the concurrent task affected WM's capacity. There was evidence for this in several studies which showed that, when both duration and difficulty of the concurrent task were equalized among age groups, the span difference between age groups became smaller (Barrouillet et al., 2009; Gaillard et al., 2011; Gavens & Barrouillet,

2004). The TBRS explains therefore the increase in span capacity with age, by an increase in processing efficiency. Processing would recruit attentional resources for shorter time with age as it becomes more efficient. In consequence, attentional resources, which are shared between maintenance and processing in time, can be recruited for maintenance of memory traces for longer periods of time. More precisely, processing becomes more efficient with age because of the development of cognitive resources (Halford, 1993; Pascual-Leone, 1970), reflecting the increase of speed processing with age. In particular, control processes become more efficient with age (Bardikoff & Sabbagh, 2017; Blaye, 2022; Blaye & Chevalier, 2011; Chevalier & Blaye, 2008, 2009; Miyake et al., 2000), as shown by a maturation of the frontal cortex (Huttenlocher, 1979; Yakolev & Lecours, 1967), by an increase in inhibition capacities (Bjorklund & Harnishfeger, 1990, Tipper et al., 1989), and most importantly here, the development of attentional resources (Cowan, 2001; Halford et al., 2014). Adding to this, maintenance mechanisms also become more efficient with age, which allows individuals to maintain more efficiently the items in memory.

3.2.1 The development of attentional resources.

Cowan (2001, Chapter 5) explains the increase of WM capacity by the development of the focus of attention with age. According to him, an increase of the size of the focus of attention (i.e., the number of elements on which we can put our attention), would allow us to maintain more items in memory, which would in turn lead to the increasing span performance with age observed in various studies. To measure the number of items that could be maintained in the focus of attention, Cowan and colleagues used tasks that neutralized the possibility of articulatory rehearsal or other strategies like chunking, leading participants to rely on the retrieval of items present in the focus of attention. They showed that the focus of attention could maintain 2 items at 6-7 years and about 4 items in adults, confirming an increase of the focus of attention with age (see Cowan, 2001, for a review; see also Halford et. al., 1998, for similar findings).

Besides, in addition to the increase of the size of the focus of attention, there is evidence that the central executive, that controls the focus of attention, develops and matures with age. For instance, Maccoby and Hagen (1965), showed that participants' span increased between 3-4-year-olds and 9-10-year-olds on the characteristic they were asked to memorize (i.e. the color of items). Oppositely, their performance on characteristics they were not asked to memorize were lower in the oldest age group compared to the youngest. This suggests that the control of

the focus of attention via the central executive, improves with age. These results echo the findings on the maturation of the frontal cortex and inhibition capacities, presented in the previous section.

3.2.2 The development of maintenance mechanisms.

As described in section 3.1.2, the TBRS identified two processes serving the function of maintenance of information in memory: attentional refreshing and articulatory rehearsal. Several studies have shown that attentional refreshing develops with age, leading to better maintenance in WM with age. To show this, Barrouillet et al. (2009, Exp 3.) introduced to 5- and 7-year-olds, a complex span task in which children had to memorize animal names. In between the presentation of each name, a concurrent task was introduced in which the difficulty of the task, and more precisely the cognitive load induced varied. In a high load condition, participants had to identify the color of 4 smileys. In a medium load condition, only 2 smileys were presented for the same time interval (8 seconds). In a no-load condition, there was no concurrent task. Results show that 7-year-olds' recall performance decreased in the 4-smiley condition compared to the 2-smiley condition. On the contrary, even though the introduction of a concurrent task lowered the performance of 5-year-olds (no-smiley condition versus the two others), the cognitive load of the concurrent task did not vary their performance. It indicates that they might not benefit from situations in which they have more time to refresh items in memory at this age, like 7-year-olds would. Hence, if younger children do not use refreshing, longer delays for the concurrent task should decrease their performance. Camos and Barrouillet (2011) tested this hypothesis by varying orthogonally the cognitive load and the duration of the concurrent task in a similar design than that of Barrouillet et al. (2009). The concurrent task was therefore composed of 1 or 2 smileys for a duration of 2 seconds or of 2 smileys for a duration of 4 seconds. Results indicate that performance of 7-year-olds depended on the cognitive load of the concurrent task (1 versus 2 smileys) and not on the duration of the task (2 versus 4 seconds), while it was the opposite for 5-year-olds. This brings further evidence that 5-year-olds do not use refreshing to prevent the decline of memory traces with time. Hence, it seems that refreshing emerges around 7-years of age (but see Bertrand & Camos, 2015, for earlier manifestation of refreshing use under playful conditions). Furthermore, refreshing keeps on developing with age and its efficiency increases between 7 and 14-year-olds (Barrouillet et al., 2009, Exp.2; Gaillard et al., 2011). Indeed, attentional refreshing gets faster and more efficient, therefore allowing to refresh more information. Gaillard et al., (2011, Exp.3) showed this by adapting the concurrent task difficulty to each tested age group (8 and 11-year-olds), in

a complex span paradigm. The younger age group had to process fewer items during the concurrent task than the older children and were also given more time to do so. Results indicate that under these conditions, span performance of both age groups was equalized. The performance of the youngest age-group increased indeed to the level of older children's, leading to the interpretation that under natural conditions, refreshing is slower and less efficient in younger children. These results were also replicated by a recent study (i.e., Barrouillet et al., in prep.), comparing wider age groups: 9 and 15-year-olds.

Articulatory rehearsal was also shown to contribute to quantitative changes in WM's performance with age. The main constraint of this mechanism is that it relies on verbal material. Therefore, its use depends on the ability to recode the studied material with verbal labels when it is not presented under verbal form. Verbal recoding capacity was shown to be used from 5 years and increases with age (Hitch, et al., 1989), even though children of less than 7-years showed a preference for visual than verbal codes (e.g., Hitch et al., 1988; Frick, 1988a, 1988b; Logie et al., 2000; Longoni & Scalisi, 1994; Palmer, 2000a). Congruently with this switch between 5 and 7-years in coding preferences, various studies indicated indirectly, that rehearsal starts to be used between those ages and that the frequency at which it is used develops with age. These studies are contrasted about the exact emerging age of rehearsal. Flavell, Beach and Chinsky (1966) showed for instance that older children had more lip movements than younger children. Lip movement or overt speech could be a relevant indicator of the presence of articulatory rehearsal, in that it uses the same phonological articulatory system. Moreover, Flavell et al.'s (1966) study was recently replicated over 17 laboratories across the world and showed that overt speech appeared from 5-year-olds and increased between 5 and 10-years of age (Elliott et al., 2021). They also showed that at 5-6-years-old, children who employed overt speech, had higher memory span performance compared to children that did not. Moreover, because articulatory rehearsal relies on verbal maintenance, speech rate could also indirectly give indications about the use and efficiency of articulatory rehearsal. Some studies showed no correlation between speech rate and span performance in children before age 7 (e.g., Ferguson, et al., 2002; Gathercole & Adams, 1993), while the correlation occurred in older children and adults (Baddeley, et al., 1975; Jarrold, et al., 2004). Moreover, the typical word length effect (i.e., better recall of lists of smaller words) and phonological similarity effect (i.e., better recall of phonologically dissimilar than similar words or labels) observed under articulatory rehearsal do not appear in young children (4-5-years) unlike in older children (from 6-years) and adults (e.g., Allik & Siegel, 1976; Henry et al., 2012, but see Conrad, 1971). In other studies, a more

direct measure has shown that when articulatory rehearsal was impaired by the implementation of a concurrent articulation during retention intervals of a WM task, children's performance decreased, indicating that they used this strategy to maintain information in WM. This effect was present in 6-year-olds, which was the youngest tested age group in these studies, and in older children (Oftringer & Camos, 2016, 2018). Besides, when younger children were instructed to use rehearsal, their performance did not increase, which leads the belief that it is difficult for them to implement this strategy or at least, benefit from it (Ornstein et al., 1975; Naus et al., 1977). However, more recent work showed that children as young as 5 years of age could benefit from training in articulatory rehearsal (Miller et al., 2015). Altogether, these elements imply that the use of articulatory rehearsal for verbal information maintenance emerges between the ages of 5 and 7. Qualitative changes also operate during rehearsal's acquisition. At first, children can use simple rehearsal, where they repeat one word, which is usually the last one they heard. Later on, they start to use cumulative rehearsal, which consists of repeating all the studied words from the start. This development underlies span performance improvement (Naus et al., 1977).

Despite refreshing and rehearsal, other mechanisms have been identified to allow maintenance of information in WM. Belletier et al. (2021) have regrouped the strategies that participants reported using during maintenance. Aside from articulatory rehearsal, participants reported using strategies that intended to regroup information such as temporal clustering (i.e. regrouping items based on their successive presentation in time), or memory reduction strategies (i.e. dropping parts of the items to learn, such as some letters of a word to be able to remember the most items possible). Another reported strategy consisted of trying to memorize the surface form of the stimuli in the sensorial trace they were presented (e.g. 'visually memorize the letters to learn', 'memorize the sound of the letter by focusing on the sound itself'). Finally, a last type of reported strategy consists of trying to give meaning to the studied information, by using LTM, and to make categories or establish meaningful relationships between items (e.g., associating letters with words, make a sentence with words, etc.). This strategy is called: elaboration and chaining. There was evidence in the literature that elaboration was not used before 8-years-old (Cowan et al., 2005; Towse, et al., 1999). Most of those strategies are measured through questionnaires, in which participants are asked to describe the strategies they used to maintain information or to reply by yes or no if they used the strategies described by the experimenter. Some of these strategies are indicators of

qualitative rather than quantitative changes in the child's WM development. The emergence of a new strategy could be considered as a progression in the way children apprehend WM tasks.

3.2.3 Coordination of maintenance mechanisms

Articulatory rehearsal was shown to be the default maintenance mechanism used by children from about 6-years-old (e.g. Oftinger & Camos, 2016). Indeed, this is a mechanism that's easy to use because it does not rely on attentional resources, being therefore independent of the cognitive load of the concurrent task. A few studies have tested whether both rehearsal and refreshing could be used by children to maintain information in WM. Oftinger and Camos (2016) varied orthogonally the possibility to use rehearsal and refreshing during the maintenance interval of a complex span task, by implementing a higher (color discrimination) versus lower (categorization task) cognitive load concurrent task (i.e. refreshing manipulation), with and without aloud articulation (i.e. rehearsal manipulation). Results reveal that a concurrent articulation decreased word or letter recall performance of children of all age groups (6 to 9-years). Independently, a higher cognitive load was synonymous of reduced performance of children from all age groups (see Oftinger & Camos, 2017, 2018, for similar results). It suggests for one, that children are able to use both rehearsal and refreshing in a task. Secondly it reinforces the idea mentioned in section 3.1.2 that both mechanisms are independent of each other, as they did not show any interaction effect on recall performance (see also, Magimairaj, & Montgomery, 2012; Mora & Camos, 2015). They did however have additive effects on recall. Indeed, it was found that preventing both rehearsal and refreshing simultaneously, decreased recall performance more than when only one mechanism was prevented (Tam et al., 2010). In Oftinger and Camos's study (2018), rehearsal and refreshing was varied orthogonally like in their 2016 study, but this time in a Brown-Peterson task. A correlation between the measure of the availability of attention for refreshing and recall performance was found in 7- and 8-year-olds but not in 6-year-olds. This correlation was found in 7-year-olds only when rehearsal was prevented. These findings confirm that rehearsal seems to be the preferred strategy and that when it can be used, children do so. When rehearsal cannot be used, from 7-years of age, they can adaptively switch toward the use of refreshing, a capacity also shown in adults (Camos, 2015, 2017, for reviews). These findings are also backed up by Barrouillet et al.'s (2009) and Camos and Barrouillet's (2011), which indicated an emergence of attentional refreshing around age 7.

3.2.4 The role of long-term memory in the development of working memory.

As described in section 3.1 of this Chapter, WM interacts with and relies on LTM, which in some cases can lead to more efficient performance in WM tasks, compared to when it does not rely on it. For instance, as noted by Cowan (2001), more elements could be maintained in the focus of attention when they can be regrouped under meaningful chunks (see also Miller, 1956). Other effects observed in WM tasks such as the semantic similarity effect, the imageability effect, the frequency effect or even the better recall of words than non-words, are caused by reliance on LTM. As explained in Chapter 2, because knowledge in LTM develops during childhood, it could partly account for the performance increase in WM tasks with age. An example of this is Roodenrys, et al.'s study (1993), who showed that older children (11-12-years) benefitted from LTM's knowledge and recalled more words than non-words in WM tasks, while younger children (5-6-years) did not. A study from Cowan et al. (2015) supports this idea but shows that it is however not the only determinant of WM development with age, as age increase in span was not abolished when unfamiliar characters had to be recalled. Ironically, if LTM facilitates the occurrence of correct recall with age in WM tasks, it could also facilitate the occurrence of semantic false memories.

3.3 Short- and long-term false memories in working memory tasks.

3.3.1 False memories at short-term.

One legitimate question to ask regarding false memories is whether they occur at short-term like they would in LTM tasks. Indeed, given that memory traces decline with time, especially verbatim memory, one could think that false memories do not occur at short term because verbatim memory traces are preserved at that time. This question started to be investigated recently.

First, two studies conducted by Sugrue and colleagues (2006, 2009), compared the rate of false recall of the critical item, occurring immediately after word-list presentation in classical 14-word DRM lists and shorter 7-word DRM lists. They showed that 10 year-olds like adults, produced on average 20-25% of false recall of the critical item for short-lists against the classical 45% on longer lists. These results are consistent with Coane et al.'s study (2007) who presented 5 to 7 words per DRM list to young adults and found around 20% of false recognition, of the critical item, in a test occurring 1 second after stimuli presentation. Besides, they found

that correct rejection of a critical lure took longer than of unrelated distractors (around 100 ms more). This suggests that false memories can occur in short-term memory tasks but at lower rates than in LTM tasks. One potential explanation to this, which was the one given by Sugrue and colleagues, would be that when presenting more words related semantically, gist processing increases and therefore leads to an increased rate of false memories. Complementarily to this interpretation, we could think that, when all the word-lists can be maintained by WM maintenance mechanisms, detailed traces in memory are preserved, leading in consequence to fewer false memories.

Atkins and Reuter-Lorenz (2008) were the first to show that false memories could occur in WM tasks. They conducted a study in which four semantically related words were presented, followed by a 3-4 second retention interval, during which they had to do a concurrent task. Following this, was presented an immediate single probe recognition test or a recall test. Results revealed 31% of false recognition on the critical item against 16% for unrelated distractors, and 11% of semantic errors in recall, which was more than half of the total error rate. Semantic errors consisted mainly of the recall of the critical item. Like in Coane et al. (2007), correct rejection of a critical distractor took longer than of unrelated distractors. Therefore it seems that false memories can occur in WM tasks at fairly high rates. One particularity of WM tasks, as mentioned in section 3.1.2, is that maintenance mechanisms can be used to hold information temporarily. Hence, studies that tested false memories in WM tasks, have measured the effect of those mechanisms on such memory illusion.

3.3.2 The role of maintenance mechanisms.

First, some studies conducted in young adults assessed the effect of articulatory rehearsal on short-term false memories. Atkins et al. (2011, Exp 1.) presented a task similar to that of Atkins and Reuter-Lorenz (2008). During the retention interval, participants were asked in half of the trials to repeat the numbers ‘one, two, three’ continuously to prevent the use of the articulatory rehearsal of the to-be-learned words. In the other half of the trials, participants were not instructed to do any task during the retention interval, allowing the use of articulatory rehearsal. Results indicate that the rate of false alarms on critical items, increased when participants had to do a concurrent articulation during the retention interval (i.e., 10% compared to 2%). This suggests that using articulatory rehearsal prevents the occurrence of immediate false memories by reducing their rate to quasi floor levels. In another study, Macé and Caza (2011, Exp. 2) presented lists of 6 words containing 4 DRM associates and 2 unrelated words. Participants

were presented two successive lists and had to judge whether they were similar or not. In this design, the two successive lists could be either identical, or changed by one word. In this second case, either a related word could be replaced by the critical item or an unrelated word could be replaced by another unrelated word. The possibility to rehearse the studied words of the list was also manipulated. In half of the trials, participants had to do a concurrent articulation during the presentation of the two successive lists, by counting aloud from one to ten until they gave their response about the match between the two lists. Results show that more often participants identified two lists as matching when the critical item was introduced, compared to when an unrelated lure was introduced. This was especially true in the condition where they had to do a concurrent articulation. In fact, there was 36% of false identification of a list composed with the critical item under articulatory suppression against 8 % when rehearsal could be used. Therefore, it seems that articulatory rehearsal has a role in short-term false memory prevention.

In 2019, Abadie and Camos tested the conjoint and independent effects of articulatory rehearsal and attentional refreshing on short-term false memories. They used the same paradigm than the one used by Atkins and Reuter-Lorenz (2008) and manipulated the availability of rehearsal and refreshing orthogonally through the 4 seconds retention interval implemented between the 4 word-related-lists⁴ presentation and the recognition probe. During this interval, in one experiment, participants had to do a low attentionally silent detection task, which allowed to use both mechanisms for word maintenance. In a second experiment, participants had to do an aloud operation verification task (e.g., ‘ $2+3 = 4$?’), to block both refreshing and rehearsal for word maintenance. In a third experiment, participants only had to read aloud but not verify those operations, which impaired only articulatory rehearsal for maintenance. Finally in a fourth experiment, an attentionally demanding spatial fit task (decide whether a horizontal bar can fit in the gap between two dots) was introduced during the retention interval to block refreshing. Compared to when both maintenance mechanisms could be used (~17%), false alarm rates of critical items were higher under the aloud operation verification task (~37%) and the aloud reading task (~30%), meaning that when neither maintenance mechanisms or only refreshing could be used, false memory increased. It however did not increase when only attentional refreshing was impaired by the spatial fit task (~11%). Altogether, it suggests that when

⁴ Note that in this experiment, unrelated lists were also presented to measure the effect of gist activation on false recognition, when compared to related lists.

articulatory rehearsal is preserved, false memory is partly prevented, which is congruent with other previous findings (i.e., Atkins et al., 2011; Macé & Caza, 2011). In this study, half of the lists were tested with a delayed recognition test occurring 2 minutes after the end of the list presentation. During those two minutes, participants were asked to count backward to eliminate the last items in WM. Results indicate quite different patterns on that test. Indeed, rates of false alarms on critical items were very low when attentional refreshing was blocked by a concurrent task (~1% in both operation verification and spatial fit tasks) and were higher when it could be used for word maintenance (~25%). Hence, it seems that the use of refreshing does not impact immediate false memories, but promotes them at longer delays in WM tasks. In contrast, articulatory rehearsal prevents false memories in immediate and not delayed tests.

3.3.3 Dissociation between false memories in immediate and delayed tests.

The findings from Abadie and Camos (2019), in which rehearsal impacts false recognition in an immediate test, whereas refreshing impacts false recognition in a delayed test, suggests a dissociation between false memories at short and long-term. In other words, false memories at short-term might be ruled by different factors than false memories at long-term. This was investigated in two studies, which established the role of two different LTM moderators on false memory in immediate and delayed tests.

First, in a study from Flegal et al. (2014), the level of stimuli processing during encoding, a variable known to affect false memories in delayed tests in classical LTM tasks (see Chapter 2) was manipulated in a Brown-Peterson task. To manipulate this variable, participants in one condition had to do a shallow processing of the studied words by counting the numbers of ascender (e.g., t, l) and descender (e.g., p, q) letters in the four presented word associates. In another condition they had to do a deeper processing by judging whether they liked the word they studied or not. This second condition is considered as triggering deeper processing because it necessitates the understanding of the word's meaning to answer the asked question, while the first condition asks for local letter processing, tapping surface morphological processing. This manipulation did not impact the production of either correct recognition of target probes or false recognition of related distractors in the immediate recognition test, occurring 3-4 seconds after item study. However, in the 20 minutes delayed test, both correct recall of targets and false recognition of related probes increased in the deep processing condition compared to the shallow condition, whereas there was no effect of the LOP on false recognition of unrelated

probes. This is what is also typically found in classical DRM tasks with longer lists of words. This contrasted pattern between immediate and delayed tests is congruent with the results of a previous study from Rose et al. (2010), which also showed no effect of LOP on immediate correct recognition, whereas deeper LOP increased delayed correct recognition. One simple explanation offered by the authors, to the absence of LOP effect in immediate recognition, could be that verbatim being preserved enough at that time, a deeper level of meaning processing (i.e., gist) does not affect recognition in such a short-period.

In another study, Olszewska et al. (2015) tested the effect of stimuli modality on immediate and delayed recognition. Lists of 4 associates were presented either as visual words, like in the previous studies using a similar paradigm, or auditorily. The retention interval was filled by a concurrent task targeting the phonological loop. Participants were asked to do a concurrent articulation during 3 seconds before immediate or delayed recognition. Results indicate that auditory presentation leads to higher correct recognitions and lower false alarm rates on related distractors, compared to the visual presentation, in the immediate test. Conversely, correct recognition was higher and false memory lower in the visual modality in the delayed recognition test. These results on the delayed test are congruent with those of studies using LTM tasks (see Gallo, 2006 for a review), and likely translate the increased distinctive perceptual features in visual stimuli, allowing for more correct rejection of unstudied words (see Hunt, 2013; Hunt et al., 2011). It seems however to work differently in the immediate test. Olszewska et al., suggest that the memory traces available in the immediate test (i.e. gist and verbatim) are not similar in quantity to those in the delayed test in the two different tested modalities. Indeed, auditory stimuli would create more immediate verbatim traces (see Nairne, 2002; Neath & Nairne, 1995; Penney, 1989), but they would fade away quickly in time (see, Bigelow & Poremba, 2014). Hence, the advantage of auditory modality in short delays because of a greater availability of verbatim memory, would fade away and even reverse with time, letting the visual modality take the advantage with time.

The dissociation between false memories in immediate and delayed tests, highlighted by a difference of effects of maintenance mechanisms, level of processing and of stimuli modality on each type of test, questions whether similar processes underly false memories at short and longer delays.

3.3.4 Underlying processes and phenomenological experience.

To answer this question, Abadie and Camos, investigated how the FTT would account for their findings obtained from the manipulation of rehearsal and refreshing. They used the simplified conjoint recognition model described in Chapter 2 (Stahl & Klauer, 2008), to assess on which type of memory traces, the responses to the recognition tests were based (i.e. gist memory, verbatim memory or guessing processes). They showed first that false memories were based on gist memory both in the immediate and the delayed test. More specifically, Abadie and Camos compared the gist retrieval rate in short-DRM lists and in lists of 4 unrelated words. They showed that related word lists led to higher rates of gist memory on related distractors than unrelated lists. This increased rate underlied an increased rate of false memories in related lists. Secondly, they showed that when articulatory rehearsal could be used to maintain the word-lists, verbatim memory was increased, which allowed intact detailed memories to be preserved, and caused in turn a false memory decrease in immediate recognition. Oppositely, the use of refreshing fostered gist memory traces, and caused the increase of false memories in a delayed test. It seems therefore that short and long-term false memories in WM tasks are underlied by a similar process: gist memory. The dissociation between short and long-term false memories observed in some studies is most likely due to a modification of the balance between gist and verbatim memory with time. Indeed, verbatim memory decreasing faster in time, it does not allow to counteract the effect of gist memory as much in the long than in the short term. Moreover, it seems then, that maintenance mechanisms have a moderator role in false memory occurrences in WM tasks, in that each of them allow a specific type of memory trace to be preserved.

Complementary to the objective measures collected by Abadie and Camos, a study from Flegal et al. (2010) measured participant's subjective descriptions regarding the quality of their memories. Participants were presented with a task similar to that of Atkins and Reuter-Lorenz (2008) with lists of 4 associates followed by a 3-4 seconds operation verification task in all trials. Half of the lists were paired with an immediate recognition probe test while the other half of lists were tested 20 minutes later with a surprise recognition test. During recognition, participants were asked to estimate their confidence in their response on a scale from 1 (very low) to 4 (very high), in the first experiment. In the second experiment, they had to associate their response to either a familiarity judgment (i.e., know), which corresponded to when they recognized the word without specific detail of its study, to a 'remember' judgment, which corresponded to when they recollected something distinctive about studying the word, or to a

guessing judgment when their response had been a guess. Results indicated that participants in Experiment 1, had lower confidence rates for false alarm on the critical item than on correct identification of target probes, but also that this confidence rate was similar across tests in time. It suggests that the illusory experience of false memory occurs from the first few seconds after encoding and remains stable in time. Moreover, in the second experiment, correct recognition of target probes were more often associated with remembering judgements than false alarms. Nonetheless, false alarms on critical items were more associated with remembering judgments compared to false alarms on unrelated probes. This means that like in a classical LTM DRM task, false recognition of semantically related distractors are accompanied more often by the retrieval of a detailed memory, compared to other error types. Besides, the proportion of false alarms of critical items related to a remembering judgment was stable across test delays, which is supplementary evidence that immediate and delayed false memories are based on similar memory traces. Regarding 'know' judgments, results indicate that they were associated with similar rates to target and related probes and were also stable over time. According to the authors of the study, it suggests that gist memory exerts equivalent influence in immediate and delayed tests, when context information is inaccessible. The results from Flegal et al. (2010) were replicated in a following study (i.e., Flegal et al. 2014). Moreover, false recognition of related probes, were descriptively associated as much to 'remember', 'know' and 'guessing' judgements. This last point shows that false recognition does not always trigger the same subjective experience. Some are associated to more vivid detailed experiences than others. These findings are in line with Brainerd et al. (2001) who described different false memory types, i.e. those based on phantom recollection and those based on fuzzier gist traces (see Chapter 2) where phantom recollection could be assimilated to remember judgements and fuzzier gist based false memories to 'know' judgments. Hence, the stability of subjective experience over time associated to false memories questions the dissociation found between immediate and delayed tests. On one side, there is evidence that false memories are dissociated at short and long-delays in WM tasks, when some moderators are introduced (i.e., LOP, WM maintenance mechanisms, stimuli modality). On the other side, studies that assessed the underlying processes and subjective experience, indicate that false memories in immediate and delayed tests are underlied by similar processes. There is currently too little work on the topic to fully understand the comparison between false memories in WM tasks at short and long-term.

In this chapter, I introduced the differences between long-term and working memory, and presented the sources of development of WM during childhood. I introduced fairly recent findings showing that false memories occurred in WM tasks, and reviewed the few existing studies on the matter. The next chapter will aim at introducing the experimental part of this thesis.

Summary

In this Chapter, were presented the distinctions between three main memory systems: sensory, WM and LTM. Sensory memory captures sensorial stimuli and then transfers to WM, where only a few pieces of information can be successively maintained and processed for the purpose of a task. I presented two models of WM, namely the Time-Based-Resource-Sharing model and the Embedded-Process model. Attentional resources are at the heart of both models as they are essential to maintenance and processing. Moreover, according to the TBRS, two mechanisms allow the maintenance of information in WM: attentional refreshing and articulatory rehearsal. After information is held in WM for a few seconds, it then ultimately transfers to LTM and consolidates in time. LTM can contribute to WM activities, as highlighted by the presence of specific effects in WM performance, such as the semantic relatedness effect. Besides, WM develops during childhood, as shown by the increase in span performance with age. Efficiency in processing speed increases with age, which leaves more time for maintenance of information via attentional refreshing and articulatory rehearsal. The emergence of those mechanisms starts between 5 and 7-years of age and keep on increasing in efficiency until adulthood. Cowan also describes an increase of the focus of attention with age, allowing the maintenance of more information. The study of false memories in WM tasks is a very recent research topic. The few existing experiments were conducted in adults and have shown that this illusion could occur only 3-4 seconds after the study of 4 stimuli. This type of false memory was shown to be based on gist traces as in classical LTM task. Besides, false memories in immediate and delayed tests were shown to share similar phenomenological experiences. Some are based on remembering phenomenology, while others rely on familiarity. There was also evidence that WM maintenance mechanisms could modulate the occurrence of false memories. Rehearsal prevents the occurrence of false memories in immediate tests by preserving verbatim traces, conversely, there was evidence that refreshing promoted the maintenance of gist memory, increasing therefore delayed false memories (Abadie & Camos, 2019). Moreover, some studies using WM tasks showed some dissociations of effects between immediate and delayed tests (Flegal et al. 2014; Olszewska et al., 2015). It is very likely that this difference could be explained by a difference in verbatim and gist availability at different times of tests. In longer delays, some effects could appear (e.g., LOP, switch of modality of stimuli that leads to higher false memories) because verbatim memory is not high enough to counter gist activation effect on false memories.

Résumé

Dans ce chapitre, je fais la distinction entre trois systèmes de mémoire : la mémoire sensorielle, la mémoire de travail et la mémoire à long-terme. La mémoire sensorielle capture les stimuli et les transfère en mémoire de travail, où seulement quelques informations peuvent être successivement maintenues et traitées dans le but de réaliser une tâche. J'ai présenté deux modèles de mémoire de travail, à savoir : le modèle du partage temporel des ressources et le modèle des processus emboîtés. Les ressources attentionnelles sont au cœur de ces deux modèles et sont essentielles pour le maintien et le traitement de l'information. De plus, d'après le modèle du partage temporel des ressources, deux mécanismes permettent le maintien de l'information en mémoire de travail : le rafraîchissement attentionnel et la répétition articulatoire. Après que l'information ait été maintenue durant quelques secondes, elle est transférée en mémoire à long-terme et se consolide dans le temps. La mémoire à long-terme contribue aux activités de la mémoire de travail, ce qui a été mis en évidence par la présence d'effets spécifiques sur les performances en tâche de mémoire de travail, tel que l'effet de similarité sémantique. Par ailleurs, la mémoire de travail se développe durant l'enfance, comme l'indique l'augmentation des performances d'empan avec l'âge. La vitesse de traitement augmente avec l'âge, ce qui laisse plus de temps pour le maintien de l'information par le rafraîchissement attentionnel et la répétition articulatoire. L'émergence de ces mécanismes de maintien commence entre 5 et 7 ans et continue d'augmenter en efficacité jusqu'à l'âge adulte. Cowan décrit également une augmentation de la taille du focus attentionnel avec l'âge, permettant un maintien de plus d'informations. L'étude des faux souvenirs en tâche de mémoire de travail est un sujet de recherche récent. Le peu d'études existantes a été réalisé auprès d'adultes et a montré que cette illusion pouvait survenir 3-4 secondes après l'étude de 4 stimuli. Ce type de faux souvenirs se basait sur des traces gist, comme dans les tâches de mémoire à long-terme. Par ailleurs, l'expérience phénoménologique d'un faux souvenir est similaire à court et long-terme. Certains faux souvenirs sont associés à de la recollection alors que d'autres sont associés à de la familiarité. Par ailleurs, il a été montré que les mécanismes de maintien en mémoire de travail pouvaient moduler le taux de faux souvenirs. La répétition prévient les faux souvenirs dans les tests immédiats en préservant les traces verbatim. A l'inverse, le rafraîchissement promeut le maintien des traces gist en mémoire, ce qui augmente les faux souvenirs dans les tests différés (Abadie & Camos, 2019). Par ailleurs, certaines études utilisant

des tâches de mémoire de travail, ont montré une dissociation entre des effets en tests immédiats et différés (Flegal et al. 2014; Olszewska et al., 2015). Il est probable que cette différence s'explique par une différence de disponibilité des traces verbatim et gist en fonction du moment du test. Certains effets pourraient apparaître lors de délais plus longs (e.g., effet de profondeur de traitement, effet du changement de modalité des stimuli, qui conduit à plus de faux souvenirs) car le verbatim n'est pas suffisant en mémoire pour contrer l'effet de l'activation des traces gist sur les faux souvenirs.

Chapter 4. Overview of experimental chapters

The theoretical part of this thesis summarized how false memories have been the object of many studies in the past. From these studies, we learned about the different factors impacting this illusion. The literature has differentiated ‘suggested’ from ‘self-generated’ false memories, showing the involvement of specific and common factors. Various studies have also investigated how false memories evolved with age during childhood, showing that false memories tend to increase with age. According to the FTT, the development of gist memory would be responsible for such increase; these findings have been demonstrated using mainly LTM tasks. In these tasks, recognition or recall tests occur minutes after stimuli were studied, and the number of stimuli to study usually is above 8 items. However, a few more recent studies have shown that false memories occurred in mere seconds, after the study of only four semantically related words (e.g., Atkins et Reuter-Lorenz, 2008). This finding opens a wide range of questions. Indeed, one could wonder whether this type of short-term false memory is the same as the one observed in LTM tasks. Do these memories rely on similar memory traces (i.e., gist memory) ? Are they impacted by the same factors as in LTM tasks ? How does this type of false memory evolve with age ?

It would take an extensive amount of research to answer all these questions. Indeed, the literature on short-term false memories is nascent when compared with the literature on long-term false memories. Nonetheless, the few existing studies have shown the involvement of mechanisms specific to WM in short-term false memories. Indeed, the mechanisms that allow us to maintain information for a few seconds were shown to modulate the occurrence of false memories in adults in a WM task. In particular, articulatory rehearsal decreased false memories in immediate tests (Abadie & Camos, 2019; Atkins et al., 2011; Macé & Caza, 2011). In contrast, attentional refreshing was shown to favor false memories in delayed tests (Abadie & Camos, 2019). Moreover, it was also found that these maintenance mechanisms impacted memory traces. In particular, rehearsal preserved verbatim memory, whereas refreshing preserved gist memory. These findings raise several lines of questioning, including the existence of a potential dissociation between false memories in immediate and delayed tests in WM tasks; the impact of WM maintenance mechanisms on memory traces, and the involvement of LTM in false memories in WM tasks. For this last question, Abadie & Camos’ study had in fact shown that presenting semantically related lists of 4 words increased false memories, in

comparison with the presentation of 4 unrelated word-lists. These findings suggested that gist memory would be responsible for false memories in WM tasks as they seem to be in LTM tasks.

No studies have yet investigated the developmental trend of false memories in WM tasks. We do not know whether we would find in these tasks the increase of false memories with age, usually observed in classical LTM tasks. Additionally, both gist memory and WM maintenance mechanisms develop with age. Therefore, comparing the occurrence of false memories in different age groups could contribute to our understanding of how these mechanisms are involved in false memories at short and longer delays. It could be expected that the more WM maintenance mechanisms are developed, the greater their impact on false memories. Age could also be a quasi-manipulation of gist memory activation, in that older children would be expected to have more gist memory than younger children. Thus, comparing different age groups would indirectly assess the effect of gist memory on false memories. This comparison would be combined with complementary measures of gist memory.

Therefore, the goal of the following series of experiments was to adopt a developmental perspective to find out more about the role of both LTM traces (i.e., gist and verbatim) and WM maintenance mechanisms in false memories in WM tasks. We conducted three studies in which we compared different age groups, ranging from young children to young adults, to measure the development of false memories in WM tasks with age (see Table 1 for an overview). We measured the effect of WM maintenance mechanisms on short and long-term false memories in different age groups by manipulating the demands of the concurrent task introduced during retention intervals of the WM task. We also measured the influence of LTM on false memories in recognition and recall, by manipulating the relatedness of words within lists (i.e., related versus unrelated words). We also assessed the processes underlying false memories (i.e., gist and verbatim), by using the simplified conjoint recognition model of the FTT. Finally, we measured the phenomenological experience associated with false memories.

Even though this work is mainly fundamental, and contributes to our general knowledge on how memory works, it could also influence two main societal fields. The first one is the field of justice with the topic of eye-witness testimonies. There have already been some guidelines regarding how to question a witness in order to avoid the production of false memories, which were established based on the findings of previous studies (e.g., Bruck & Ceci, 1995; Otgaar et al., 2017; Reyna et al., 2002). The present work could inform on practical behavior that one might have in the first few seconds after witnessing an event, to limit the production of false

memories. The second field is education. Indeed, understanding whether false memories are likely to occur in WM tasks, which represent nearly all classroom tasks (e.g., taking notes, reading, doing maths, etc.), and at which ages, could provide educational guidelines in teaching methods.

4.1 Study 1

Our first study (i.e., Chapter 4) aimed at assessing whether false memories would occur at short- and long-term in a WM task in young and older children like they were shown to occur in young adults (Abadie & Camos, 2019; Atkins & Reuter-Lorenz, 2008; Atkins et al., 2011; Flegal et al. 2010, 2014). Four- and eight-year-olds were therefore presented a Brown-Peterson task, in which 3 to 5 associates were to be memorized, followed by a 6-8 second retention interval and a single probe immediate recognition test that could be either a target, a semantically related or unrelated probe. A delayed recognition test occurred at the end of the task. During the retention interval, rehearsal was prevented in one condition via the implementation of a concurrent articulation. In another condition, it was not prevented because no concurrent articulation was introduced. Refreshing was also prevented or not via the variation of the cognitive load in the concurrent task (i.e., high versus low). This manipulation aimed at testing the effect of WM maintenance mechanisms on false memories in immediate and delayed tests. We compared 4-year-olds with 8-year-olds because a developmental shift in WM maintenance mechanisms occur precisely between those ages. Hence, the manipulation of WM maintenance mechanisms was expected to impact 8-year-olds, who were shown to use both rehearsal and refreshing (e.g., Oftinger & Camos, 2016, 2018), but not 4-year-olds, for which there is evidence against the use of WM maintenance mechanisms (e.g., Elliott et al., 2021). Abadie and Camos (2019) showed that false memories in WM tasks in adults were underpinned like in DRM classical LTM tasks by an increase in gist memory. In this study, we used the simplified conjoint recognition model to assess memory traces underlying false memories. Results revealed that both age groups produced similar rates of false recognition. Moreover, false recognition was underlied by gist memory, which suggests that false memories occur at short and longer delays in WM tasks from age 4. As expected, the manipulation of maintenance mechanisms did not impact false recognition at 4-years. However, contrary to our expectations, it did not impact false or correct recognition in 8-year-olds either. Two interpretations were offered for this unexpected result. Either, maintenance mechanisms do not impact performance in 8-year-old children, or our paradigm did not trigger active maintenance in 8-year-olds, as the test used was recognition (e.g., Uittenhove et al., 2019).

4.2 Study 2

In a second study (i.e., Chapter 5), we adapted the DRM paradigm to a WM task. The methodology of this paradigm is considered as reliable and provides therefore a comparative baseline for new studies. As in the first study, this study aimed at assessing the effect of WM maintenance mechanisms on false memories in young and older children, but this time in an immediate recall test and on a delayed recognition test, following the DRM classical procedure. Five-year-olds and eight-year-olds had to learn lists of three or four semantically related words in each age group, respectively. To manipulate the possibility of rehearsing and refreshing the words, the concurrent task introduced in the retention intervals of 5 seconds between each word was either an aloud attentionally demanding task or on the contrary, a silent low attentionally demanding task. We compared an age group for which rehearsal is an emerging maintenance mechanism (e.g., Elliott et al., 2021; Henry et al., 2012) with an age group that was known to use both rehearsal and refreshing. To get a qualitative measure about the memory associated to semantic errors, in the immediate recall test, we assessed the phenomenological experience associated to recall responses by using a 3-option subjective judgement. This tool was developed for the purpose of the study. We used the simplified conjoint recognition model to estimate underlying memory traces of delayed false recognition. Results revealed a dominance of semantic errors in immediate recall in 8-year-olds, whereas 5-year-olds produced more non-semantic errors. Semantic recall errors increased with age, as found in the classical long-term DRM task. Most importantly, the suppression of WM maintenance mechanisms increased the dominant error type of each age group in immediate recall. Indeed, semantic errors were increased in 8-year-olds when under combined articulatory suppression and high cognitive load, while non-semantic errors increased under similar conditions in 5-year-olds. Phenomenological experience was similar across age groups and semantic errors were associated to all three subjective experience, i.e.- detailed experience, fuzzier experience or guessing. Moreover, as in our first study, 5 and 8-year-olds produced as many delayed false recognitions, which were underpinned mainly by gist traces. WM maintenance mechanisms were not found to impact false memories in the delayed test. These findings on recall tests challenge the findings from Study 1, using immediate recognition. Besides, as previous studies on false memories were conducted with adult participants and that Study 1 and 2 were conducted exclusively with children, it does not grant a direct comparison between the findings from previous studies and the ones from Study 1 and 2.

4.3 Study 3

In our final study (Chapter 6), we compared the occurrence of short-term false memories in older children (9-years) and young adults. We also compared short-term false memories in recognition and recall tasks. In addition to testing the impact of WM maintenance mechanisms on false memories, like we did on the two previous studies, we added a manipulation of semantic relatedness. This manipulation aimed at varying gist memory activation to assess its impact on false recognition and recall. In a complex span task, 6 word lists were interspersed with a concurrent task that prevented the use of both or neither WM maintenance mechanisms. The 6 words were composed of 3 semantically related and 3 unrelated words, a traditional manipulation known to increase gist memory in LTM (e.g., Abadie et al., 2021; Brainerd et al., 1999; 2001). Based on evidence showing that semantically related words boost false memories in WM tasks, because it boosts gist memory (i.e., Abadie & Camos, 2019), one could expect that false memories would be produced mainly on related associates in our study. Moreover, we expected a greater impact of WM maintenance mechanisms in adults than children, as the efficacy of these mechanisms develop with age. Finally, we measured gist, verbatim memory and guessing underlying false recognition and measured phenomenological experience associated to false recall, using a four option-subjective judgement response. Main results revealed that children produced as many false recognitions as adults. By contrast, semantic error became the dominant error type in adults, while children produced as many semantic and non-semantic recall errors. As expected, there were more false recognitions and semantic errors on related than unrelated associates, which confirms the role of gist in short-term false memories. Finally, there was a dissociative pattern between the effect of WM maintenance mechanisms on recognition and recall. Indeed, the conjoint suppression of these mechanisms increased semantic recall errors equally as much in children and adults. By contrast, it did not impact false recognition.

The general discussion provides a summary of the findings from experimental part, discusses the implications of such findings and offers some future perspectives of research.

Table 1.

Overview of the studies conducted in the experimental part, with the detail of each experiment per study (i.e., age and number of subjects, task used, list length, stimuli, concurrent task, manipulated variables and collected measures).

Study	Exp.	Subjects' age	Nb	Task	List length	Stimuli	Concurrent task	Manipulated variables	Measures
1 (CH5)	1	4- and 8-year-olds	N = 32	Brown-Peterson with immediate single probe recognition and delayed recognition	Three thematically related words	Images and auditory words	'baba' rehearsal versus none	Age	Response to target, semantically related and unrelated recognition probes (i.e., true and false recognition); gist, verbatim memory and guessing with the SCR
			N = 33						
	2		N = 33						
			N = 34						
	3		N = 31						
			N = 36						
2 (CH 6)	1	5- and 8-year-olds	N = 46	Complex span with immediate recall and delayed recognition	Three or 4 thematically related words for each age	Auditory words	Color aloud identification versus detection task	Age	Response to immediate recall, to delayed recognition (i.e. true and false memories).
			N = 43						

				group, respectively			WM maintenance suppression	Phenomenological experience associated to immediate recall (3 choices). Gist, verbatim memory, guessing associated to delayed recognition with the SCR
	2	N = 47 N = 48			Auditory words	Color aloud identification versus detection task		
	1	N = 46 N = 49	Complex span with immediate recognition				Age	Responses in recall and its associated phenomenological experience (4- options)
3 (CH 7)				6-word lists (3 semantically and 3 unrelated)	Auditory words	Parity judgment task: fast versus slow pace	WM maintenance suppression	
	2	N = 49 N = 49	Complex span with immediate recall				Words relatedness within lists	Responses to recognition probes (i.e. true and false memories); gist, verbatim memory and guessing with the SCR

Note. 'baba' rehearsal task aimed at preventing the use of articulatory rehearsal for stimuli maintenance; color aloud identification task aimed at preventing conjointly rehearsal and refreshing for stimuli maintenance; aloud parity judgment task, where participants had to say whether digits were odd or even, aimed at preventing the use of both rehearsal and refreshing. The faster digits were presented within a time interval, the more maintenance mechanisms were impaired. SCR = simplified conjoint recognition model; Nb = number of subjects per group respectively; CH = chapter.

Summary.

In this Chapter, I introduced the three studies conducted in the following experimental part and underlined the interest of adopting a developmental perspective to assess false memories in WM tasks.

Résumé.

Dans ce chapitre, j'ai introduit les trois études menées dans la partie expérimentale qui suit. J'ai souligné l'intérêt d'adopter une perspective développementale pour étudier les faux souvenirs en tâche de mémoire de travail.

Experimental part

Chapter 5. Children's Gist-based False Memory in Working Memory Tasks.

5.1 Introduction

Research on false memories in children has been the subject of intense debate over the past 30 years. It emerged from concerns about false reports by young children in certain types of legal cases (e.g., *Buckey v. County of Los Angeles*, 1992). The question that has received much attention is whether young children are more vulnerable to false memories than older children or adults. Research has first shown that children's memories, especially of younger children, are easily distorted by suggestions, whether they are blatant and deliberate or subtle and involuntary (for a review, see Howe, 2000). In addition to these early studies, much evidence has accumulated showing that young children are generally less susceptible to the well-known Deese-Roediger-McDermott illusion (DRM, Deese, 1959a; Roediger & McDermott, 1995) than older children and adults (e.g., Anastasi & Rhodes, 2008; Brainerd et al., 2018). False memories in the DRM paradigm are well-established episodic LTM phenomena: within minutes or days of learning lists composed of many semantic associates, other unstudied semantic associates are confidently and erroneously remembered as studied items. This powerful illusion does not depend on providing children with suggestive materials or leading questions, but is rather the consequence of the development of associative processes that enables older children to make meaning-based inferences about experienced events. Recent studies in young adults have shown that, surprisingly, the DRM illusion could also occur in WM tasks for lists of only a few items when a short 4-second interval was given between study and test (e.g., Abadie & Camos, 2019; Atkins & Reuter-Lorenz, 2008). However, this illusion in WM has never been studied in children. This question is of particular interest since WM is not fully developed in children (e.g., Camos & Barrouillet, 2018; Cowan, 2014). The aims of the present study were, first, to examine whether false memories could also occur in WM tasks in young children and whether they are prone to the same age-related increase as long-term false memories in the DRM paradigm and, second, to determine the role of WM in their occurrence.

The development of long-term false memories

In the DRM paradigm, participants study lists composed of the 15 strongest associates (e.g., “climber, peak, ski, valley, etc.”) to a given theme word (e.g., “mountain”) that is not itself presented, and are then given a recall or a recognition test that may occur minutes or days later. This paradigm typically produces high levels of false recall or recognition of the critical theme word in young adults, often rivaling the acceptance rate for presented words (e.g., Reyna & Lloyd, 1997; Roediger & McDermott, 1995). According to the Fuzzy-Trace theory (FTT, Brainerd & Reyna, 2002a; Brainerd et al., 2008a; Chang & Brainerd, 2021), subjects store two types of episodic representations as they encode the items in DRM lists: verbatim traces of item presentations and gist traces of item meanings, especially semantic relations that connect the different items of a given list. Verbatim traces are representations of surface features of items, i.e., auditory, visual, or other details associated with the item presentation (e.g., the color, font, size, position, etc. of an item; Brainerd et al., 2019). Gist traces are representations of the semantic content of items and other relational information such as taxonomic, synonymous, or situational relations. Access to verbatim and gist memories declines over time, but verbatim traces are faster to fade away (Brainerd & Reyna, 2002b) and are also more fragile and sensitive to cognitive load and interferences than gist traces (Abadie & Waroquier, 2020; Abadie et al., 2013; 2017). The FTT states that the retrieval of verbatim traces supports true memory for list words and suppresses false memories for critical distractors (remembering “climber”, “peak”, produces rejection of “mountain”; Brainerd et al., 2003), while gist retrieval supports both true and false memories (remembering that the gist of the list was “mountain” produces acceptance of “climber” but also of “mountain”; Brainerd & Reyna, 2005). Because they are associated, DRM-list words repeatedly cue list themes (e.g., “mountain”) resulting in very strong gist memory of these themes, which are especially likely to be falsely recalled or recognized.

Many studies have investigated the occurrence of the DRM illusion in children in relation to the development of verbatim and gist traces. Brainerd and Reyna (1998) and Ceci and Bruck (1998) were the first to propose that false memories increase with age. Soon, several investigators found that the false memory effect is at near floor levels in young children aged 5 to 7 years and intensifies throughout childhood and adolescence (Anastasi & Rhodes, 2008; Brainerd et al., 2018; Sugrue & Hayne, 2006; Wimmer & Howe, 2009). The FTT accounts for these findings by stating that acquisition, retention and retrieval of both gist and verbatim memories improve from childhood to adulthood (see Brainerd & Reyna, 2002a; Brainerd & Reyna, 2015, for reviews). Although both verbatim and gist memories increase from early

childhood onwards, the increase is sharper for gist than for verbatim traces (Brainerd et al., 2002c; Reyna & Kiernan, 1994). Gist memory improves during childhood, because the ability to process meaning of an item and to connect meaning across different items improves. Semantic clustering in long-term recall or recognition increases during childhood, and intrusions become increasingly semantic with age. It is well established that the magnitude of this trend can be influenced by many theoretically driven manipulations. For instance, encouraging 7- and 11-year-olds to search for meaningful relations when items of DRM lists are encoded or when DRM lists are repeatedly tested boosts reliance on gist, leading to an increase in long-term false recognition for both age groups. By contrast, focusing on processing superficial characteristics of the stimuli, such as spelling, reduces false memories in both age groups (Holliday et al., 2011; Odegard et al., 2008). Other studies presented DRM-list words in stories that revolved around the lists' themes (Dewhurst et al., 2007; Howe & Wilkinson, 2011). Stories increased long-term false memories in 5-year-olds but had little effect in 8- and 11-year-olds.

Toward a new account of short- and long-term false memories

Although the phenomenon of false memories has been widely studied in the field of LTM, a few studies in young adults have reported that false memories can also occur in WM tasks within seconds of studying a small number of semantically associated words (e.g., Atkins & Reuter-Lorenz, 2008; Coane et al., 2007; Flegal et al., 2010; Flegal & Reuter-Lorenz, 2014). At first glance, these findings suggest that processes that give rise to false memories are not exclusive to LTM, but may also be active in WM, a limited capacity system for maintaining relevant representations over the short term. To thoroughly examine the role of WM on the false memory effect, a series of experiments were conducted in young adults using a Brown Peterson paradigm (Brown, 1958; Peterson & Peterson, 1959) with an immediate and a delayed recognition test of DRM-like lists (Abadie & Camos, 2019). Participants had to maintain four highly or weakly semantically related words for 4 seconds before recognizing them among related and unrelated distractors. The 4-second retention interval was filled by a concurrent task of varying difficulty to manipulate the possibility that the items could be actively maintained in WM. Two experiments demonstrated the role of WM in the occurrence of false memories: active maintenance of word-lists in WM prevented false recognition of related distractors in the immediate test but elicited false recognition in the delayed test. Abadie and Camos (2019) therefore proposed that active maintenance in WM would prevent the occurrence of false

memories in short-term tests. It would also participate in their formation in long-term tests such as in the classical DRM paradigm.

In the time-based resource-sharing model (TBRS, Barrouillet & Camos, 2015), two mechanisms have been described as responsible for the maintenance of information in WM: articulatory rehearsal, which operates by articulatory repetition of memory items, and attentional refreshing, a domain-general attention-based mechanism that operates by briefly thinking back to recently active memory items (Camos, 2015, 2017). Studies demonstrated that articulatory rehearsal and attentional refreshing had independent and additive effects on immediate recall. Manipulating the availability of one mechanism while controlling for the other led to poorer immediate recall performance. To disentangle the role of each maintenance mechanism on the emergence of false memories, Abadie and Camos (2019) varied the attentional demand of the concurrent task and the presence (or not) of a concurrent articulation during the retention interval of a Brown-Peterson task to impede either refreshing or rehearsal, respectively. Results revealed that using rehearsal prevents false recognition of related distractors in the immediate test whereas refreshing increases false recognition of related distractors in the delayed test.

By integrating the functional description of WM from the TBRS model with the FTT, Abadie and Camos (2019) proposed a new model that allows for a precise depiction of how each WM maintenance mechanism affects the creation of verbatim and gist traces and thus the formation of false memories. Consistent with the predictions of the FTT, this model predicts that false memories in both the immediate and delayed tests are underpinned by the retrieval of gist traces. Specifically, the use of rehearsal should promote the retrieval of verbatim traces that oppose gist traces in the immediate test, hence reducing the occurrence of short-term false memories. By contrast, the use of refreshing is assumed to foster the creation of both verbatim and gist traces. Gist traces are counteracted by verbatim traces in immediate tests, but as they are less sensitive to loss than the latter (e.g., Brainerd & Reyna, 2005), they can support memory retrieval after long delay of retention. Thus, impairing refreshing should reduce both verbatim and gist retrieval, resulting in fewer correct recalls or recognitions in immediate tests. However, in delayed tests, as performance relies more on gist retrieval, preventing the use of refreshing should decrease both true and false memories. These predictions on correct and false recognition as well as on measures of verbatim and gist representations were verified (Abadie & Camos, 2019).

Extending this model to childhood, a seemingly self-evident prediction is that the occurrence of false memories in a WM task will increase with age as it does in the classical DRM episodic memory task due to the development of gist memory. However, WM capacity is characterized by a significant age-related increase during childhood (Camos & Barrouillet, 2018; Cowan, 2014, for reviews). Memory span, the number of items successfully maintained in WM, increases throughout childhood, until early adolescence (e.g., Gathercole et al., 2004). Moreover, research showed that children of about 4-years of age and younger do not spontaneously attempt to refresh or rehearse memory items when engaged in a concurrent task. WM maintenance mechanisms become efficient from age 7 onwards and keep increasing in efficiency until adolescence, which underpins a sizable part of WM increase with age (Henry, 2012, for a review). Thus, if, as predicted by Abadie and Camos' (2019) model, WM plays an important role in preventing short-term false memories, young children under the age of 7 may not be able to maintain information correctly and may produce semantic errors in WM tasks. Regarding the use of each main WM maintenance mechanism, there is substantial evidence that rehearsal starts to be efficient between the ages of 5 and 7 (Henry et al., 2012; Tam et al., 2010), while children would start to use refreshing at age 7 (Barrouillet et al., 2009; Camos & Barrouillet, 2011). In addition, since the use of refreshing is attention-demanding, children favor the use of rehearsal whenever they can. Recent studies suggested that by the age of 6, children use rehearsal as a default maintenance strategy, while from age 8 onwards they could adaptively switch between rehearsal and refreshing (Oftinger & Camos, 2018). So far, the literature on WM development has focused on true memory, and no studies have yet investigated the role of these mechanisms in the formation of false memories in children.

The present study

We report two experiments⁵ that were designed to examine, first, whether false memories can occur in a WM task in children and whether they increase with age as in the classical DRM paradigm. Secondly, these experiments tested Abadie and Camos' (2019) model regarding the role of WM maintenance mechanisms in the occurrence of false memories and the underlying verbatim and gist representations. We compared false memories in 4- and 8-year-old children since the shift in the use of WM maintenance mechanisms happens between

⁵ A third experiment is reported as a supplementary material on the OSF. Its objectives and results are outlined in the discussion of Experiment 2.

these ages. Therefore, 4-year-olds should not yet be using WM maintenance mechanisms, whereas 8-year-olds should spontaneously use maintenance strategies. Children were presented with DRM-like lists of 3-5 items represented by pictures and/or auditorily-presented words. The number of memory items was determined based on previous studies in children of similar ages (e.g., Barrouillet et al., 2009; Hitch et al., 1989). The presentation of each list was followed by a retention interval of a few seconds filled with a concurrent task in a Brown-Peterson paradigm. The type of concurrent task varied in each experiment to manipulate the possibility of implementing WM maintenance mechanisms. As in previous studies in adults, following the retention interval, an immediate recognition test composed of target items, related and unrelated distractors was proposed to the children. In Experiment 2, following the completion of all trials, participants completed a delayed recognition test similar to the immediate test. This test was added in order to compare the results with those obtained in the classical DRM paradigm. Finally, the simplified version of the conjoint recognition (SCR) model of the FTT (Brainerd, et al., 1999; Stahl & Klauer, 2008) was used to directly measure the verbatim and gist representations that underlie true and false memory in the immediate and delayed test. This model delivers precise quantitative measurements of the contributions of verbatim and gist memory for the responses to the different types of probes.

Based on findings on the development of verbatim and gist memory (e.g., Brainerd et al., 2018), we expected that 8-year-old children would retrieve more verbatim and gist traces than 4-year-olds, which would increase both true recognition of target probes and false recognition of related distractors in the immediate and delayed tests in 8-year-olds. Because WM maintenance mechanisms are efficient from 7 (e.g., Barrouillet et al., 2009; Camos & Barrouillet, 2011), correct recognition in the immediate test should be reduced in 8-year-olds when the use of WM maintenance mechanisms was prevented relative to when it was not, whereas such a manipulation should not affect correct recognition in 4-year-olds. Most importantly, results in young adults showed that false recognition increased in the immediate test when the use of WM maintenance, specifically rehearsal, was blocked (Abadie & Camos, 2019). Therefore, we also expected that the rate of false recognition of related distractors would increase in 8-year-olds when they could not use any WM maintenance mechanism. In addition, as predicted by Abadie and Camos (2019), verbatim and gist traces retrieval should be affected by the manipulation of WM maintenance mechanisms, but only in 8-year-olds. For this age group, verbatim retrieval in the immediate test should be reduced when WM maintenance was prevented, while gist retrieval should increase. Finally, the present study also examined whether

the use of WM maintenance strategies affects long-term information retention in young children. Studies indicated that the use of refreshing increased long-term true and false memories in young adults (Abadie & Camos, 2019; Camos & Portrat, 2015). However, no study has yet examined the long-term effects of using refreshing in children. Therefore, the delayed test in Experiment 2 aimed at examining this issue.

5.2 Experiment 1

Four- and 8-year-old children performed a Brown-Peterson task in which they had to maintain lists of auditory-presented words, each of them being associated with a drawing. The words in a given list were all related to one another and to a common theme, just as in the DRM lists, with the only difference being that relevant themes and words were selected so that they would be familiar to children as young as 4 –years old. The number of words presented was determined based on previous studies showing that 5- to 6-year-olds are able to recall 2-3 items correctly and that 8-year-olds can recall 3-4 items correctly in a WM task (e.g., Barrouillet et al., 2009; Hitch et al., 1989). Since recognition test is less difficult than recall, lists of three words were presented. Children performed the immediate recognition test after a short retention delay. In one condition, they performed a concurrent articulation task during the retention interval to prevent rehearsal, while in the other both rehearsal and refreshing were preserved. Only the use of rehearsal was manipulated in this experiment since this is the most influential mechanism in preventing false memory at short term in adults (Abadie & Camos, 2019).

5.2.1 Method

Sample size

We computed a meta-analytic effect size across studies that reported an interaction between age and the availability of WM maintenance mechanisms on recall performance in children aged 5-14 years using a procedure similar to the present one (Barrouillet et al., 2009; Camos & Barrouillet, 2011). The meta-analytic effect size was $\eta_p^2 = .131$; it was calculated by weighting the η_p^2 of the interaction obtained in each study by the number of participants in the study. Power analysis using G*Power (Faul, et al., 2009) indicated that an overall sample of 56 participants would be needed to achieve an 80% power. In the experiments reported here, we systematically collected larger samples to accommodate for potential data loss. This study was not preregistered.

Participants

Sixty-seven French kindergartners and 3rd graders were recruited from several preschools and elementary schools, respectively, in France to participate to the study. Two of them were excluded from the data analysis because they did not understand the instructions. Out of the 65 remaining children, 32 were kindergartners (the 4-year-old group; 18 girls; *mean age* = 52.92 months; *SD* = 3.16 months) and 33 were 3rd graders (the 8-year-old group; 19 girls; *mean age* = 102.3 months; *SD* = 4.10 months). All children were native French speakers. Although we were not able to collect any other demographic variables, we suspect the majority of participants to be from European origin and coming from medium to medium-high socioeconomic backgrounds given the location where the data was collected. The study was conducted in accordance with the local institutional guidelines and APA Ethics Code. For each participant, a written parental consent was obtained. Ethic approval was obtained for both experiments from the institutional review board of Aix-Marseille University (“Working memory and false memories during childhood”, protocol number: 2019-12-12-003).

Material

For the study phase, two subsets of 18 lists were created corresponding to the following thematic categories: circus, farm, school, supermarket, bedroom, soccer game, pirates, fire station, birthday party, Christmas, princesses, beach, zoo, forest, hospital, bathroom, snack time, kitchen (see <https://osf.io/9gfmd/> for the lists development). Each list included three memory items and the average word frequencies of the two subsets did not differ significantly ($t = .08, p = .94, d = .02$). Children were exposed to one of two subsets.

For the recognition phase, two additional lists of 18 words were constructed. Each of them included six targets (i.e., words from the presented subset), six related distractors (i.e., unrepresented words from the other subset), and six unrelated distractors (i.e., unrepresented concrete words that are not related to any of the thematic categories; the same unrelated distractors were used in both lists). For each of the 18 three-word lists studied, a target probe, a related distractor or an unrelated distractor from the two recognition lists was presented at test. The type of probe presented at test was counterbalanced for each studied list.

Memory items were recorded via ‘audacity’ software (Audacity, 2016) to be auditory presented. Each word was associated with its representation drawn in color. Drawings were copyright-free images found on the internet, pre-selected by the authors of this study to match

the words they were to represent. Additionally, three training lists in which all the words pertained to the category “vegetables” were created in the same manner.

Procedure

The study took place in a quiet room of the school. Children were asked to memorize the lists of words related to each category presented as a ‘world’ (e.g., ‘the world of pirates’; ‘the world of princesses’) for a subsequent recognition test. In each trial, children first heard the name of the category, then the 3 items of a list were sequentially presented on screen for 2250 ms per word with a 250 ms interstimulus interval (ISI)⁶. Words were presented in one of three predefined orders. The three presentation orders were counterbalanced across participants.

Next, for half of the 18 trials (i.e., with articulatory suppression), children had to repeat the syllable “ba” aloud at a pace of 1 syllable per second throughout a 6 s retention interval. The experimenter repeated the syllables simultaneously to encourage the children to maintain the repetition until the end of the retention interval. For the other half of the trials (i.e., without articulatory suppression), children remained silent. In each condition (i.e., with and without articulatory suppression), the lists were presented in random order, the conditions as a block, and their order was counterbalanced across participants.

Then, at the end of the trial, children were presented with a probe (i.e., a target, a related or an unrelated distractor) and asked to judge whether or not the probe was in the studied list (Figure 6). A “yes” response was classified as a “target” response. If a “no” response was given, the experimenter would then ask whether or not this probe could belong to the studied category. A “yes” response to this second question was classified as a “related” response, and a “no” as an “unrelated” response. This procedure inspired by the SCR paradigm (Stahl & Klauer, 2008) was used to estimate the contribution of gist and verbatim representations to recognition performance.

A training phase preceding the experiment, in which the two conditions (with and without concurrent articulation) were seen in a random order, was presented to the children. At the end of the experiment, children were rewarded with a diploma.

⁶ Presentation times of memory items and duration of the retention interval were mapped on previous studies examining WM in children (Barrouillet et al., 2009; Camos & Barrouillet, 2011; Fitamen et al., 2019; Oftringer & Camos, 2018)

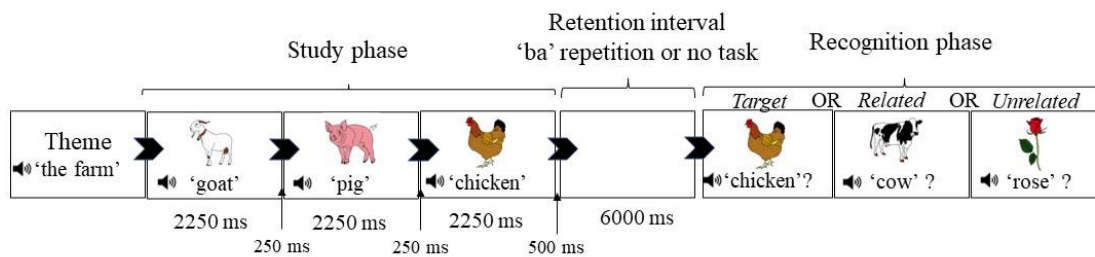


Figure 6. Illustration of one trial of the Brown-Peterson task in Experiment 1. Images were retrieved from “Multipic” open access database (Duñabeitia et al., 2018).

5.2.2 Results

Data for all experiments are available at <https://osf.io/9gfm/>

Memory accuracy

We computed discriminability indexes (Pr) to eliminate potential developmental differences in ‘yes-saying’ bias (i.e., the tendency of young children to answer ‘yes’ to a question more often than older children, e.g., Otgaar et al., 2014). True and false recognition were conditionalized by subtracting the baseline false recognition rate of unrelated distractors (‘yes’ responses to unrelated distractors) from the rate of correct recognition of targets (‘yes’ responses to targets; true recognition) and from the rate of false recognition of related distractors as targets (‘yes’ responses to related distractors; false recognition). True and false recognition data are shown in Figure 7.

Mixed-measure ANOVAs were conducted on true and false recognition with age group as between-subject factor (4 vs. 8 years) and type of concurrent task performed during the retention interval (“ba” repetition vs. no task) as within-subject factor. As expected, the main effect of age on true recognition was significant ($F(1, 63) = 5.81, p = .019, \eta_p^2 = .084, 90\% \text{ CI } [.008; .204]$). Eight-year-olds made more true recognition than 4-year-olds. By contrast, age had no significant effect on false recognition ($F(1, 63) = 3.48, p = .067, \eta_p^2 = .052, 90\% \text{ CI } [0; .161]$). Contrary to our prediction, the false recognition rate tended to be even higher in 4- than in 8-year-olds. However, a Bayesian paired-sample t-test provided weak evidence for an age-

related difference ($BF_{10} = 1.09$). The presence of a concurrent articulation during the retention interval did not significantly impact correct nor false recognition ($F_s < .09$). No interaction with age group was significant ($F_s \leq .18$).

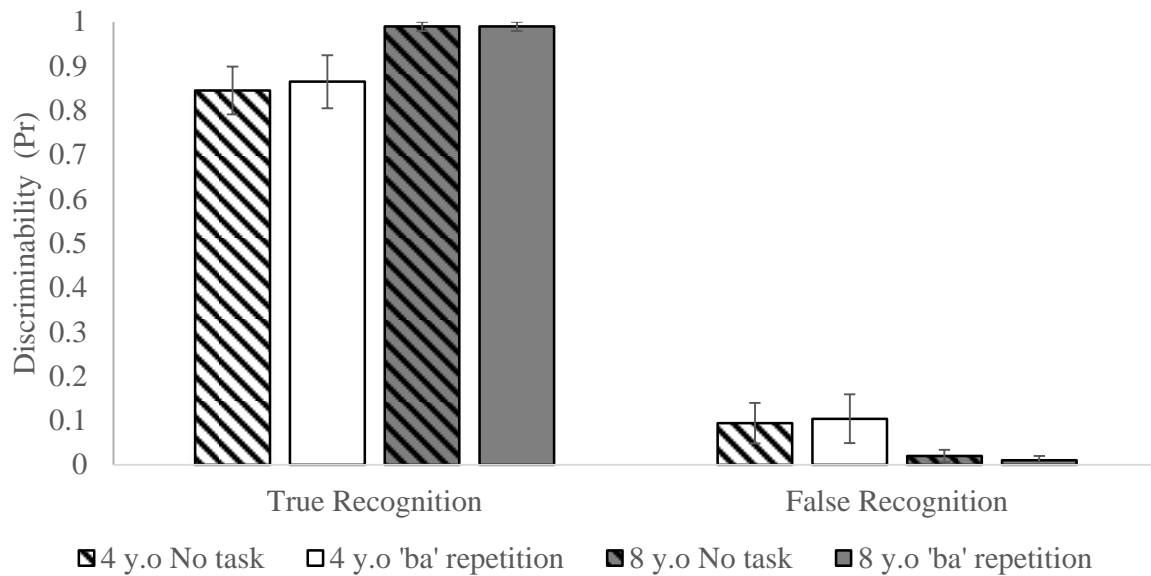


Figure 7. True and false recognition accuracy as a function of age group and type of concurrent task in Experiment 1. Error bars represent standard errors.

Gist and verbatim memory

We used the SCR (Stahl & Klauer, 2008) of the FTT to compute four memory parameter estimates and two guessing parameters from the classification of responses as “target”, “related” or “unrelated” for each probe type. Two parameters V represent the probability of retrieving a verbatim trace of a target when either a target probe (V_t) or a related distractor (V_r) was presented at test. Verbatim memory traces lead to correct identification of target and related probes. The two parameters G represent the probability of participants retrieving a gist trace of a target for either a target probe (G_t) or a related distractor (G_r), given that they have not retrieved its verbatim trace. When participants retrieve gist memory, they cannot remember whether the probe itself or a related word with the same gist had been presented at study. This leads them to guess whether the probe is a target (with probability a) or a related distractor (with probability $1 - a$). Therefore, gist retrieval can lead to correct identification of target and related probes or to error in identification of a target as a related distractor or of a related probe

as a target item. When neither verbatim nor gist is available, a participant can still guess that the probe meaning is old with the probability b . The decision between “target” and “related” responses is again modelled by the parameter a . Otherwise, the participant guesses that the probe is new with the probability $1 - b$. Verbatim and gist traces do not intervene when responding to unrelated distractors because these probes do not trigger the retrieval of verbatim or gist representations in the study phase. Therefore, the responses to unrelated distractors are based entirely on b . The processing tree representation of the model is given in Figure 8. The model equations are provided in the appendix of Stahl and Klauer’s (2008) article.

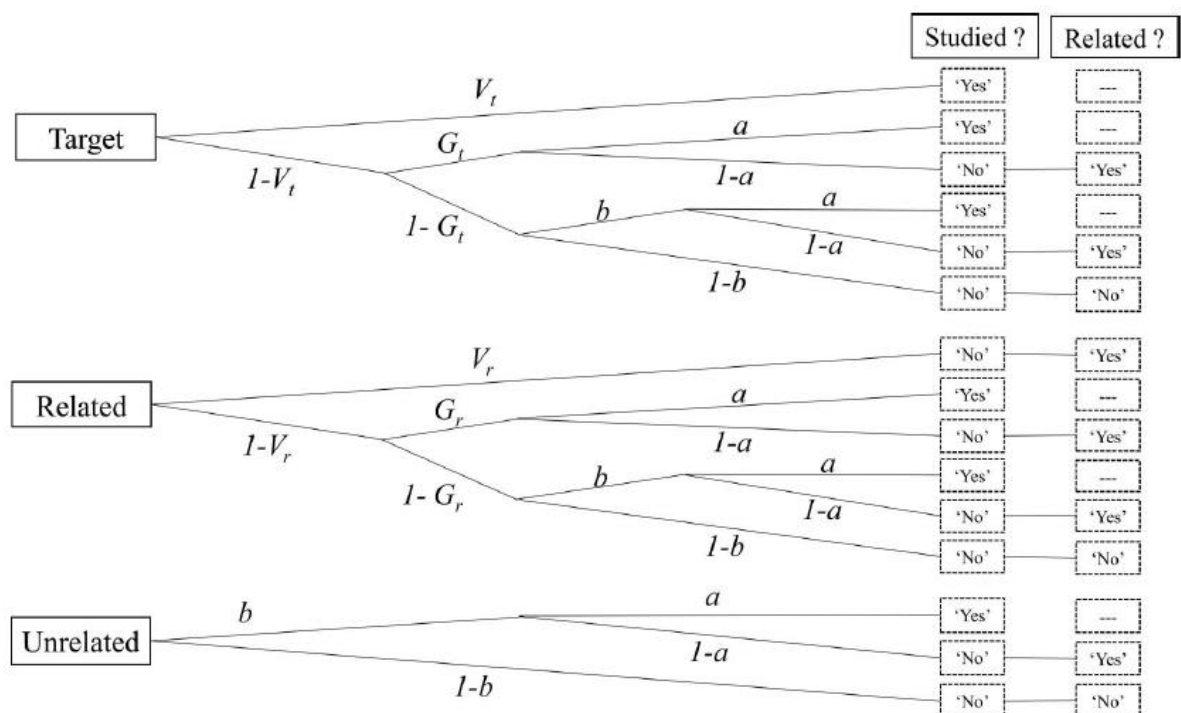


Figure 8. Multinomial Processing Tree for the simplified conjoint recognition paradigm based on Stahl and Klauer (2008). Rectangles on the left represent probe types and rectangles on the right represent responses to questions “Did you studied this word?” and “Could it be related of the presented list?”. They are connected by branches of the processing tree that represent the combination of cognitive processes postulated by the model. V_t = probability of retrieving a target’s verbatim trace given a target probe; V_r = probability of retrieving a target’s verbatim trace given a related probe; G_t = probability of retrieving a target’s gist trace given a target probe; G_r = probability of retrieving a target’s gist trace given a related probe; b = probability of guessing that an item is either a target or a related probe; a = probability of guessing “target”. The figure is adapted from Abadie and Camos (2019).

Verbatim, gist and bias parameter estimates are given in Table 2. We used MultiTree Software (Moshagen, 2010) to perform parameter estimations and hypotheses tests. To test the goodness of fit of our data with the model, we tested whether there were significant violations of the model's four inequality constraints, following Stahl and Klauer's explanations (2008) (see model fit section). Sixteen tests were conducted, none of them revealed a significant violation of the model.

Table 2.

Estimates for the parameters of the simplified conjoint recognition model as a function of age group and type of concurrent task in Experiment 1

Parameter	4 y.o		8 y.o	
	No Task	'Ba' repetition	No Task	'Ba' repetition
V_t	.86 [.39, 1.33]	.88 [.78, .97]	.99 [.97, 1.01]	.99 [.97, 1.01]
V_r	.47 [-.94, 1.88]	.52 [.18, .86]	.00 [-3.00, 3.00]	.00 [-4.52, 4.52]
G_t	1.00 [-3.08, 5.08]	.80 [.53, 1.07]	1.00 [-1.21, 3.21]	1.00 [-1.18, 3.18]
G_r	.79 [.24, 1.34]	.82 [.65, .99]	.97 [.86, 1.07]	.98 [.87, 1.08]
b	.16 [.05, .26]	.13 [.07, .20]	.11 [.05, .17]	.10 [.04, .16]
a	.33 [-.73, 1.39]	.38 [.12, .65]	.02 [-.04, .08]	.01 [-.03, .05]

Note. These estimates are the proportion of responses to the target, related and unrelated probes in the recognition task based on verbatim memory, gist memory or guessing. 95% confidence intervals are in brackets.

As expected, verbatim memory for targets was greater in 8- than in 4-year-old children ($\Delta G^2(2) = 20.23, p < .001$). Neither verbatim memory for related distractors nor gist memory for targets or for related distractors were significantly impacted by age ($\Delta G^2(2) = .56, p = .76$; $\Delta G^2(2) = .37, p = .83$; $\Delta G^2(2) = 2.61, p = .27$, respectively). As mentioned above, according to the SCR model, the retrieval of gist traces occurs when verbatim traces of the studied items are not accessible (i.e., $1 - V_t$ for target probes or $1 - V_r$ for related probes) and can result in two

types of responses: the “target” response with probability $(1 - V_t) \times G_t \times a$ for target probes and $(1 - V_r) \times G_r \times a$ for related probes, or the “related” response with probability, $(1 - V_t) \times G_t \times (1 - a)$ for target probes and $(1 - V_r) \times G_r \times (1 - a)$ for related probes. Thus, considering responses given to related probes on the basis of gist retrieval, it can be noted that gist trace retrieval led 8-year-olds to correctly identifying related distractors in most cases (96%). By contrast, 4-year-olds identified related distractors either as related distractors (26% of cases) or as targets (14% of cases) on the basis of gist memory. Interestingly, 4 year-olds did not guess significantly more often than 8-year-olds that an item meaning was old (b ; $\Delta G^2(2) = 1.42, p = .49$). However, when they identified an item as old on the basis of guessing or gist memory, 4-year-olds were more likely to identify it as a target than as a related distractor (a ; $\Delta G^2(2) = 12.49, p = .002$). The concurrent task manipulation had no significant impact on the memory and guessing parameters ($\Delta G^2s \leq 3.01$).

5.2.3 Discussion

As expected, true recognition rate increased with age. Contrary to our predictions, false recognition did not vary significantly as a function of age. Interestingly, in our short-term DRM-like paradigm, 4-year-olds tended to make as many (or even more) false recognitions of related distractors as 8-year-olds. Moreover, memory accuracy of children in both age groups was not significantly affected by articulatory suppression introduced during the retention interval. The analysis of verbatim and gist representations provided a very good account of the recognition performance.

First, verbatim memory increased with age, which perfectly fits the increase in true recognition. Second, gist memory was not significantly affected by age, which directly explains the absence of significant age-related differences in false recognition rates. There was, however, a qualitative difference in the use of gist traces between 4- and 8-year-olds. Indeed, gist retrieval primarily led 8-year-olds to correctly identify related probes as such, whereas it led 4-year-olds to produce both correct and false recognitions. In addition, the tendency of younger children to judge probes with familiar meaning as targets (parameter a) seems to be explained by a different use of gist memory rather than by response bias (i.e., the probability of guessing that an item is either a target or a related probe, which is modeled by parameter b). Indeed, the parameter b did not differ between age groups and only accounted for 5% of the target responses in younger children. Hence, it seems that even 4-year-old children were able to process the gist of our short

DRM-like lists. Although it can be surprising that such young children were able to extract the gist of DRM lists, it should be remembered that our lists were specifically tailored for them, which could have facilitated gist encoding. Nevertheless, 4-year-olds also appeared to use gist traces differently than 8-year-olds.

In the present experiment, neither memory accuracy nor verbatim and gist representations in 8-year-olds were affected by our manipulation of rehearsal. This might suggest that older children have used strategies other than rehearsal. Moreover, the task was easier for older than younger children as list lengths were identical between age groups. This could explain the absence of a significant difference in false recognition rate between the age groups despite a qualitatively different use of gist memory. Hence, in Experiment 2, our manipulations of the concurrent task targeted not just rehearsal, but also refreshing, while adjusting the difficulty of the task to age group.

5.3 Experiment 2

The method of Experiment 2 was similar to that of Experiment 1, with the exception of the number of memory words and their presentation time being adapted to the age of the children, as older children have larger WM capacity and faster processing speed (Camos & Barrouillet, 2018; Cowan, 2014, for reviews). Moreover, the memory items were presented only auditorily to facilitate the use of rehearsal (e.g., Baddeley, 1986; Henry, 2012). As in Experiment 1, we contrasted two conditions that differed in the concurrent task implemented during retention interval. However, in this study the availability of the two WM maintenance mechanisms, rehearsal and refreshing, was manipulated. In one condition, the use of both WM mechanisms was impaired by a concurrent task involving both articulatory suppression and high attentional demand. Children were asked to name aloud the color of smileys while pressing a corresponding-colored key. This concurrent task is highly attention-demanding for children and prevents the use of rehearsal (Barrouillet et al., 2009; Camos & Barrouillet, 2011; Fitamen et al., 2019). In the other condition, the concurrent task was a simple detection task performed silently. This task did not involve any concurrent articulation and had a low attentional demand, thus allowing the use of the two WM maintenance mechanisms. Finally, a delayed recognition test was added to assess true and false recognition at longer delays in order to compare our results with those obtained in the classical DRM paradigm.

5.3.1 Method

Participants

Thirty-three French kindergartners (4-year-old group; 19 girls; *mean age* = 57.96 months, *SD* = 3.06 months) and 34 3rd graders (8-year-old group; 19 girls; *mean age* = 102.26 months, *SD* = 4.10 months) were recruited from several schools in France. All children spoke French as their first language and did not participate to Experiment 1. As in the first experiment, the majority of participants were assumed to be mainly from European origin and coming from medium to medium-high socioeconomic backgrounds given the location where the data was collected. For recruitment purposes, parents were sent a written description of the study and signed a consent form. Data of one kindergartner were excluded from the analysis because his performance in the high-demanding concurrent task was below two standard deviations of the group's average performance. Moreover, due to technical problems, two 3rd graders did not achieve the delayed recognition test, and their data were therefore not included in the analysis of this test.

Material

Word lists in this experiment were similar to the ones used in Experiment 1. The two subsets from Experiment 1 were mixed to get 6 words per thematic category. These lists were validated by two tasks conducted prior to the present experiment (see the OSF page).

Procedure

The procedure was similar to that of Experiment 1, but changes were made to adjust the task difficulty to each age group (Figure 9). In each trial, children first heard the name of the category, then four or five words were presented sequentially and in a random order on screen to 4- and 8-year-old children, respectively. Words were presented for 2000 ms for 8-year-olds and 2500 ms for 4-year-olds with an ISI of 250 ms in both age groups. Next, a 500 ms “beep” sound signaled the start of the concurrent task. The attentional demand of the concurrent task was adjusted to each age group based on the results of the titration procedure described on the OSF. In order to vary the availability of refreshing, the concurrent task was either high or low attention demanding. In the high attention demanding condition, three or five smileys appeared sequentially on screen for 2167 ms or 1100 ms followed by an ISI of 500 ms for the 4- and 8-

year-old children, respectively. To prevent the use of rehearsal, children had to name aloud the color of each smiley while pressing a corresponding-colored key. In the low attention demanding condition, circles appeared sequentially on the screen and children had to press the space bar for each one of them. The number of circles was equal to the number of smileys presented in the high demanding concurrent task. Conditions were presented in blocks, the order of which was counterbalanced across participants. At the end of all trials, children completed a delayed recognition task. Six targets, 6 related distractors and 6 unrelated distractors were sequentially presented in a random order. Children had to identify each of them using the same procedure as in the immediate recognition task.

A training phase preceded the experimental phase. First, children completed the two concurrent tasks alone on 12 stimuli each. Then, they performed two trials of the Brown-Peterson task for each condition.

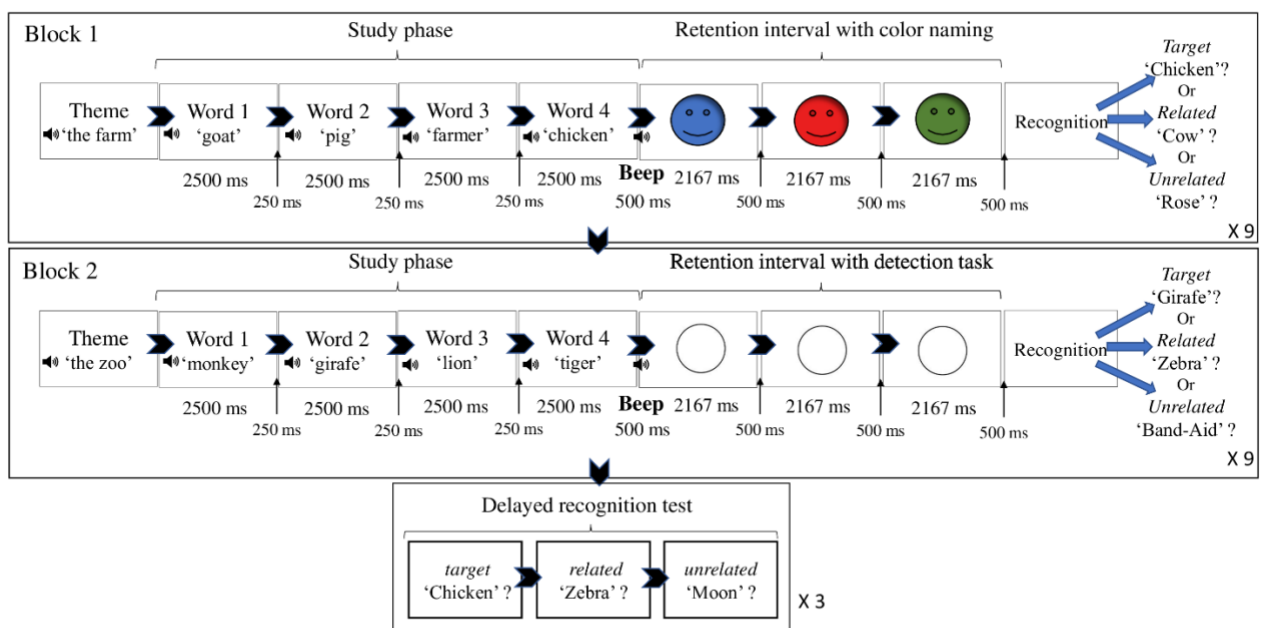


Figure 9. Illustration of trials in Experiment 2. Eight-year-old group was shown one extra-word during the study phase and had to process one more stimulus than the 4-year-olds during the retention interval. The order of presentation of the blocks was counterbalanced among participants.

5.3.2 Results

Similar analyses as in Experiment 1 were conducted on both immediate and delayed recognition tests. Mean performance in both concurrent tasks were above 80% accuracy in each age group (detection $M_{4years} = 85.5\%$, $SD = 9.3$; $M_{8years} = 93.2\%$, $SD = 7.8$; color naming: $M_{4years} = 82.9\%$, $SD = 13.06$; $M_{8years} = 87.7\%$, $SD = 10.0$), which showed that children followed the instructions.

Memory accuracy

Mixed-measure ANOVAs were conducted on Pr indexes for true and false recognition with age group as between-subject factor (4 vs. 8 years old) and type of concurrent task (detection vs. color naming) as within-subject factor for each probe type and time of test separately. There was a significant main effect of age on true recognition rate in both immediate and delayed tests ($F(1, 64) = 18.31$, $p < .001$, $\eta_p^2 = .23$, 90% CI [.087; .354]; $F(1, 64) = 4.85$, $p = .03$, $\eta_p^2 = .07$, 90% CI [.003; .185], respectively). As predicted, the true recognition rate was higher in 8-year-olds than 4-year-olds in both tests (Figure 10). However, as in Experiment 1, false recognition rate was not significantly affected by age in either the immediate or delayed tests ($F(1, 64) = .16$, $p = .69$, $\eta_p^2 = .003$, 90% CI [0; .055]; $F(1, 64) = .05$, $p = .83$, $\eta_p^2 = 7.58 \times 10^{-5}$, 90% CI [0; .037], respectively). Moreover, a Bayesian t-test provided substantial evidence for an absence of age-related difference in false recognition in both tests ($BF_{10} = .27$; $BF_{10} = .26$, respectively). Finally, the type of concurrent task did not significantly affect correct nor false recognition and there was no interaction with age ($F_s \leq 1.52$).

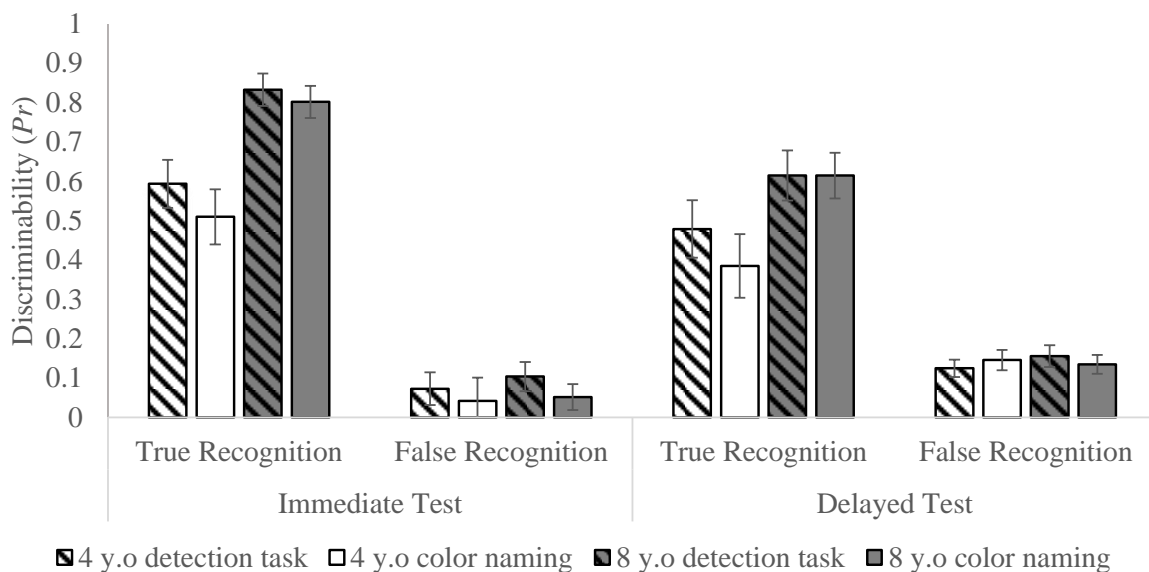


Figure 10. True and false recognition accuracy as a function of age group and type of concurrent task for immediate and delayed tests in Experiment 2. Error bars represent standard errors.

Gist and verbatim memory

Verbatim, gist and bias parameter estimates for immediate and delayed recognition tests are given in Table 3. As in Experiment 1, we tested whether there were significant violations of the model's four inequality constraints. Sixteen tests were conducted for the immediate and delayed test, none of them revealed a significant violation of the model.

Table 3.

Estimates for the parameters of the SCR model as a function of age groups and concurrent tasks in both immediate and delayed tests in Experiment 2

Time of test	Parameter	4 y.o		8 y.o	
		Detection task	Color Naming	Detection task	Color Naming
Immediate	V_i	.67 [.51, .83]	.60 [.41, .78]	.79 [.70, .88]	.80 [.71, .88]
	V_r	.25 [-.02, .52]	.22 [-.02, .46]	.00 [-1.25, 1.25]	.00 [-2.64, 2.64]
	G_i	.77 [.54, 1.00]	.68 [.41, .95]	.95 [.84, 1.05]	.95 [.85, 1.05]
	G_r	.65 [.43, .86]	.51 [.25, .77]	.95 [.88, 1.02]	.97 [.88, 1.06]
	b	.45 [.35, .55]	.51 [.41, .61]	.11 [.05, .17]	.05 [.01, .09]
	a	.46 [.31, .61]	.51 [.37, .65]	.09 [-.03, .22]	.05 [-.08, .18]
Delayed	V_i	.54 [.32, .77]	.41 [.12, .71]	.55 [.42, .67]	.57 [.45, .69]
	V_r	.00 [-.16, .16]	.01 [-.15, .16]	.00 [-.58, .58]	.00 [-.64, .64]
	G_i	.48 [.13, .83]	.39 [.02, .76]	.70 [.52, .87]	.68 [.50, .86]
	G_r	.22 [-.04, .48]	.35 [.13, .58]	.74 [.56, .92]	.74 [.55, .93]

<i>b</i>	.47 [.37, .57]	.42 [.32, .52]	.22 [.13, .30]	.22 [.13, .29]
<i>a</i>	.65 [.53, .78]	.72 [.59, .86]	.16 [.04, .27]	.16 [.04, .28]

Note. 95% confidence intervals are in brackets.

Immediate test. As expected and observed in Experiment 1, verbatim memory for targets was greater in 8- than in 4-year-olds ($\Delta G^2(2) = 6.72, p = .034$). There was a significant effect of age on verbatim memory for targets when the concurrent task was the color naming task ($\Delta G^2(1) = 4.67, p = .031$), but not for the detection task ($\Delta G^2(1) = 2.05, p = .15$). Moreover, 8-year-olds exhibited greater gist memory for targets than 4-year-olds ($\Delta G^2(2) = 6.20, p = .04$). This effect was also significant when the concurrent task was the color naming task ($\Delta G^2(1) = 4.08, p = .04$), but not for the detection-task ($\Delta G^2(1) = 2.13, p = .14$). Furthermore, when decomposing the response type (target or related) resulting from the retrieval of gist memory for targets, we observed that 8-year-olds identified targets as related distractors (18%) more often than as targets (1%). By contrast, 4-year-olds equally identified targets as either targets (13%) or related distractors (13%). In addition, older children had greater values of gist memory for related distractors than younger children ($\Delta G^2(2) = 14.85, p < .001$). The effect was present whether the color naming task ($\Delta G^2(1) = 9.24, p < .002$) or the detection task ($\Delta G^2(1) = 8.43, p < .004$) was used as the concurrent task. As in Experiment 1, gist retrieval led 8-year-olds to more often correctly identify related distractors as such (89%) rather than as targets (7%). By contrast, 4-year-olds identified related distractors as related (23%) as often as targets (21%) based on gist traces retrieval. Finally, the two guessing parameters (*a* and *b*) were significantly greater in younger children ($\Delta G^2(2) = 15.61, p < .001$; $\Delta G^2(2) = 86.47, p < .001$, respectively). Four-year-olds correctly identified targets on the basis of guessing (parameter *b*) in 2% of cases and they identified related distractors as such in 8% of cases and as targets also in 8% of cases. This never occurred in 8-year-olds for whom this parameter virtually did not contribute to these types of responses. Moreover, 4-year-olds were more likely than 8-year-olds to guess that an unrelated distractor was a target or a related probe (24% and <1% respectively).

Delayed test. Verbatim and gist memory for targets were lower in this test compared to the immediate test and they were not significantly affected by age ($\Delta G^2(2) = .68, p = .71$; $\Delta G^2(2) = 3.63, p = .16$, respectively). Gist memory for related distractors was significantly higher in 8- than 4-year-olds ($\Delta G^2(2) = 22.30, p < .001$). Gist retrieval led 8-year-olds to more often correctly identifying related probes as such (62%) rather than as targets (12%). By contrast 4-

year-olds were more likely to identify falsely related probes as targets (19%) than as related distractors (9%) based on gist retrieval. As in the immediate test, the two guessing parameters (a and b) were greater in the younger children group ($\Delta G^2(2) = 59.84, p < .001$; $\Delta G^2(2) = 22.74, p < .001$, respectively). Parameter b accounted for 9% of correct identification of targets, 10% of correct identification of related distractors and 22% of false identification of related distractors as targets in 4-year-olds. Parameter b contributed only marginally to the responses to target and related probes of 8-year-olds (at most 4% for the correct identification of related distractors). In addition, 4-year-olds were more likely than 8-year-olds to guess that an unrelated distractor was either a target or a related probe (22% and 11% of cases, respectively). All other effects were not significant.

5.3.3 Discussion

Results of Experiment 1 were replicated even though the task difficulty (i.e., memory load and concurrent task load) was adapted for each age group. In both the immediate and the delayed tests, the true recognition rate was higher in 8- than 4-year-olds, whereas the false recognition rate did not differ as a function of age group. In both age groups, memory accuracy was not affected by the manipulation of WM mechanisms availability. As in Experiment 1, the analysis of the contribution of verbatim and gist representations provided a good account of the recognition performance.

In the immediate test, verbatim and gist memory for targets were greater in 8- than 4-year-olds, particularly when the concurrent task was more demanding. This result accounts for the increase in true recognition with age. Older children also exhibited greater gist memory for related distractors than 4-year-olds. Here, gist memory retrieval led 8-year-olds to more often correctly judge that the item they were judging was a related distractor. By contrast, gist memory in 4-year-olds led them to make more false identification of related distractors as targets. These qualitative differences in the use of gist memories explain why, although they retrieved more gist memories, older children did not make more false memories than younger children. The same pattern of results for gist memory for related distractors was found in the delayed test. Congruently with Experiment 1, these results suggest that gist-based false memories can occur in children as young as 4 years old.

Finally, as in Experiment 1 and despite the adaptation of the task difficulty to the children's age group, recognition performance was not affected by the type of concurrent task. This could be due to the auditory presentation of words, which may have encouraged the use of rehearsal

and to the fact that uttering the names of colors did not sufficiently block the use of this mechanism. We conducted an additional experiment, the method and results of which are described in detail on the OSF, in which we intended to minimize the use of rehearsal and encourage the use of refreshing to maintain the information in WM. However, false recognition rate was at near floor in this additional experiment, rendering hazardous any conclusion. This could have resulted from the use of pictures; previous studies showed that visual supports weaken the DRM illusion in older children (e.g., Brainerd et al., 2018; see Brainerd & Reyna, 2012, for a review). These different points were addressed in the general discussion.

5.4 General Discussion

The present study examined how the age-related changes in verbatim and gist representations, and in the use of WM maintenance mechanisms affect children's correct and false recognition in a WM task. The availability of WM maintenance mechanisms did not affect recognition performance in any of the experiments, contrary to our expectations. However, we demonstrated two robust phenomena. First, 4- and 8-year-old children made false recognition of related distractors in both immediate and delayed recognition tests. Second, congruently with the FTT's predictions, false recognitions were underpinned by gist memory in both age groups. We will discuss in turn the false recognition effect obtained in 4- and 8-year-olds, the absence of significant effect of the manipulation of WM maintenance mechanisms in young children's recognition, and the role of gist memory in the formation of children's short- and long-term false memories.

False recognition effect in 4- and 8-years old children

Our study demonstrated that semantic distortions could occur in a WM task in 4- and 8-years old children. Previous studies have shown that semantically related distractors can be falsely recognized by young adults within seconds of the study phase (e.g., Abadie & Camos, 2019; Atkins & Reuter-Lorenz, 2008). To our knowledge, our study is the first to examine and evidence this phenomenon in children.

In addition, although false memories occur at short term in children, their incidence did not vary with age as in the classical DRM illusion. Younger children are as prone to false memory at short term as their older counterparts. These findings are both interesting and unexpected since developmental LTM studies using the DRM paradigm typically show an age-related increase in the occurrence of false memories, with younger children being less

susceptible to DRM false memories than older children (e.g., Anastasi & Rhodes, 2008; Brainerd et al., 2002c). At first, this might be taken as evidence that it is much easier for young children to connect and use semantic gist in WM tasks than in LTM tasks.

However, it is also possible that giving the theme of each list prior to the word presentation might have facilitated the retrieval of the gist of the lists. Studies indeed showed that gist-cuing before the presentation of DRM lists increased false recall and recognition of critical lures in children as young as 5 years old (see Brainerd et al., 2008a, for a review). In our study, this may have particularly benefited young children's ability to process the meaning of the items, favoring the creation of visual scenes based on the items. For example, the presentation of 'the farm' theme may prompt the image of a farm in which each element (pig, farmer, etc.) is easily integrated, as this type of visual scene is very common in 4-year-olds' books and cartoons and probably more so than for 8-year-olds. Furthermore, the presentation of a theme word may have a different impact on the processing of the common meaning (i.e., gist) from multiple items in young and older children. Indeed, older children are better at extracting the gist from multiple items (e.g., Brainerd et al., 2002c) and the occurrence of false memory in these children may not be influenced by the presentation of the theme word. By contrast, the theme word likely scaffolded the retrieval of common meaning in younger children, reducing the age-related difference in the incidence of false memories. This interpretation is partially supported by previous studies. For instance, Brainerd et al. (2004) have shown that telling 5- and 11- year-olds that the studied list exemplars would all be part of a specific taxonomic category before list-presentation (e.g., "all the words on the next list will be animals"; gist-cuing condition) increased false recognition in both age-groups. In the same vein, Holliday et al. (2008) showed that the increase in false recall between 7- and 15-year-olds was smaller in a gist-cuing condition compared to a control condition. Hence, these findings suggest that age-related increases in false memory may be attenuated by gist cuing in our study, but gist cuing might not be the only reason.

Other studies have indeed shown that the standard developmental trend in LTM tasks can be attenuated or even reversed when using modified versions of the DRM paradigm in which the DRM words are embedded in stories (e.g., Ghetti, et al., 2002). Story contexts enhanced 5-year-olds' vulnerability to the DRM illusion (e.g., Dewhurst et al., 2007; Howe & Wilkinson, 2011). Using scene-like visual stimuli that are associatively related with each other also resulted in higher false memory rates in 7- and 8-year-old children than in adults (Otgaar et al., 2014). The word lists used in the present study were specifically created for young

children, i.e., they contained themes that were well known and relevant to them. Carneiro et al. (2007) showed that when the studied lists were adapted to age groups by selecting the words rated as most associated to the critical lure according to the target group, false memories increased in the youngest group (3–4-years old) and the usual developmental increase trend was reduced. Hence, list of words designed for and by children could also be responsible for such a high rate of false memories in the youngest group. However, other studies have documented an increase in false memories with age using DRM lists created by children (e.g., Anastasi & Rhodes, 2008, Exp. 3; Metzger et al. 2008). It is worth noting that the latter studies used classical DRM themes, which was not the case in our study in which we used themes that are well known and more related to young children's daily life (such as the princesses, the farm, etc.) than DRM themes. In addition, the use of shorter list lengths in our WM task may have contributed to differences with the above-mentioned studies which implemented the long list lengths, typical of research of false memory in LTM. Further studies should aim at disentangling the impact of theme word presentation and of the type of lists used for generating false memories in WM tasks across different age groups.

The impact of WM maintenance on recognition performance

Previous findings have suggested that children from age 7 onwards can use rehearsal and/or refreshing to maintain information in WM, whereas younger children do not spontaneously use these mechanisms (e.g., Camos & Barrouillet, 2011; Henry, 2012; Oftinger & Camos, 2018; Tam et al., 2010). In the present study, as expected, 4-year-olds did not seem to engage in information maintenance strategies. Indeed, their responses to recognition tests were not impacted by the type of concurrent task in any of the experiments. Similarly, the 8-year-olds' responses were also not impacted by the type of concurrent task, which was rather unexpected.

There are two possible reasons for this. First, it is possible that our concurrent tasks did not properly suppress rehearsal of memory words or were not attentionally demanding enough to prevent refreshing. However, the concurrent tasks used in our experiments were similar to those implemented in previous studies in which they had affected the memory performance of children of the same age (Barrouillet et al., 2009; Camos & Barrouillet, 2011; Oftinger & Camos, 2018). Nevertheless, it may be noted that Langerock et al. (2020) showed that, in adults, the manipulation of concurrent attentional demand (i.e., the cognitive load effect) has a weaker effect (or even none) in Brown-Peterson tasks than in complex span tasks used in the previously

mentioned studies. Thus, one might expect that a stronger manipulation of cognitive load would be required to observe its effect in a Brown-Peterson task, especially in children. However, the princeps studies showing short-term false memories in adults implemented a Brown-Peterson task (Abadie & Camos, 2019; Atkins & Reuter-Lorenz, 2008), and therefore called for the use of this type of task in the first study in children, for the sake of comparisons.

Second, another explanation could be that 8-year-olds did not use WM maintenance mechanisms in our recognition task. As recently shown by Uittenhove et al. (2019) in adults, recognition test responses are supported by passive storage of memory items in LTM. This phenomenon is likely to have been even stronger in the present study among older children for whom the recognition test may have been perceived as particularly easy and effortless. Moreover, studies showed that compared to adults, children tend to rely more on reactive control, which involves passively engaging in a task (Chevalier, et al., 2014; Lucenet & Blaye, 2014; Munakata, et al., 2012). In contrast, proactive control is required to actively maintain information in WM and implement WM maintenance mechanisms (Braver 2012). It should be noted that previous experiments showing effects of WM maintenance mechanisms suppression in children's memory performance have used recall tasks and not recognition tasks (e.g., Barrouillet et al., 2009; Camos & Barrouillet, 2011; Oftinger & Camos, 2016; Tam et al., 2010). However, in our study, recognition was preferred over recall, because the use of a recognition task seemed more appropriate in 4-year-old children to avoid floor effect. Finally, to assess gist and verbatim, we implemented the SCR model, which can only apply to recognition tasks.

The role of gist memory in false memories in children

First of all, it is important to point out that, in all the experiments, the SCR model of the FTT fitted the data of both age groups very well, demonstrating that children in both age groups understood and responded well to our recognition tasks.

The FTT predicts that gist memory improves during childhood, because the ability to process the meaning of items and to understand the relationships between different items improves (e.g., Brainerd & Reyna, 2015). Improvements in gist memory increase the tendency to falsely accept semantically related lures as having been previously studied in LTM (e.g., Brainerd et al., 2002c). In the present study, both 4- and 8-year-olds seem to be able to make meaning connections among the items of our short DRM-like lists, which led them to falsely recognize related distractors in the immediate and delayed tests. Although there were no

significant differences in gist memory use between the two age groups in Experiment 1, we found an increase in gist memory with age in Experiment 2. In the latter experiment, memory lists were presented orally, which is the most common methodology used to study the DRM illusion. Studies showed that using visual supports, as it was the case in Experiment 1, weaken the illusion in older children (e.g., Brainerd et al., 2018; see Brainerd & Reyna, 2012, for a review). As shown in adults (e.g., Abadie et al., 2017), the format in which items are presented seems likely to modulate the tendency to form and retrieve gist traces in children.

However, these methodological differences cannot explain why gist memory increases with age, while older children do not make more false memories than younger children. A better explanation might lie in the qualitatively different use of gist memory depending on age. Combining the results of both experiments, we found that 8-year-olds used gist memory in most cases to correctly identify related distractors as such (i.e., an average of 96% of the time in immediate tests and of 84% of the time in the delayed test). By contrast, gist retrieval led 4-year-olds to identify related distractors either as such (on average, 58% of cases) or as target probes in an undifferentiated manner in the immediate tests and mostly as targets in the delayed test (on average, 70% of cases). Similarly, gist memory for target probes led the 8-year-olds to identify them correctly in the majority of cases (93% and 86% of cases in immediate and delayed tests, respectively), while it led the 4-year-olds to identify them almost as often as target probes (67% and 59%, respectively) as related distractors.

These findings suggest that, although young children were able to make meaningful connections between items of each list, which may explain the absence of age-related difference in false memory rates, there are still qualitative differences between 4- and 8-year-old children in the use of gist memory.

5.5 Conclusion

The present study aimed at examining how the development of LTM and WM can impact the emergence of false memory in 4- and 8-year-old children in WM tasks. Despite some evidence in adults (Abadie & Camos, 2019; Atkins & Reuter-Lorenz, 2008), this issue had never been examined in children. This study, in which short DRM-like lists specifically tailored for young children were presented in a Brown-Peterson task, found that younger children are as prone to false memories as their older counterparts in the immediate recognition test, and in

a delayed recognition test. In addition, using the state-of-the-art mathematical model of the FTT, we demonstrated for the first time in the literature that the retrieval of gist traces of the list themes is responsible for the emergence of short-term false memories in 4- and 8-year-olds. Gist memory also underpinned the occurrence of false recognition in the delayed test, replicating the classical results obtained in LTM with the DRM paradigm. These findings suggest that young children are as likely to make gist-based false memories as older children in WM tasks.

Summary

False memories are well established episodic memory phenomena. Recent research in young adults has shown that semantically related associates can be falsely remembered as studied items in WM tasks for lists of only a few items when a short 4-second interval was given between study and test. The present study reported two experiments yielding similar effects in 4- and 8-year-old children. Short lists of semantically related items specifically tailored for young children were retained over a brief interval. Whether or not the interval was filled with a concurrent task that impeded or not WM maintenance, younger children were as prone to falsely recognize related distractors as their older counterparts in an immediate recognition test, and also in a delayed test. In addition, using the conjoint recognition model of the fuzzy-trace theory, we demonstrated that the retrieval of gist traces of the list themes was responsible for the occurrence of short-term false memories in 4- and 8-year-old children. Gist memory also underpinned the occurrence of false recognition in the delayed test. These findings suggest that young children are as likely to make gist-based false memories as older children in WM tasks.

Résumé

Les faux souvenirs représentent un phénomène de mémoire épisodique bien établi. De récentes études chez de jeunes adultes ont montré que des items reliés sémantiquement à un petit nombre d'items réellement étudiés pouvaient être faussement reconnus comme étudiés dans des tâches de mémoire de travail, avec un délai de seulement 4 secondes entre la phase d'étude et de test. La présente étude rapporte deux expériences montrant des effets similaires chez des enfants de 4 et 8 ans. Des listes courtes d'items sémantiquement reliés, spécifiquement calibrées pour de jeunes enfants étaient retenues durant un intervalle de temps bref. Cet intervalle était rempli ou non par une tâche concurrente qui altérait ou non les mécanismes de maintien en mémoire de travail. Les jeunes enfants étaient autant enclins aux fausses reconnaissances que les enfants plus âgés, en test de reconnaissance immédiat et différé, peu importe la tâche remplissant l'intervalle entre la phase d'étude et la phase de reconnaissance. De plus, l'utilisation du modèle de reconnaissance conjointe simplifié de la théorie des traces floues a permis de montrer que la récupération de traces gist en mémoire était responsable de la survenue des faux souvenirs à court-terme chez les enfants de 4 et de 8 ans. Les traces gist sous-tendaient également les faux souvenirs dans les tests de reconnaissance différés. Ces résultats suggèrent que les jeunes enfants sont autant enclins aux faux-souvenirs basés sur des traces gist dans des tâches de mémoire de travail, que les enfants plus âgés.

Chapter 6. The development of false memory in working memory tasks: implication of maintenance mechanisms for 5- and 8-year-olds.

6.1 Introduction

The development of false memories during childhood has been widely investigated over the last half-century to shed light on societal issues, notably the credibility of witnesses when testifying during or about their childhood, but also to contribute to our general knowledge about human memory. Contrary to some intuitive beliefs, research has shown that young children are actually less prone to false memories than their older counterparts. The Deese-Roediger and McDermott paradigm (DRM, Deese, 1959a; Roediger & McDermott, 1995) is one of the most popular for studying the occurrence of false memories because of its simplicity and representativeness of false memories in everyday life. Participants are presented with lists of semantically related words (e.g., “climber, peak, ski, valley, etc.”), all of which are associates of a critical unrepresented word (“mountain”). On memory tests, that critical word is falsely recalled and recognized at reliable levels (i.e., well above chance level), resulting in a robust memory illusion. Extensive literature using this paradigm with children now exists, and several reviews of this literature have consistently shown that this semantic illusion increases dramatically with age (see Brainerd, 2013; Brainerd & Reyna, 2012; Brainerd et al., 2011, for reviews). Studies have generally shown that false recall increases about fivefold (i.e., from 10% to 53%) and false recognition increases about 1.2-fold (i.e., from 77% to 92%) from early childhood to young adulthood (Brainerd et al., 2002c). Two key conditions have been identified for this age-related increase in false memories to occur (Chang & Brainerd, 2021). First, false memories are based on semantic relations that connect critical unrepresented and studied items. This condition is particularly important because young children often lack the ability to form these connections spontaneously during encoding (e.g., Waroquier et al., 2020; Wimmer & Howe, 2009). Second, a high rate of false memories is typically obtained when many semantically related items are presented, and memory is tested after a significant delay (i.e., at least several minutes and often a day or more). However, a few more recent studies have challenged this second condition showing that the false memory illusion could also occur in WM tasks for only a few related items when a short 4-second time lapse was introduced

between study and test (e.g., Abadie & Camos, 2019; Atkins & Reuter-Lorenz, 2008; Coane et al., 2007; Flegal et al., 2010, 2014). Only one study examined the occurrence of false memories in WM tasks in children and found that children as young as 4 years old produced a relatively high rate of false recognition of critical unrepresented items (about 10%) in this type of task (Rousselle et al., 2023). These results raise questions about the origin of these errors, which have long been thought to be the hallmark of LTM. Could they also be formed in WM, a limited capacity system devoted to the temporary maintenance of representations relevant to the ongoing task (Barrouillet & Camos, 2022), or would they reflect an influence of LTM on responses to WM tasks? To address these questions, the present study examined the role of mechanisms dedicated to the maintenance of information in WM on the occurrence of false memories at different ages in childhood. If these mechanisms play a role in the formation of false memories, hindering them should thus affect their occurrence. In addition, as WM develops with age (e.g., Camos & Barrouillet, 2018; Cowan & Alloway, 2009, for reviews), children who actively maintain information should be more likely to produce false memories when maintenance mechanisms are hindered. Before presenting the current study, we present the Fuzzy trace theory (FTT, Brainerd & Reyna, 2002a; Brainerd et al., 2008a; Chang & Brainerd, 2021) as a theoretical framework for the development of false memories in LTM tasks. Then we review the findings regarding false memories in WM tasks. Finally, we underline the importance of assessing them both in young and older children.

The development of false memories in children.

Many studies have shown that false memories increase with age using DRM-like paradigms (see Brainerd & Reyna, 2012). In those studies, false recall and recognition become increasingly semantic between young childhood and early adulthood (e.g., Brainerd et al. 2002c; Dewhurst & Robinson, 2004). The dominant error type becomes semantic intrusions, and rivals the rate of correct recall and recognition. It indicates first that false memories are based on the processing of the meaning of items, and second that this type of processing seems to increase with age. The FTT supports this interpretation and states that false memories are based on the retrieval of gist traces, which are the representations of the semantic content of items and other relational information such as taxonomic, synonymous, or situational relations (Brainerd & Reyna, 2005). Retrieving such traces can indeed lead to the recall of information that shares the same semantic content (e.g., retrieving having seen something ‘from the farm’, that ‘gives milk’ can lead to recalling having seen a goat instead of a cow). According to the FTT, such traces develop with age, and would underpin the development of false memories

with age. Some studies provide evidence for this by showing that the increase in false memories with age was shown to occur on word lists that are semantically related, but not phonologically related (e.g., Khanna & Cortese, 2009; Dewhurst et al., 2012; Holliday & Weekes, 2006). Moreover, when favoring gist extraction for younger children by presenting lists adapted in meaning to age groups (e.g., Anastasi & Rhodes, 2008; Carneiro et al., 2007; Metzger et al., 2008) or by using visual stimuli (Ghetti et al., 2002; Howe, 2006; Howe et al., 2004; Khanna & Cortese, 2009) or ecological paradigms (e.g., Howe & Wilkinson, 2011; Lyons et al., 2010; Odegard et al., 2009), the increase in the rate of false memories was reduced.

In parallel with the development of gist memory, a second type of LTM trace, called verbatim memory, was also described by the FTT as improving with age (see Brainerd & Reyna, 2002a; Brainerd & Reyna, 2015, for reviews). Verbatim memory relates to representations of the surface features of an item, i.e., auditory, visual, or other details associated with the item's presentation (e.g., the colour, font, size, position, etc. of an item; Brainerd et al., 2019). Verbatim traces were shown to prevent false memories because they represent specific details of the perceived stimuli (e.g., retrieving having seen the exact shape of a cow with its exact colors allows us to rule out the goat, Brainerd et al., 2003). For instance, when 7- and 11-year-olds had to focus on processing the superficial characteristics of a stimuli, such as spelling, false memories were reduced in both age groups (Holliday et al., 2011; Odegard et al., 2008). As a parallel consequence of both gist and verbatim memory development with age, correct recall and recognition increase with age. False memories also increase in tasks in which verbatim is not strong enough to counter the gist memory effect on recall or recognition. This is the case in DRM and other tasks in which the stimuli to process are strongly semantically related and activate strong gist memory; in those tasks, recall and recognition are delayed and verbatim memory was shown to fade away faster than gist memory (Seamon et al., 2002) and to be less robust to interference (Abadie & Waroquier, 2020; Abadie et al., 2013; 2017). Verbatim memory is therefore not strong enough to counter the gist memory effect on false memories.

Additionally, some research took interest in the phenomenological experience associated with false memories during childrens' development, using a 3-point-confidence scale. Semantic intrusions were associated with lower confidence when compared to correct recall between 6 and 18 years (Ghetti et al., 2002; Ghetti & Angelini, 2008); 5-year-olds associated semantic errors with confidence rates similar to those of correct recall (Ghetti et al., 2002). However, one particularity of false memories is that they were reported with higher

confidence rates than other error types, both in older children (10-year-olds) and in young adults, but not in younger children (6-year-olds), whose confidence rate did not depend on the type of errors (Lyons et al., 2010). In some studies, an estimation of recollection (i.e., souvenirs recollected with their contextual details) and familiarity (i.e., memories of stimuli that are familiar but not recollected with their encoding context, Mandler, 1980) was collected among age groups. To do so, participants were asked to give a ‘remember judgment’ when their memory was retrieved with contextual details (i.e., recollection), or a ‘know’ judgment when it was not (i.e., familiarity). They also had the option to use a ‘guess’ judgment when they gave their response based on guessing. Brainerd et al. (2004) showed that recollection of semantic errors increased between 7- and 11-year-olds while familiarity reports remained stable, suggesting an increase of strong false memories with age. In line with these results, another study showed that 10-year-olds had more recollection than 6- to 8-year-olds, whereas familiarity remained stable across age groups.

Short term false memory.

A few studies have shown that false memories in young adults could occur within the scope of a few seconds (4 seconds) after the study of a very small number of stimuli (4 items) (Atkins et al., 2011; Atkins & Reuter-Lorenz, 2008; Abadie & Camos, 2019; Flegal et al., 2010, 2014). When the phenomenological experience of these short-term false memories was assessed, semantically related distractors were more often associated to ‘remember’ judgments when compared to semantically unrelated distractors. More precisely, in one third of cases, semantic associates were associated with remember judgements, in one third of cases to ‘know’ judgments and the last third was associated to guessing (Flegal et al., 2010, 2014). Interestingly, similar findings were observed in a delayed recognition test, which suggests consistency between the subjective phenomenological experience of false memories at short and longer delays. In another study, subjective judgments were collected during a recall test (Abadie et al., in rev., Exp. 4). Results show that semantic errors were associated mainly with the judgment ‘I heard this word, or a word related in meaning to this word’. This judgment refers to a fuzzy meaning-based trace and contrasts the other two judgments of the scale in this study, which were either ‘I heard this word’, referring to precise strong memory, or ‘I don’t know if I heard this word’, referring to guessing.

Very recently, the question of whether false memories could occur in WM tasks in children was first tested in a study using a Brown-Peterson task, in which lists of words adapted

to 4 year old children and 8 year old children were presented, followed by a 6-8 second retention interval (Rousselle et al., 2023). They used the simplified conjoint recognition model (SCR, Sthal & Klauer, 2008) of the FTT, a multinomial processing tree that estimates the contribution of gist, verbatim and guessing to true and false recognition. Results showed that gist-based false recognitions occurred at both ages. Therefore, it seems that false memories can occur in WM tasks in children as young as 4 years of age. This evidence is groundbreaking considering that young children were sometimes found to produce very low rates of false memories in LTM tasks (e.g., Anastasi & Rhodes, 2008; Brainerd et al., 2018); as false memories were mainly shown to occur in LTM tasks, these findings question the role of WM maintenance in false memories.

The role of working memory maintenance mechanisms' development

WM is a complex system organized in different mechanisms, which are dedicated to maintaining and processing a limited amount of information for a short period of time (a few seconds). Two main WM maintenance mechanisms have been described by the prominent Time-based resource sharing model (TBRs, Barrouillet & Camos, 2022) as developing with age and impacting performance in WM tasks. The first mechanism is called 'articulatory rehearsal' and consists of a recirculation of memorial traces to conserve its phonological form (Barrouillet & Camos, 2021). In the literature, there is substantial evidence that articulatory rehearsal starts to be efficient between the ages of 5 and 7, such as the emergence of the phonological similarity effect at those ages (Henry et al., 2012), the emergence of overt speech (Elliott et al. 2021; Flavell et al., 1966), and the drop of recall performance in WM tasks when introducing a concurrent articulation (Oftinger & Camos, 2016, 2018; Rosselet-Jordan et al., 2022 ; Tam et al., 2010). The second WM maintenance mechanism is 'attentional refreshing' and consists of refocusing attentional resources on temporarily stored information in WM, to restore and preserve the decline with time, with a sufficient activation level to avoid forgetting (Barrouillet & Camos, 2021). Attentional refreshing starts to be used efficiently from age 7, and keeps on increasing in efficiency until age 14 (Gaillard et al., 2011). This was mainly evidenced by studies showing that varying the attentional demand of a concurrent task impacted children's performance from 7 years of age (e.g., Barrouillet et al., 2009; Camos & Barrouillet, 2011; Oftinger & Camos, 2016, 2018). Those studies have shown that maintenance mechanisms promoted correct immediate recall in WM tasks in children as soon as they started using them. However, no studies have yet assessed the effect of maintenance mechanisms on children's false recall.

A few studies assessed the role of articulatory rehearsal on false memories in adults (Abadie & Camos, 2019; Atkins et al., 2011; Macé & Caza, 2011), using a WM task in which four semantic related words were presented. They showed that allowing rehearsal, rather than preventing its use by introducing a concurrent articulation during retention delays of a few seconds, prevented the occurrence of false memories. Regarding refreshing, the study that assessed its role in false memories revealed that refreshing did not impact short-term false memories but tended to increase long-term false recognitions (Abadie & Camos, 2019). Indeed, when refreshing was hindered by a high attentionally demanding concurrent task, the rate of false memory in a delayed test was decreased. Abadie & Camos' study went further and showed that maintenance mechanisms impacted traces in memory by using the SCR model. They showed that rehearsal preserved verbatim memory for immediate recognition, which explained the decrease of false memories in this test. Conversely, refreshing promoted gist memory traces, which favored the occurrence of delayed false memories.

WM maintenance mechanisms represent a pillar of many daily activities (e.g., taking notes during class, doing mental calculations, having a conversation, etc.). It was shown to be a predictor of both math and reading performance in school (Peng et al. 2016; Peng et al. 2018, respectively; Swanson & Alloway, 2012). Given the central role of WM in cognitive activities and the evidence that its maintenance mechanisms moderate the production false memories in adults (e.g., Abadie & Camos, 2019; Atkins et al., 2011; Macé & Caza, 2011), there is real interest in understanding how its development impacts the occurrence of false memories in children. Yet the literature on the development of WM in children is limited compared to the work conducted with adult populations. The existing studies in children usually focus on true memory without looking at errors. There was indeed to our knowledge only one study conducted on false memories in children in WM tasks (Rousselle et al., 2023). Looking at the role of WM maintenance mechanisms in false memory in those tasks could be quite informative, especially for choosing the most appropriate teaching methods for targeted age groups, but also for our general understanding of memory development with age.

The present study

In the present study, two experiments were conducted in which 5- and 8-year-olds completed a complex span task, in which DRM-like short lists of words containing 3 or 4 words respectively for the younger and older group, were presented. In between each word, a concurrent task lasting a few seconds was introduced. This task was either highly attentionally

demanding and required a concurrent articulation, or was less attentionally demanding and did not ask for a concurrent articulation. This manipulation aimed to affect both WM maintenance mechanisms to manipulate the ability to maintain the words to be learned. As in the classic DRM task, an immediate recall test was presented at the end of each list presentation and a delayed recognition test was introduced at the end of each of the two blocks of lists. During recall, participants were asked to specify their phenomenological experience of each recalled word among 3 subjective judgements, which were similar to that used by Abadie et al. (in rev.).

First, this study aims to test the effect of WM maintenance mechanisms on false memories in children in which those mechanisms were barely used compared to children who already used them. We expected that 8-year-olds would be more impacted than 5-year-olds by the manipulation of WM maintenance mechanisms through a concurrent task. Hence, immediate semantic errors in recall should increase, and correct recall should decrease in older children when maintenance mechanisms cannot be used during retention intervals. Based on Abadie & Camos' study (2019) which showed that delayed false memories increased when refreshing could be used, we expected that 8-year-olds would produce more false recognitions and fewer correct recognitions when WM maintenance mechanisms could be used compared to when they were prevented by the concurrent task. This effect was not expected in younger children who usually have fewer gist memory traces and do not seem to use refreshing.

Second, our study seeks to test the FTT in WM tasks by looking at the evolution of false memories between 5- and 8-year-olds in such tasks. We expected that, as in classic DRM tasks, semantic errors and correct recall would increase with age in immediate recall. In contrast, based on Rousselle et al.'s (2023) findings, we expected higher rates of true delayed recognition in older children but similar rates of false delayed recognition between 5- and 8-year-olds when both WM maintenance mechanisms were impaired.

Third, based on the idea that older children have more gist memory, we expected that they would more often associate their correct recall and semantic errors to a judgment reflecting the retrieval of a fuzzy meaning-based trace (i.e., 'I heard this word or a resembling one') than younger children. Moreover, because verbatim also develops with age, we expected older children to have more precise and strong memory judgments (i.e., 'I heard this word') on correct recall and semantic errors than younger children. In contrast, younger children should produce more guessing judgments (i.e., 'I don't know if I heard this word'). Additionally, we expected that the suppression of WM maintenance mechanisms would decrease precise strong 'I heard

this word' judgements and should increase guessing judgments on correct recall and semantic errors in older children, who use such mechanisms.

Finally, in delayed recognition, as in Abadie & Camos' study (2019), the SCR model of the FTT was used to assess memory traces on which true and false recognition were based. We expected that older children would have more verbatim and gist memory.

6.2 Experiment 1

In Experiment 1, 5- and 8-year-old children performed a complex span task in which a theme was presented, and then they had to retain lists of thematically related words, adapted for children of their age (Rousselle et al., 2023). Three words per list were presented to younger children and 4 words to older children. Studies have shown that children's memory span was of four words at age 5 and just over four at age 8 (Dempster et al., 1981, for a review). Five-year-olds were found to be able to repeat 3 words back, and 9-year-olds retained 5 words (Hitch, et al., 1989). A concurrent task was introduced between each presented word list to impede the possibility of using WM maintenance mechanisms. In the highly demanding group, children were asked to name colors out loud, a task aimed at impairing both rehearsal and refreshing. Indeed, the task required concurrent articulation, and was attentionally demanding enough to prevent the use of refreshing; we picked this task because it is a task that can be performed by both age groups. In the less demanding condition, we used a detection task aimed at preserving both mechanisms, as it did not ask for specific concurrent articulation nor draw much attentional resources. An immediate recall test followed each list presentation. During this task, children were asked to choose their subjective experience for each recalled word between three options. There has been evidence that young children (3- to 5-year-olds) can use similar three-options scales (i.e., Coughlin et al., 2015; Ghetti & Alexander, 2004; Ghetti et al., 2002; Hembacher & Ghetti, 2014). At the end of each block, which included half of the lists to study, a delayed recognition test was introduced after a 2-minute distraction task. For recognition, we used the SCR model, an objective measure of gist and verbatim memory.

6.2.1 Method

Material and data for the two experiments can be found at: <https://osf.io/c6m3x/>.

Participants

Forty-six 1st graders (the 5-year-old group; 23 girls; *Mage* = 67.26 months; *SD* = 3.68 months) and 43 3rd graders (the 8-year-old group; 26 girls; *Mage* = 102.77 months; *SD* = 3.24 months) were recruited from several preschools and elementary schools in France to participate in the study. Two subjects from the 5-year-old group were excluded from the data analysis because they did not follow the instructions. They were all French native speakers. For each participant, written parental consent was obtained. Ethical approval was obtained from the institutional review board of Aix-Marseille University.

Material

Word lists. We selected 12 lists of words from those used in Rousselle et al., (2023). There were two versions of each word list which included words that are semantically related (e.g., “chicken, goat, pig, tractor” or “chicken, goat, pig, cow”) to a theme (e.g., “the farm”). They were recorded by a female voice via Audacity software (2016). We ensured that the words in each list did not share semantic similarities with words in other lists and that the length of the selected words did not vary by more than one syllable. Each list consisted of three semantically related words for 5-year-olds and four semantically related words for 8-year-olds. Two additional lists were added for the training phase.

The 12 lists were separated into two groups of six lists. Each group was randomly assigned to each of the two experimental conditions (i.e., with or without suppression of WM maintenance mechanisms). The order of presentation of the lists within a given group was also randomized.

Subjective experience. During the recall phase, children were asked to choose a subjective judgment among three options for each recalled word. When they recalled a word at random, they were to qualify that recall with the judgment- ‘I don’t know if I heard this word’. If they thought that the recalled word was one of the studied words, they were to give it the judgment - ‘I know I heard this word’. Finally, if they thought the recalled word might be either a studied word or a word that was similar to a studied word, they were to give the recall judgement - ‘I heard this word or a similar one’. To ensure that children understood the three options, each statement was accompanied by a visual representation. Option 3 was accompanied by an image of an ear, option 2 by an image of a fuzzy ear and option 1 by a drawing of an indecisive man. These images were found freely available on the internet.

Delayed recognition test. A list of 18 probe words was constructed including six targets (i.e., randomly selected words from lists of the study phase), six related distractors (i.e., unpresented words semantically related to a studied word list), and six unrelated distractors (i.e., unpresented words that were not related to any of the word list). Each of the 12 studied word lists was associated with either a target or a related probe. There were two variants of a same probe type for a given list (e.g., for the list “chicken, goat, cow, tractor” the target probe could be ‘chicken’ for one participant and ‘goat’ for another). Half of each probe type were used in each of the experimental condition.

Procedure

The experiment was displayed on a computer using E-Prime 3.0 software (version 3.0.3.82, 2018). It took place in a quiet room in each school. Figure 11 illustrates the entire procedure. Children were asked to memorize the 12 lists of words for a subsequent recall test. In each trial, children first heard the name of the word list category (e.g., “the farm”). Then, each word of the list was presented sequentially orally for 1750 milliseconds for 8-year-olds and 2000 milliseconds for 5-year-olds, interspersed with a 5300-millisecond concurrent task. This latter task began with a 500-millisecond period during which a ‘beep’ signaled the start of the concurrent task. Next, for half of the 12 trials (i.e., WM maintenance mechanisms suppression condition), children performed a color naming task. They had to identify the color of three smileys displayed for 1100 milliseconds for 8-year-olds or two smileys displayed for 1900 milliseconds for 5-year-olds. The ISI between two smileys was of 500 milliseconds for both age groups. Children had to name aloud the color of each smiley (i.e., green, blue, or red) while pressing a corresponding-colored key. For the other half of the trials (i.e., without WM maintenance mechanisms suppression condition), children performed a detection task. Circles appeared sequentially on the screen and children were asked to press the spacebar for each one of them. The number and presentation time of each circle was similar to the number and presentation time of each smiley for each age group.

At the end of each trial, children were asked to recall the studied words. The experimenter recorded their answers in the exact order they were given. If the children did not recall all the words, the experimenter told them how many words were missing. If they did not fully complete their recall following this comment, they were encouraged to give the words they thought they heard. Finally, if their recall was still not complete after this time of reflection, they were invited to give any word they could think of. The experimenter then repeated their

responses and asked them to give a subjective judgment score to each word using the 3-point scale. The trials were presented blocked by condition. The order of presentation of each block was randomized for each participant.

At the end of each block, children did a distraction task. They were asked to copy Rey’s Figure (1941) for 2 minutes. Then, they performed a surprise delayed recognition test in which the different types of probes were presented auditorily in a sequential and random manner. For each of them, the children were asked if they had heard it among the six studied lists. A “yes” response to this question was considered and recorded as a “target” response. If they answered ‘no’, they were asked if the probe could be related to a thematic category being studied. A “yes” response to this question was recorded as a “related distractor” response and a “no” response was recorded as an “unrelated distractor” response.

The whole procedure lasted 35 minutes on average. The experimental session was preceded by a training phase. At the beginning of each block, children were trained to the corresponding concurrent task alone first without, and then with a time limit. They were then trained to complete the complex span task. At the end of the experiment, children were thanked for their participation and received a medal.

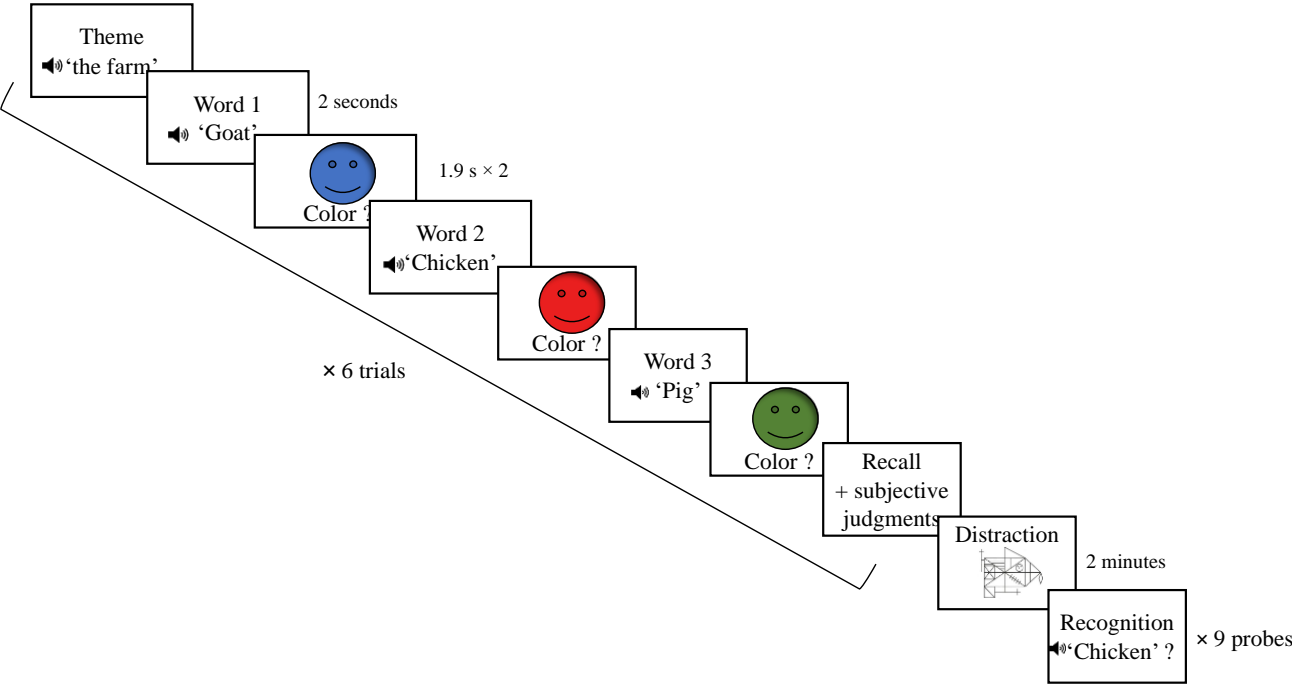


Figure 11. Illustration of one trial of the complex span task in Experiment 1, for the 5-year-old group. For the other half of the trials, smileys were replaced by black circles (detection task).

In the 8-year-old group, four words were presented for 1.75 seconds, and three smileys were presented for 1.1 seconds.

6.2.2 Results

Bayesian analyses were conducted using JASP Version 0.16.4 (Jasp Team, 2022) with the default mode settings. The Bayes factor (BF_{10}) of each possible model was obtained by comparing its likelihood to the likelihood of the null model. First, we reported the best model, the model with the largest BF_{10} . Then, we reported the $BF_{inclusion}$ value for each effect included in the best model (e.g., a main effect or an interaction effect), which indicates the likelihood of each model including a given effect compared to all models stripped of that effect. The strength of the Bayes Factors was interpreted based on Kass and Raftery (1995) norms.

Concurrent task performance

Mean performance in the two concurrent tasks were above 80% accuracy in each age group (89.6%, $SD = 13.7$; 91.7.1%, $SD = 11.4$, for 5- and 8-year-olds in the detection task, respectively, and 85.5%, $SD = 15.8$; 85.2%, $SD = 13.1$, for 5- and 8-year-olds in the color naming task, respectively), showing that children followed the instructions adequately.

Immediate recall accuracy

Two independent and trained raters classified the responses to the recall test. They identified six different types of responses. The first type was correct recall. We considered a response as correct when the recalled word was similar to a studied word, no matter whether the position of the word was correct or not. All other responses were errors. The first type was semantic errors, i.e., words semantically related to the theme of the studied list. The second type was the recall of the theme of the list, which was given before the presentation of each list (note that children were asked not to recall it). The third was phonological errors, i.e., words phonologically related to one of the words of the studied list. The fourth was intrusions by words from previously studied lists or recall errors from previous lists. The fifth and last type was other errors that had no obvious link with the studied words. Interrater agreement was 94.4% before discussion among raters and full interrater agreement was obtained after discussion. Table 4 shows the percentages of each type of response as well as the decomposition of subjective judgments for correct recall, semantic and aggregated non-semantic errors by age group and type of concurrent task.

Table 4.

Mean percentage of correct recall and recall errors (semantic errors, theme recall, phonological errors, intrusions and other errors) and decomposition of subjective judgments (H, HR and DK^a) for correct recall, semantic and non-semantic errors as a function of age group and type of concurrent task.

Age Group	Concurrent task	Correct Recall	Semantic errors	Theme recall	Phonological errors	Intrusion errors	Other errors
		H/HR/DK	H/HR/DK		H/HR/DK		
5	Detection	79.7	7.2	5.8	0.8	3.5	3
		(17.1)	(7.4)	(7.9)	(1.9)	(7.3)	(5.7)
	Color naming	75.4/ 2.3/ 2	4.2/ 0.5/ 2.5		8.6/ 0 / 4.5		
		(21.5/ 7/ 7.7)	(6.6/ 2/ 3.9)		(12.6/ 0/ 8.9)		
8	Detection	90.8	7.4	0.7	0.4	0.6	0.2
		(9.2)	(7.9)	(2.4)	(1.5)	(1.5)	(0.9)
	Color naming	89.1/ 0.9/ 0.8	2/ 2.6/ 2.7		1.5/ 9.7 x 10 ⁻⁴ / .3		
		(9.6/ 1.9/ 1.9)	(4.8/ 4.2/ 3.7)		(2.9/ .6/ 1.1)		
Color naming	79.4	15.8	3.1	0.4	1	0.4	
	(12.2)	(9.2)	(5.5)	(1.5)	(3)	(1.5)	
		74/ 4/ 1.4	3.8/ 5.5/ 6.5		3.1/ 1.4/ .4		
		(12.4/ 5.8/ 2.5)	(5/ 7/ 7.3)		(4.8/ 3/ 2)		

Note. Standard deviations are in brackets. ^a H = ‘I know I heard this word’, HR = ‘I heard this word or a resembling one’, DK = ‘I don’t know if I heard this word’.

Correct recall. A mixed-measures Bayesian ANOVA with age group (5 vs. 8-year-olds), concurrent task type (detection vs. color naming) and subjective judgment type (H, HR, DK) was conducted on correct recall. The model that included the main effects of age, concurrent task type and subjective judgment as well as the interactions between subjective judgment and concurrent task and between subjective judgment and age, was the best ($BF_{10} = 2.91 \times 10^{135}$). As predicted, there was decisive evidence that correct recall was higher in 8- than in 5-year-olds ($BF_{inclusion} = 563.54$) and that correct recall was higher in the detection task than in the color naming task ($BF_{inclusion} = 8.85 \times 10^4$). There was also decisive evidence for

differences in subjective judgments ($BF_{\text{inclusion}} = 6.58 \times 10^{100}$). Post-hoc comparisons revealed that there was more correct recall associated with judgment “I heard this word” (i.e. ‘heard’), than with judgments “I heard this word or a resembling one” (i.e. ‘heard or related’), or “I don’t know if I heard this word” (i.e. ‘don’t know’) ($BF_{10} = 1.74 \times 10^{95}$; $BF_{10} = 1.56 \times 10^{92}$, respectively). The interaction between subjective judgment and age ($BF_{\text{inclusion}} = 587.02$) was decomposed using independent t-tests with age as a between-subject factor on each judgment separately. The analyses revealed that 8-year-olds made more ‘I heard this word’ judgments than 5-year-olds ($BF_{10} = 97.74$), whereas there were no age differences for judgments ‘heard or related’ and ‘don’t know’ ($BF_{10} = .27$; $BF_{10} = .29$, respectively). The interaction between subjective judgment and concurrent task type ($BF_{\text{inclusion}} = 2.96 \times 10^{25}$) was decomposed using paired sample t-tests with concurrent tasks as a within-subject factor on each judgement separately. The analyses showed that there were more ‘I heard this word’ judgments in the detection than the color naming task ($BF_{10} = 4.03 \times 10^{12}$), whereas there were no differences for ‘heard or related’ and ‘don’t know’ judgments ($BF_{10} = .47$; $BF_{10} = .13$, respectively).

Recall errors.

An initial Bayesian repeated-measures ANOVA compared the rate of each type of recall error (semantic, theme recall, phonological, intrusions and other errors). The analysis provided decisive evidence for different error rates ($BF_{10} = 1.24 \times 10^{32}$). As expected, post-hoc comparisons indicated that participants made more semantic errors (10.5 %) than any other error type (theme: 4.6%, $BF_{10} = 9.88 \times 10^4$; phonological: 0.5%, $BF_{10} = 6.4 \times 10^{21}$; intrusion: 2.7%, $BF_{10} = 2.07 \times 10^{12}$; other: 2.9%, $BF_{10} = 2.95 \times 10^9$). They also more frequently recalled the theme of the list than they made phonological or intrusion errors ($BF_{10} = 2.54 \times 10^5$; $BF_{10} = 3.06$; respectively). They made as many intrusions as other errors ($BF_{10} = .12$). Those two error types were more frequent than phonological errors ($BF_{10} = 222.3$; $BF_{10} = 88.4$, respectively).

As our predictions were mainly about the comparison between semantic and non-semantic errors, we aggregated non semantic errors (i.e., phonological, other errors, theme recall, intrusions) and conducted a Bayesian mixed measures ANOVA with concurrent task, age group and type of errors (semantic vs. non semantic errors). The best model included the main effects of age group, concurrent task and error type and the interaction between age group and error type ($BF_{10} = 10.41 \times 10^{16}$). The second-best model also included the three-way interaction between age, concurrent task and error type ($BF_{10} = 8.62 \times 10^{16}$). Five-year-olds

produced more errors than 8-year-olds ($BF_{\text{inclusion}} = 676.57$) and there were more errors in the color naming than in the detection condition ($BF_{\text{inclusion}} = 8 \times 10^{10}$). There was substantial evidence against a main effect of error type ($BF_{\text{inclusion}} = .22$). Analyses of effects provided substantial evidence for the three-way interaction between age, concurrent task and error type ($BF_{10} = 29$).

Thus, further mixed-measures ANOVAs with age and error type were conducted separately on each concurrent task. In the detection-task condition, there was decisive evidence for an interaction between age and error type in the detection task ($BF_{\text{inclusion}} = 809.77$). Paired sample t-tests with error type as a within-subject factor were conducted separately for each age group to decompose the interaction. There was weak evidence that 5-year-olds produced more non-semantic (13.1%, $SD = 15.2$) than semantic errors (7.2%, $SD = 7.4$; $BF_{10} = 1.94$) and decisive evidence that 8-year-olds produced more semantic (7.4%, $SD = 7.9$) than non-semantic errors (1.8%, $SD = 3.2$; $BF_{10} = 777.58$). In the color naming task, there was decisive evidence for an interaction between age and error type ($BF_{\text{inclusion}} = 2.54 \times 10^8$). Paired-sample t-tests showed that there was strong evidence that 5-year-olds produced more non-semantic (22.7%, $SD = 17.9$) than semantic errors (11.6%, $SD = 8.7$; $BF_{10} = 23$). Conversely, there was decisive evidence that 8-year-olds produced more semantic (15.8%, $SD = 9.2$) than non-semantic errors (4.8%, $SD = 6.7$; $BF_{10} = 9.58 \times 10^4$).

Finally, we conducted two mixed-measures Bayesian ANOVAs separately for semantic and non-semantic errors by adding the subjective judgment factor to the previous analyses. Concerning semantic errors, the best model included the main effects of concurrent task, subjective judgment and age, and the interaction between age and subjective judgment ($BF_{10} = 1.16 \times 10^5$). There was weak evidence against a main effect of subjective scale ($BF_{\text{inclusion}} = .73$, respectively). To decompose the interaction between age and subjective judgment ($BF_{\text{inclusion}} = 35.32$), independent sample t-tests were conducted separately for each judgment. Analyses revealed that 8-year-olds more often associated the judgement 'I heard this word or a resembling one' with a semantic error than did 5-year-olds ($BF_{10} = 127.24$). There was no difference by age group for judgments 'don't know' and 'heard' ($BF_{10} = .47$; $BF_{10} = 1.52$, respectively).

For non-semantic errors, the best model included the main effects of concurrent task, age and subjective judgment and an interaction between age and subjective judgment and between concurrent task and age ($BF_{\text{inclusion}} = 2.38 \times 10^{15}$). Post hoc comparisons of the main effect of

subjective judgment ($BF_{\text{inclusion}} = 1.40 \times 10^4$) revealed that judgment ‘heard’ was used more often than judgment ‘don’t know’ ($BF_{10} = 9.08$) and that judgment ‘don’t know’ was used more often than judgment ‘heard or related’ ($BF_{10} = 60.36$). There was decisive evidence for an interaction between age and subjective judgment ($BF_{\text{inclusion}} = 106.39$). Independent sample t-tests conducted separately on each subjective judgment type revealed that 5-year-olds used judgment ‘heard’ and ‘don’t know’ more often than 8-year-olds ($BF_{10} = 934.31$; $BF_{10} = 90.23$, respectively), while judgment ‘heard or related’ was used equally across age groups ($BF_{10} = .23$).

Delayed recognition accuracy

We computed discriminability indexes (Pr) for correct recognition of target probes and for false recognition of related probes to eliminate potential differences between age groups that would be related to the ‘yes’ response bias (e.g., Brainerd et al., 2002c; Otgaar et al., 2014). Pr indexes were obtained by subtracting the baseline false recognition rate of unrelated distractors (i.e., responses “target” to unrelated distractors) from the rate of correct recognition of targets (correct recognition) and from the rate of false recognition of related distractors as targets (false recognition)⁷. Pr indexes as a function of age group and concurrent task are shown in Figure 12. Two separate Bayesian mixed measure ANOVAs were conducted on correct and false recognition with age and concurrent task type. Concerning correct recognition, the additive model including the main effects of age group and concurrent task type was the best ($BF_{10} = 34.98$). Older children produced more correct recognition than the younger ones ($BF_{\text{inclusion}} = 9.95$), and there was more correct recognition in the detection than in the color naming condition ($BF_{\text{inclusion}} = 3.61$). For false recognition, the null model was the best ($BF_{10} = 1.00$).

⁷ You can find the individual data of all response type (target related, unrelated) to each probe type (target, related, unrelated) on the OSF. Analysis on correct identification of target probes and false recognition of related probes are consistent with than of their discriminability indexes in both experiments.

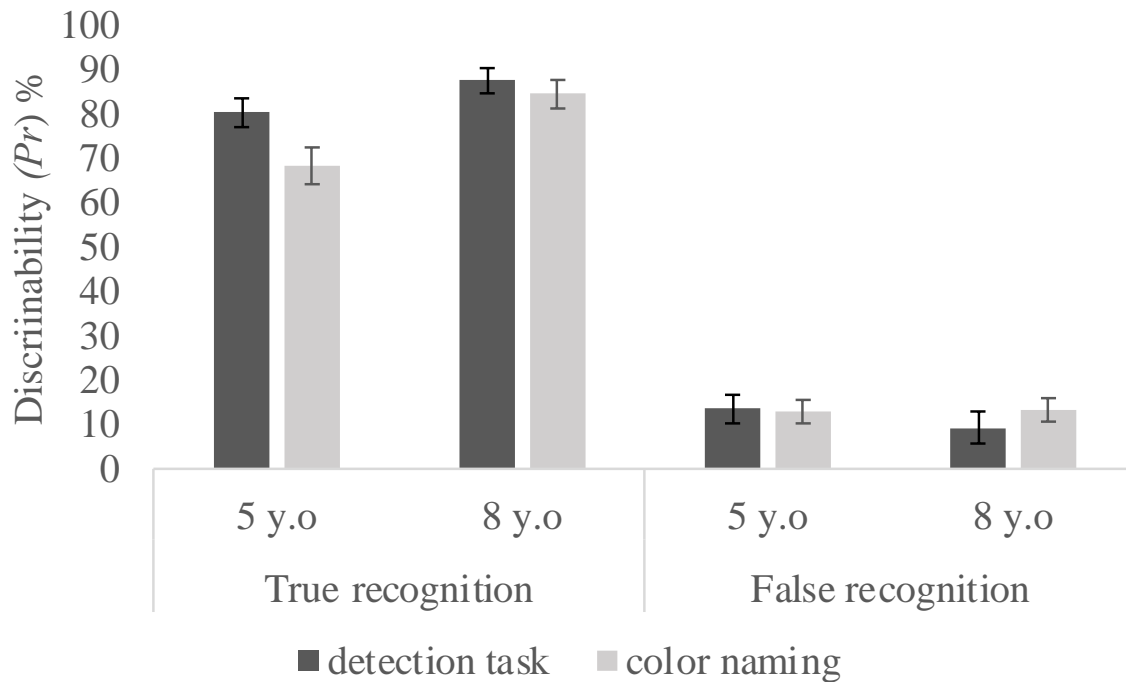


Figure 12. Correct and false recognition as a function of age group and type of concurrent task. Error bars represent standard errors.

Verbatim, gist memory and guessing processes. The multinomial model of the simplified conjoint recognition model (Stahl & Klauer, 2008) was used to compute parameter estimates for verbatim and gist memory for targets (V_t , G_t , respectively) and for related probes (V_r , G_r , respectively) and guessing parameters (b and a). The parameters V_t and G_t represent the probability of retrieving a verbatim (V_t) or a gist trace (G_t) of a target when a target probe is presented at test. The parameters V_r and G_r correspond to the probability of retrieving a verbatim (V_r) or a gist trace (G_r) when a related probe is presented at test. According to the model, the retrieval of a verbatim trace of targets leads to correct identification of target and related probes. When verbatim retrieval fails, however, participants can retrieve gist memory. They are familiar with the meaning of the target or related probe, but they do not remember whether the probe itself or a related one with the same gist was presented at study. They are then assumed to guess whether the probe is a target (with probability a) or a related distractor (with probability $1 - a$). When neither verbatim nor gist memory is available, a participant can still guess that the probe meaning is old with the probability b . The decision between the target and related responses is again modelled by the parameter a . Otherwise, the participant guesses that the probe is new with the probability $1 - b$. Verbatim and gist traces do not intervene in the responses to unrelated distractors because these probes do not trigger the retrieval of memory representations of the

study phase. Therefore, the responses to unrelated distractors are based on a combination of guessing processes a and b (see Stahl & Klauer, 2008 for a representation of the model).

Memory and guessing parameter estimates are given in Table 5. Hierarchical analyses were conducted using the latent-trait approach (Klauer, 2010) implemented in the TreeBUGS software (Heck et al., 2018). We adapted the analysis script for the Multinomial Processing Tree (MPT) developed by Greene and Naveh-Benjamin (2020).

Model fit was assessed with T_1 which is the distance between the observed and the expected mean frequencies, and T_2 which is the summed distance between the observed and the expected covariance statistics. The data fit the model well, as indicated by nonsignificant test results in 5-year-olds ($T_1: p = .24, T_2: p = .37; T_1 p = .10, T_2: p = .49$; for the detection and color naming task, respectively) and in 8-year-olds ($T_1: p = .31, T_2 p = .55; T_1 p = .36, T_2: p = .70$; for the detection and color naming task, respectively).

Table 5.

Mean parameter estimates of the SCR model as a function of age and concurrent task type

	5-year-olds		8-year-olds	
	Detection task	Color naming task	Detection task	Color naming task
V_t	.77 (.04)	.65 (.08)	.85 (.04)	.82 (.05)
G_t	.49 (.12)	.46 (.28)	.46 (.18)	.59 (.12)
V_r	.13 (.13)	.12 (.07)	.31 (.23)	.19 (.05)
G_r	.37 (.26)	.35 (.20)	.44 (.06)	.64 (.05)
b	.24 (.14)	.22 (.21)	.16 (.12)	.18 (.05)
a	.41 (.18)	.47 (.36)	.25 (.09)	.24 (.01)

Note. Standard deviations are in brackets.

We conducted a Bayesian mixed-measures ANOVA with age group as a between-subject factor and concurrent task type as a within-subject factor on individual estimates of each parameter.

Concerning verbatim memory for targets (V_t), the full model was the best ($BF_{10} = 3.58 \times 10^{37}$). Verbatim memory for targets was higher for 8-year-olds than 5-year-olds ($BF_{inclusion} = 2.20 \times 10^{19}$) and when the detection task rather than the color naming task was used as the concurrent task ($BF_{inclusion} = 3.52 \times 10^{13}$). There was also decisive evidence for the interaction between age and concurrent task type ($BF_{inclusion} = 2.73 \times 10^8$). To decompose the interaction, we conducted paired sample t-tests separately on each age group with concurrent task as a within-subject factor. The concurrent task effect was larger in 5-year-olds ($BF_{10} = 1.15 \times 10^{12}$) than 8-year-olds ($BF_{10} = 44.76$).

The full model was also the best for gist memory for targets (G_t ; $BF_{10} = 7.1$). However, there was weak evidence against the main effects of age and concurrent task type ($BF_{inclusion} = .57$; $BF_{inclusion} = .56$). Importantly, there was strong evidence for the interaction between age group and concurrent task type ($BF_{inclusion} = 22.2$). Follow-up paired samples t-tests revealed that gist memory for targets was higher in 8-year-olds when the color naming rather than the detection task was used as a concurrent task ($BF_{10} = 182.48$) but it was not the case in 5-year-olds ($BF_{10} = .22$).

Concerning verbatim memory for related distractors (V_r), the full model was the best ($BF_{10} = 2.14 \times 10^7$). Verbatim memory for related distractors was greater for 8-year-olds than for 5-year-olds ($BF_{inclusion} = 1.48 \times 10^4$) and when the detection rather than the color naming task was used as a concurrent task ($BF_{inclusion} = 26.46$). There was also substantial evidence for the interaction between age and concurrent task type ($BF_{inclusion} = 7.77$). Follow-up analyses provided strong evidence for the effect of concurrent task type in 8-year-olds ($BF_{10} = 24.48$) and substantial evidence against this effect in 5-year-olds ($BF_{10} = .20$).

The full model was also the best for gist memory for related distractor (Gr ; $BF_{10} = 2.64 \times 10^{11}$). Gist memory for related probes was higher for 8-year-olds than for 5-year-olds ($BF_{inclusion} = 2.30 \times 10^5$) and in the color naming rather than in the detection task ($BF_{inclusion} = 55.84$). There was also decisive evidence for the interaction between the two factors ($BF_{inclusion} = 2.72 \times 10^3$). Follow-up analyses indicated that the effect of concurrent task type on the Gr parameter was present only in the 8-year-old group ($BF_{10} = 1.16 \times 10^{20}$) and not in the 5-year-old group ($BF_{10} = .18$).

The best model for the guessing parameter b only included the main effect of age group ($BF_{10} = 2.98$). There was weak evidence that b was larger in 5- than 8-year-olds. The best model

for guessing parameter a also included only the main effect of age group ($BF_{10} = 7.55 \times 10^3$). Five-year-olds were more likely than 8-year-olds to identify a probe that seems familiar to them as a target rather than as a related distractor based on the guessing parameter a .

6.2.3 Discussion

In immediate recall, as expected, older children produced more correct recall than younger children and the use of WM maintenance mechanisms increased the rate of correct recall for all children. Moreover, both age groups produced equally numerous semantic errors, which does not correspond exactly to our prediction, as we expected 8-year-olds to produce more semantic errors. However, when comparing semantic and non-semantic errors, younger children produced more non-semantic than semantic errors in the condition where maintenance mechanisms were impaired, while there was a clear dominance of semantic over non-semantic errors in the 8-year-old group, suggesting a switch of error type dominance with age, from non-semantic to semantic. This switch is very likely underpinned by the development of gist memory, which creates increased activation of semantically related candidates in memory. Moreover, as expected, semantic errors increased when maintenance mechanisms were not available for use, and did so equally in both age groups. These findings suggest two things. First, that WM maintenance mechanisms are used from age 5, which was not clearly shown in the literature before. Second, that short-term false memories could be prevented by the use of WM maintenance mechanisms, starting from age 5. Regarding subjective judgments, correct recall was mainly associated with strong and precise memory traces in both age groups, although this rate was higher in 8-year-olds and when WM maintenance was available during the retention intervals. Semantic errors were equally associated with the three types of judgments, revealing variability in the strength and preciseness of the traces associated with this error type. Moreover, 8-year-olds more frequently chose the judgment ‘I heard this word or a similar one’, suggesting that they had more retrievals that could be associated to gist memory.

In delayed recognition, as expected, the rate of true recognition was higher in older children. Both age groups produced an equal amount of false recognitions, which is consistent with previous findings (Rousselle et al. 2023). Moreover, the introduction of a concurrent task impacted delayed true recognition. Indeed, when both maintenance mechanisms could be used, true recognition increased, which suggests that such mechanisms preserve memory traces. Accordingly, we found that the rate of verbatim memory on target probes associated with the

detection task condition was higher compared to the color naming condition. Hence it seems that WM maintenance mechanisms can preserve verbatim and impact recognition even a few minutes after they were used. Additionally, gist and verbatim estimates depicted more fine-grained details about processes underlying recognition as they showed some differences between age groups in gist memory on target probes, with 8-year-olds having more gist traces under WM maintenance suppression, consistent with a pattern found in young adults (Abadie & Camos, 2019), while 5-year-olds' estimates were not affected by this suppression. Finally, false memories did not increase with age, likely because both gist memory and verbatim memory on related probes increased with age. Hence, verbatim memory might have countered the effect of gist on related probes and favored correct rejection of this type of distractor. Eight-year-olds produced more correct identification of related probes as such than 5-year-olds did (59.7%, $SD = 26$, 37.1%, $SD = 32.9$; $BF_{10} = 43.59$). An additional argument is that, like in Rousselle et al. (2023), when decomposing the responses to which G_r led, we found qualitative differences across age groups. Indeed, G_r led 5-year-olds to produce false recognition ($(1 - V_r) \times G_r \times a$) in 13.5% of cases and correct identification ($(1 - V_r) \times G_r \times (1 - a)$) in 17.5%, while it led 8-year-olds to mainly produce correct identification (31%, against 10% for false recognition). Finally, verbatim and gist decomposition showed that older children's memory traces on related probes are more sensitive to the effect of concurrent tasks, even a few minutes after the task occurred.

Note that contrary to our expectations, we did not observe the expected interaction effect between age and concurrent task condition. One explanation would be that the smiley concurrent task did not completely prevent the use of maintenance mechanisms, which could have decreased semantic error rate in older children. Indeed, 79% of 8-year-olds and 47.7% of 5-year-olds reported having done rehearsal in the color naming condition. Furthermore, we did not anticipate that announcing the theme at the beginning of the list presentation might create an ambiguous error type- i.e. the recall of the theme. Finally, as younger children produced fewer 'I remember this word or a resembling one' judgments, we could wonder whether it was because they actually had less of this subjective experience type or if they simply did not have the metacognitive skills to use such judgment. We conducted a second experiment, designed to shed light on those three last points.

6.3 Experiment 2

The second experiment was similar to the first one except for three updates which addressed some questions left unanswered in the first experiment. First, in order to limit the suspected use of maintenance mechanisms during the ‘smiley’ task, we increased the pace of presentation of stimuli in both concurrent tasks. Second, we suppressed the announcement of the thematic category at the beginning of each list to avoid any ambiguity if it was produced as an error during recall. Finally, we introduced a supplementary training phase for the scale with a different task to enhance younger children’ understanding of the scale before using it in recall.

We expected to replicate the findings from Experiment 1, in recall and recognition with two exceptions. First, we expected an interaction between age group and the concurrent task type on semantic errors in recall. Indeed, because we suppressed the ‘theme’ error type that might have introduced some noise in the results and we increased the concurrent task difficulty, we expected to better capture the effects of the development of WM maintenance mechanisms on semantic errors. Therefore, we expected to find that preventing the use of WM maintenance through a concurrent task would increase false memories in older children, while the effect would be attenuated in younger children. Second, we expected that if younger children had metacognitive difficulties in understanding the subjective judgments in Experiment 1, they might benefit from the scale training. Otherwise, they should produce a similar pattern of judgements than from Experiment 1.

6.3.1 Method

Participants

Forty-seven 1st graders (the 5-year-old group; 24 girls; *mean age* = 67.55 months; *SD* = 4.96 months) and 48 3rd graders (the 8-year-old group; 30 girls; *mean age* = 103.35 months; *SD* = 5.38 months) were recruited in several preschools and elementary schools in France to participate to this study. One participant from the 5-year-old group was excluded from the data analysis because they did not follow the instructions.

Material and procedure

The material and procedure of this experiment were similar to those of Experiment 1, except for three changes. First, the concurrent task difficulty was increased by adding stimuli (i.e., smileys or circles) to be processed. Three stimuli were displayed for 1350 milliseconds or four stimuli for 950 milliseconds followed by a 250 millisecond ISI, for 5- and 8-year-olds

respectively. Second, the name of the thematic category was no longer given at the beginning of each trial. Participants were only informed by the experimenter of the start of the next trial. Finally, we added a training session to the subjective memory scale at the beginning of the experiment to ensure that children understood the three levels of the scale and could use them appropriately. This training task was adapted from Lyons and Ghetti (2011, 2013) who successfully used a similar task with children aged 3 to 5 years. Six pictures were selected from Snodgrass and Vanderwart's database (1980). Based on Cycowicz et al.'s (1997) classification by 5- and 6-year-olds, we chose pictures that children could name properly (between 80 and 100% of correct naming), among the most familiar ($M = 3.18$, $SD = .27$ on a 5-point scale) and low in complexity ($M = 3.02$, $SD = .66$ on a 5-point scale). The pictures chosen were unrelated to the thematic categories used in the memory test. Then, children were given the name of the object, animal or body part in the picture, for example a 'dog', and asked to give a judgment of 3 if they thought it was a 'dog', a judgment of 2 if they thought it might be a 'dog' or something that resembles a dog, or a judgment of 1 if they did not know if it was a 'dog'. We pixelized some of the pictures to represent the tree-scale levels. The original picture represented judgment 3 on the scale. Eighty-five or 93% of the pixels were removed from the original picture to obtain a blurred picture representing judgment 2 and very blurred representing judgment 1, respectively. There were two pictures for scores 1 and 3 and four pictures for score 2 which was more difficult to grasp especially for younger children. These pictures represented the different mental representations that children could form during recall (i.e., either clear and precise, blurred but with retrieval of some imprecise elements, or with no retrieval possible). In this task, children had direct access to the item on which they had to make a judgment, which was not the case during recall in the experiment. This allowed the experimenter to see whether children were making an objective judgment, and to provide feedback if they were not. They gave positive feedback to children when they gave a correct answer. If they gave the wrong answer, the experimenter told them that another child had given a different answer, they stated the correct answer, and asked them what they thought about it.

6.3.2 Results

Concurrent task performance

The mean performance on the detection task was 80% or above for both age groups (80.0%, $SD = 16.5$; 90.1%, $SD = 12.5$, for 5- and 8-year-olds respectively). However, performance on the color naming task was lower in both age groups (54.9%, $SD = 19.9$; 64.3%,

$SD = 20.6$, for 5- and 8-year-olds, respectively). Note, however, that both age groups performed well above chance (33%) in the color naming task ($BF_{10} = 5.31 \times 10^7$, $BF_{10} = 1.39 \times 10^{12}$, for 5- and 8-year-olds, respectively).

Recall accuracy

Responses to the recall test were classified by the same raters as in Experiment 1. The response types were the same, except that theme recall was considered as a semantic error. Interrater agreement was 96.5% and full agreement was reached after discussion. Table 6 shows the percentages of correct recall and recall errors as well as the decomposition of subjective judgments for correct recall, semantic and non-semantic errors as a function of age group and type of concurrent task.

Correct recall. Bayesian analysis revealed that the model including the main effects of age, concurrent task type, and subjective judgment, as well as a concurrent task by subjective judgment interaction and an age by subjective judgment interaction, was the best ($BF_{10} = 1.66 \times 10^{193}$). As expected, correct recall was decisively higher in 8- than 5-year-olds ($BF_{inclusion} = 6.31 \times 10^3$) and when the detection task rather than the color naming task was used as a concurrent task ($BF_{inclusion} = 2.61 \times 10^{12}$). There were differences in the subjective judgment rates associated with correct recall ($BF_{inclusion} = 4.75 \times 10^{112}$); children more often associated correct recall with the judgment ‘I heard this word’ than with judgments ‘heard or realted’ and ‘don’t know’ ($BF_{10} = 4.36 \times 10^{87}$; $BF_{10} = 1.14 \times 10^{89}$). The interaction between age and subjective judgment ($BF_{inclusion} = 7.70 \times 10^7$) was decomposed using independent samples t-tests testing the effect of age on each judgment separately. As in Experiment 1, the analyses revealed that 8-year-olds produced more ‘I heard this word’ judgments than 5-year-olds ($BF_{10} = 6.31 \times 10^4$) and there were no differences between the two age groups for the other types of judgments ($BF_{10} = .23$; $BF_{10} = .85$, for judgment 2 and 1, respectively). The interaction between subjective judgment and concurrent task ($BF_{inclusion} = 9.27 \times 10^{62}$) was decomposed using paired samples t-tests testing the effect of concurrent task on each judgment separately. The analyses revealed that judgment ‘heard’ was used more in the detection than in the color naming task ($BF_{10} = 1.07 \times 10^{25}$), which is also consistent with the results of Experiment 1.

Recall errors.

The Bayesian repeated-measures ANOVA comparing the different error types provided decisive evidence for a main effect of this factor ($BF_{10} = 4.23 \times 10^{16}$). There was decisive

evidence that phonological errors (0.6%) were less frequent than semantic errors (9.7%), intrusions (8.6%) and other errors (7.5%; $BF_{10} = 3.25 \times 10^{20}$; $BF_{10} = 2.84 \times 10^{10}$; $BF_{10} = 2.68 \times 10^9$, respectively).

Table 6.

Mean percentage of correct recall and recall errors (semantic errors, phonological errors, intrusions and other errors) and mean percentage of each subjective judgments (H, HR and DK^a) associated with correct recall, semantic and non-semantic errors as a function of age group and type of concurrent task.

Age Group	Concurrent task	Correct Recall	Semantic errors	Phonological errors	Intrusion errors	Other errors	
		H/HR/DK	H/HR/DK		H/HR/DK		
5	Detection	80.6	6.4	0.4	6.9	5.8	
		(16.4)	(7.1)	(1.4)	(8.5)	(8.9)	
	77.7/ 1.7/ 1.2		3.7/ 1.1/ 1.6		7/ 1.8/ 4.2		
	(19.2/ 4.5/ 3.5)		(5.5/ 2.5/ 4.2)		(11.3/ 4.5/ 7.4)		
Color naming	51.6	9.2	0.8	19.9	18.5		
	(22.4)	(7.7)	(2.0)	(16.5)	(17.2)		
		47.6/ 2.2/ 1.8		4.6/ 1.6/ 3		18.4/ 8/ 12.9	
		(22.2/ 4.3/ 1.3)		(5.1/ 4.6/ 4.3)		(19.9/ 13.3/ 14.2)	
8	Detection	92.2	5.5	0.4	1.0	0.4	
		(9.2)	(5.5)	(1.8)	(5.0)	(1.3)	
	90.8/ 1.2/ .2		2/ 1.4/ 2.1		0.9/ 0.6/ 0.9		
	(9.1/ 2.4/ 1.2)		(3.1/ 2.3/ 3.7)		(3.7/ 2.7/ 1.4)		
Color naming	69.4	17.6	0.8	6.4	5.8		
	(16.0)	(10)	(2.2)	(8.4)	(6.6)		
		65/ 3.1/ 1.2		5.3/ 5.6/ 6.7		4.3/ 3.1/ 5.6	
		(15.6/ 4.1/ 2.6)		(7.4/ 7.1/ 6.3)		(6.4/ 5.4/ 8.8)	

Note. Standard deviations are in brackets. ^a H = ‘I know I heard this word’, HR = ‘I heard this word or a resembling one’, DK = ‘I don’t know if I heard this word’.

Next, a Bayesian mixed measures ANOVA was conducted with age group, concurrent task type and type of errors (semantic vs non-semantic errors). The best model was the full model including all the main effects and interactions ($BF_{10} = 3.80 \times 10^{39}$). Analyses of effects

provided decisive evidence for a main effect of concurrent task ($BF_{inclusion} = 8.32 \times 10^{16}$); more errors were produced in the color naming than in the detection task. Five-year-olds produced more errors than 8-year-olds ($BF_{inclusion} = 4.16 \times 10^3$). There was more non-semantic than semantic errors ($BF_{inclusion} = 102.99$). There was decisive evidence for an interaction between error type and concurrent task type ($BF_{inclusion} = 6.256 \times 10^3$) and between age and error type ($BF_{inclusion} = 4.35 \times 10^9$), but most importantly, there was also decisive evidence for the three-way interaction between age, concurrent task type and error type ($BF_{inclusion} = 1.22 \times 10^7$). Two separate Bayesian mixed measures ANOVAs were conducted separately in each concurrent task type with age group and error type. In the detection task condition, the full model was the best ($BF_{10} = 4.19 \times 10^4$). There was weak evidence against the main effect of response type ($BF_{inclusion} = .37$). Five-year-olds made more errors than 8-year-olds ($BF_{inclusion} = 194.42$). Most importantly, there was decisive evidence for the interaction between both factors ($BF_{inclusion} = 583.91$). Independent t-tests conducted separately on each error type revealed there was no effect of age group on semantic errors ($BF_{10} = .27$). By contrast, non-semantic errors were more frequent in 5- than 8-year-olds ($BF_{10} = 3.79 \times 10^3$). Concerning the color naming task condition, the full model was also the best ($BF_{10} = 1.49 \times 10^{17}$). There was strong to decisive evidence for the main effects of age group and error type ($BF_{inclusion} = 19.57$; $BF_{inclusion} = 1.07 \times 10^4$, respectively) and for an interaction between both factors ($BF_{inclusion} = 1.12 \times 10^{12}$). Five-year-olds produced more non-semantic errors than 8-year-olds ($BF_{10} = 2.49 \times 10^7$). By contrast, 8-year-olds produced more semantic errors than 5-year-olds ($BF_{10} = 1.25 \times 10^3$).

Finally, two Bayesian mixed measures ANOVAs were conducted separately on each error type (semantic and non-semantic) with age, concurrent task type and subjective judgments. Concerning semantic errors, the best model included main effects of age, concurrent task and an interaction between both factors ($BF_{10} = 4.90 \times 10^9$). There was no difference in the type of subjective judgments ($BF_{inclusion} = .67$), all of which were used equally frequently by the children. For non-semantic errors, the best model included main effects of age, concurrent task, and subjective judgment as well as interactions between concurrent task and age and between subjective judgment and age ($BF_{10} = 7.94 \times 10^{20}$). There was substantial evidence for a main effect of subjective judgment ($BF_{inclusion} = 7.78$). Post-hoc analyses showed that judgment ‘heard’ and ‘don’t know’ were more frequent than judgment ‘heard or related’ ($BF_{10} = 99.48$; $BF_{10} = 5.34$, respectively). Independent t-tests were conducted separately on each age group to decompose the interaction between subjective judgment and age group ($BF_{10} = 3.62$). Analyses

revealed that 5-year-olds used judgments ‘heard’ and ‘don’t know’ more often when making non-semantic errors than did 8-year-olds ($BF_{10} = 2.71 \times 10^3$; $BF_{10} = 49.12$, respectively).

Recognition accuracy

Discriminability indexes on correct and false recognition are shown in Figure 13. As in Experiment 1, we conducted mixed measures Bayesian ANOVAs on correct and false recognition with age group and concurrent task. For correct recognition, the best model included only the main effect of concurrent task ($BF_{10} = 1.34 \times 10^5$). There were more correct recognitions with the detection task than with the color naming task. For false recognition, the best model included only the main effect of age group ($BF_{10} = 1.78$). There was weak evidence that the rate of false recognition was higher in 8- than 5-year-olds.

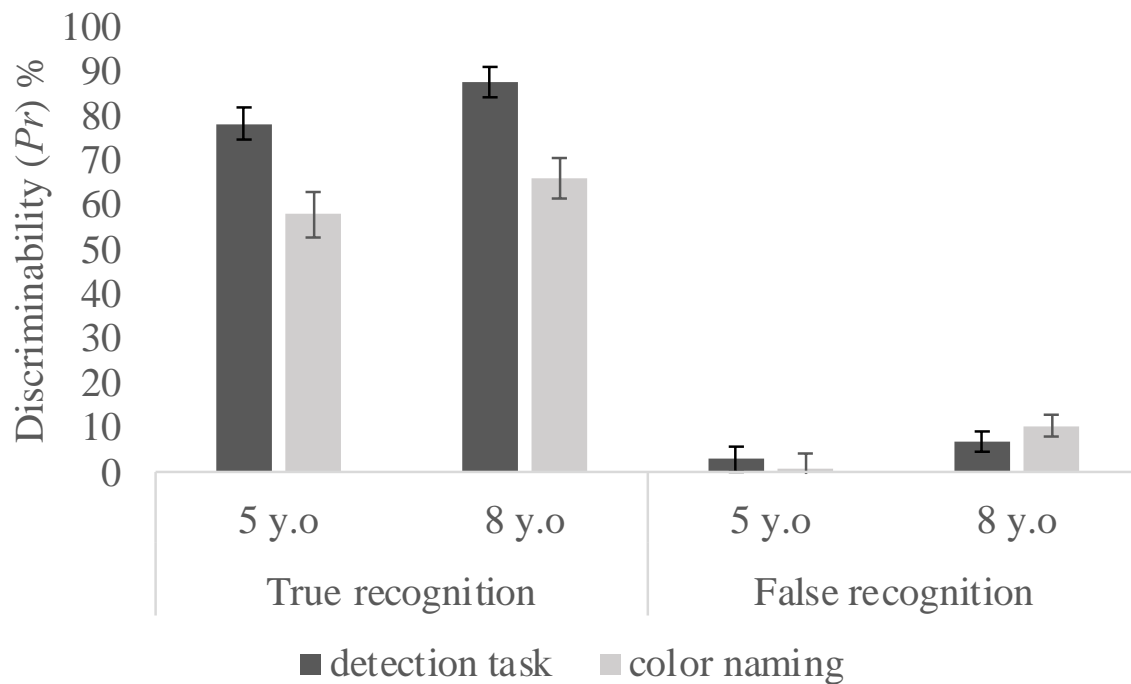


Figure 13. Correct and false recognition as a function of age group and type of concurrent task. Error bars represent standard errors.

Verbatim, gist memory and guessing processes. Verbatim, gist and guessing parameter estimates are given in Table 7. The data fit the model well, as indicated by nonsignificant test results in the 5-year-old group ($T_1: p = .57$, $T_2: p = .65$; $T_1: p = .10$, $T_2: p = .72$, for the detection and the color naming task, respectively) and in the 8-year-old group ($T_1: p = .39$, $T_2: p = .48$; $T_1: p = .11$, $T_2: p = .58$, for the detection and the color naming task, respectively).

Table 7.*Mean parameter estimates of the SCR model as a function of age and concurrent task type.*

	5		8	
	Detection	color naming	detection	color naming
V_t	.79 (.03)	.60 (.07)	.86 (.12)	.63 (.04)
G_t	.50 (.18)	.33 (.08)	.62 (.10)	.52 (.30)
V_r	.13 (.06)	.12 (.06)	.28 (.10)	.19 (.08)
G_r	.16 (.05)	.12 (.06)	.40 (.03)	.49 (.09)
b	.33 (.23)	.31 (.25)	.12 (.05)	.20 (.07)
a	.31 (.26)	.54 (.28)	.24 (.11)	.27 (.11)

Note. Standard deviations are in brackets.

For verbatim memory for targets (V_t), the full model was the best ($BF_{10} = 1.36 \times 10^{47}$). There was decisive evidence that verbatim memory for targets was greater in 8- than 5-year-olds ($BF_{inclusion} = 261.35$) and in the detection task than in the color naming task ($BF_{inclusion} = 2.63 \times 10^{46}$). The analyses of effects provided weak evidence for the interaction between age and concurrent task ($BF_{inclusion} = 1.63$).

The additive model with main effects of age and concurrent task was the best for gist memory for targets (G_t ; $BF_{10} = 6.36 \times 10^9$). Gist memory for targets was greater in 8- than in 5-year-olds ($BF_{inclusion} = 1.11 \times 10^4$) and greater in the detection task than in the color naming task ($BF_{inclusion} = 1.34 \times 10^5$).

The full model was the best for verbatim memory for related distractors (V_r ; $BF_{10} = 5.87 \times 10^{17}$). Verbatim memory for related distractors was higher in 8- than 5-year-olds ($BF_{inclusion} = 5.78 \times 10^{11}$) and higher in the detection task than in the color naming task ($BF_{inclusion} = 1.62 \times 10^3$). There was decisive evidence for the interaction between both factors ($BF_{inclusion} = 1.16 \times 10^3$). Bayesian paired-sample t-tests provided decisive evidence for the effect of concurrent task in 8-year-olds ($BF_{10} = 1.37 \times 10^4$) and substantial evidence against this effect in 5-year-olds ($BF_{10} = .17$).

The full model was also the best for gist memory for related distractors (G_r ; $BF_{10} = 1.59 \times 10^{55}$). Gist memory for related probes was higher for 8- than 5-year-olds ($BF_{inclusion} = 2.23 \times 10^{44}$) and higher in the color naming than in the detection task ($BF_{inclusion} = 3.84$). Most importantly, there was decisive evidence for an interaction between age group and concurrent task condition ($BF_{inclusion} = 2.4 \times 10^{12}$). Bayesian paired sample t-tests provided decisive evidence that gist memory for related distractors was higher in the detection task than in the color naming task for 5-year-olds ($BF_{10} = 611.14$) whereas the reverse was true for 8-year-olds ($BF_{10} = 1.49 \times 10^7$).

The best model for the guessing parameter b was the full model ($BF_{10} = 3.26 \times 10^4$). b was larger for 5- than 8-year-olds ($BF_{inclusion} = 2.75 \times 10^3$). There was weak evidence for a main effect of concurrent task ($BF_{inclusion} = 1.32$). However, there was substantial evidence for the interaction between the two factors ($BF_{inclusion} = 9.63$). Follow-up Bayesian paired t-tests provided decisive evidence that b was higher in the color naming task than in the detection task in 8-year-olds ($BF_{10} = 1.94 \times 10^8$), and substantial evidence against an effect of concurrent task in 5-year-olds ($BF_{10} = .17$).

Finally, the best model for the guessing parameter a was also the full model ($BF_{10} = 7.54 \times 10^{10}$). a was greater in 5- than in 8-year-olds ($BF_{inclusion} = 1.01 \times 10^4$) and greater in the color naming task than in the detection task ($BF_{inclusion} = 4.04 \times 10^3$). There was decisive evidence for an interaction between the two variables ($BF_{inclusion} = 280.64$). Follow-up tests provided decisive evidence for the main effect of concurrent task in 5-year-olds ($BF_{10} = 3.12 \times 10^3$), and substantial evidence against this effect in 8-year-olds ($BF_{10} = .31$).

6.3.3 Discussion

The goal of this second experiment was to replicate Experiment 1's findings except that we expected that the manipulation of maintenance mechanisms in this experiment and the removal of the theme presentation would help reveal age differences on semantic errors. We also tested whether providing supplementary training to the subjective judgment scale would change its use by children of different age groups, especially younger children. For correct immediate recall, results from Experiment 1 were replicated with increased correct recall in older children, and when maintenance mechanisms could be used during the retention interval. This response was also more often associated with a detailed strong memory trace, which was itself increased in the 8-year-olds' group and when WM maintenance could be used. Semantic recall errors were increased by the suppression of WM maintenance mechanisms and by

contrast with Experiment 1, were more frequent in older children. Even more interestingly, the effect of age on semantic error was present only when maintenance mechanisms were blocked. When they could be used, the rate of semantic errors for 8-year-olds decreased to the rate of 5-year-olds, suggesting that using such mechanisms prevents semantic errors in older children. By contrast, younger children tended to produce more non-semantic errors when they could not maintain information during the retention interval, which suggests once again that they use WM maintenance mechanisms, but that it prevents the occurrence of other errors more than semantic errors in this age group. Moreover, in this study, semantic errors were associated with all three subjective judgements, which corroborates Experiment 1's findings and suggest that such errors are based on different types of traces. There were no longer any age differences in the use of judgment 2. Regarding delayed recognition, as in the first experiment, correct recognition was increased when target probes were associated to the detection task. This effect was underpinned by more verbatim and gist memory. Notice that in contrast with Experiment 1, there was no age effect on this response type. False recognition rate as in Exp. 1 was not affected by age nor concurrent task. Older children had also more gist memory, and more verbatim on related probes, which likely countered the effect of gist on false memories. Besides gist memory on related probes, also led older children to produce more correct identification of related probes as such (25.5%) compared to false recognition (8.8%), while it led younger children to produce correct (7.2%) and false memories (5%) equally.

6.4 General Discussion

This study aimed at assessing the role of WM maintenance mechanisms and of LTM traces in children of different ages in false memories in a DRM-like paradigm adapted to WM tasks. Two experiments were conducted with 5- and 8-year-old children, who did a complex span task in which they had to learn lists or thematically related words, interspersed with either a detection task or a color naming task and followed by immediate recall test and delayed recognition. The detection task aimed at allowing the use of rehearsal and refreshing, whereas the color naming task aimed at preventing them. During recall, a subjective experience scale was implemented to assess the subjective experience associated to each recalled word. Based on a delayed recognition test, the SCR model of the FTT was used to assess the respective contribution of gist, verbatim memory, and guessing to responses in recognition. In the second experiment, the concurrent task was harder because children had to process more items in the same amount of time, the announcement of the lists' theme was removed, and there was supplementary training for subjective judgement scale.

Principal findings showed that false memories were prevented by WM maintenance mechanisms in 8-year-olds in immediate recall but not in delayed recognition. Moreover, false memories increased with age in immediate recall but not in delayed recognition. Indeed, 8-year-olds produced predominantly semantic errors in recall while 5-year-olds produced predominantly non-semantic errors including mostly unrelated errors, and intrusions from previous lists, which most likely reflects the immaturity of their executive system (see McCormack et al., 2000, for a review). Semantic errors were associated with all three subjective judgements, which reflects that this error type is not always based on similar traces. Consistent with the FTT, in delayed recognition, older children were found to have more gist and verbatim memory than younger children. Moreover, WM maintenance mechanisms impacted 8-year-olds' gist and verbatim memory on related probes, showing that they have a moderator role on LTM traces at longer delays than a few seconds in this age group.

False memories at different ages and the role of WM maintenance mechanisms.

Consistently with our predictions, correct recall increased with age and younger children produced more non-semantic than semantic errors while this trend reversed in older children. This suggests that both verbatim and gist memory's development with age impact immediate recall in WM tasks, like they impact recall in classical DRM tasks. Moreover, in the second experiment, 8-year-olds produced more semantic errors than 5-year-olds and only this age group had a reduced rate of semantic errors when WM maintenance mechanisms were available. It confirms our expectations that maintenance mechanisms play a more important role in false memories for older children in immediate recall. For younger children, in contrast, WM maintenance mechanisms mainly reduced the rate of non-semantic errors. Moreover, in both age groups, WM maintenance also boosted correct recall, which is consistent with previous findings (e.g., Barrouillet et al., 2009; Oftinger & Camos, 2016, 2018). Hence it seems that younger children are able to maintain information in WM, which support the view that articulatory rehearsal is already used at age 5 (e.g., Conrad, 1971; Elliott et al., 2021; Miller et al. 2015), and it seems that maintenance mechanisms promote correct recall in both age groups and prevents the dominant error type produced by each age group.

In delayed recognition, as expected and consistent with previous findings (Rousselle et al., 2023), 5- and 8-year-olds produced equally as many false recognitions, and WM maintenance mechanisms did not impact this error type. By contrast, older children produced more true recognition than younger children in Experiment 1 and the use of WM maintenance

mechanisms during the 5.8 second retention intervals increased true recognition in both experiments. This last finding suggests that maintenance mechanisms impact performance minutes after they were used, starting from age 5.

Subjective experience in recall.

Regarding subjective judgments associated with recall, consistent with previous findings (Flegal et al., 2010, 2014), correct recall was mostly associated with precise, strong memory traces that could be attributed to remembering phenomenology, in which a detailed memory trace is retrieved. As expected, there were more strong and precise judgements in older than younger children. Semantic errors were attributed to the three types of judgements, suggesting that this error type is not always based on similar memory traces and preciseness in children. One third is usually associated to strong memory traces, while one third is usually associated to fuzzier meaning-based traces and the last third to guessing judgements. An important aspect to note is that semantic errors in children are mainly based on memory traces (either strong or fuzzy) and not simply on guessing. Moreover, in Exp. 1 younger children associated semantic errors to fuzzy meaning-based judgments less often than older children. However, this difference disappeared in Exp. 2, suggesting that the supplementary training helped children understand the scale better. Hence, contrary to our expectations, older children did not seem to have different types of subjective experience when compared to younger children. Although, in Exp. 2 descriptively, younger children used proportionally fewer fuzzy based judgment than older children. It is therefore possible that the variability within age groups and the low rate of each response masked age differences on this fuzzy judgement.

The role of gist and verbatim in delayed recognition.

Gist and verbatim parameters were higher in older children which is consistent with the FTT showing that those traces develop with age. It is however intriguing that false recognition did not increase between age groups. As already discussed however, it seems that the increase of verbatim in older children and response differences between age groups to which gist memory led would compensate for the gist increase with age and explain the similar rates of false memory. This raises the interesting point that responses to a memory test do not always capture process differences across age groups. Hence, using the SCR revealed differences in memory traces across age groups that were not observable in raw recognition answers. In fact, another finding that was not observable directly on false recognition, was the decrease of

verbatim memory under WM maintenance suppression in 8-year-olds. Interestingly, in Experiment 1, the suppression of such mechanisms decreased gist memory in older children, which is consistent with Abadie & Camos' findings (2019), who suggested that gist memory was preserved by the use of refreshing. However, these findings were not replicated in Exp. 2, where gist memory was increased when WM maintenance mechanisms were suppressed. The main difference between Exp 1 and 2 is that the cognitive load was stronger in the second experiment and that the category theme was not introduced. However, we are not sure that it explains the differences of findings between studies, and deserves future investigation. Note that in our study, both or neither maintenance mechanisms could be used at one, hence, we cannot completely compare our results to Abadie & Camos', who isolated the effect of refreshing on gist memory alone. Finally, the decrease of delayed true recognition under WM maintenance suppression was mirrored by verbatim memory decrease under this condition and by the increase of guessing in both age groups (*b* in 8-year-olds and *a* in 5-year-olds, in Exp. 2). It supports the idea that maintaining words to learn during retention delays of a few seconds promotes verbatim memory and increases correct recall. More generally, results on gist and verbatim show that WM maintenance mechanisms can impact memory traces at delays longer than a few seconds, as shown by Abadie & Camos (2019). This impact does not however always appear on recognition responses, such as false memories.

6.5 Conclusion

In conclusion, our study showed that WM maintenance mechanisms reduced the occurrence of false memories in short-term recall in 8-year-old children. Semantic errors were dominant in this age group, while non-semantic errors predominated in 5-year-olds, which suggests a switch in error type in WM tasks between those ages. This switch is likely directly related to the development of gist memory with age. Moreover, both age groups produced as many delayed false recognitions, but older children had more gist and verbatim memory. The suppression of WM maintenance mechanisms also decreased gist and verbatim memory for older children. These results warn us about an evolution of error types with age in WM tasks. Besides, it indicates that older children's short-term false memories could be limited by using tasks that would favor the use of WM maintenance.

Summary

False memories in working memory tasks were shown to be reduced by WM maintenance mechanisms in adults. In the present study, we examine this matter in children. WM maintenance mechanisms develop between 5 and 7 years old, hence we compared the production of short-term semantic errors in 5- and 8-year-old children in a complex-span task. Children had to memorize 3 or 4 semantically related words, depending on age group. During the 4.8 second retention interval introduced between the presentation of words, a concurrent task was introduced. In one condition, it impaired the use of WM maintenance mechanisms, whereas in the other condition, the mechanisms could be used. At the end of the presentation of a list of words, participants had to recall the studied words. Phenomenological experience associated to recall was assessed, and a delayed recognition test was also introduced at the end of the presentation of several lists of words. Results revealed that semantic errors became dominant with age: younger children produced more non-semantic errors, while it was the opposite for older children. Moreover, the suppression of WM maintenance mechanisms amplified the dominant error type of each age group. The results suggest that these mechanisms reduce the dominant error type produced by each age group when they are used. Measures of phenomenological experience revealed that both age groups associated semantic errors with strong, detailed memories, but also to fuzzier gist-based memories and to guessing judgements. Moreover, in the delayed recognition test, similar rates of false recognition were produced between age groups. However, when we used the simplified conjoint recognition model to estimate memory traces underlying false recognition, we found that older children had more meaning-based (i.e., gist) and surface detail-based (i.e., verbatim) memory traces than young children. Verbatim traces were also found to be reduced and gist traces increased by the suppression of WM maintenance mechanisms in the older age group, suggesting that the effect of these mechanisms on memory traces last longer than just a few seconds.

Résumé

Il a été montré que les faux souvenirs en tâche de mémoire de travail étaient réduits par les mécanismes de maintien de la mémoire de travail chez les adultes. Dans la présente étude, nous avons examiné cet effet chez les enfants. Les mécanismes de maintien se développent entre l'âge de 5 et 7 ans. En conséquence, nous avons comparé la production d'erreurs sémantiques dans une tâche d'empan complexe, chez des 5 et 8 ans. Les enfants devaient mémoriser 3 ou 4 mots reliés sémantiquement, selon le groupe d'âge. Une tâche concurrente était introduite durant l'intervalle de rétention de 4,8 secondes, entre la présentation des mots. Dans une condition, cette tâche empêchait l'utilisation des mécanismes de maintien. Dans une autre condition, les mécanismes pouvaient être utilisés. À la fin de la présentation d'une liste de mots, les participants devaient rappeler les mots étudiés. L'expérience phénoménologique associée à ce rappel était mesurée. Un test de reconnaissance différé suivait la présentation de plusieurs listes de mots. Les résultats indiquent que les erreurs sémantiques deviennent dominantes avec l'âge. Les jeunes enfants ont produit en effet plus d'erreurs non-sémantiques, contrairement aux enfants plus âgés. De plus, la suppression des mécanismes de maintien a amplifié principalement le type d'erreur dominant de chaque groupe d'âge. Ceci suggère que ces mécanismes de maintien diminuent le type d'erreur dominant produit par chaque groupe d'âge, lorsqu'ils peuvent être utilisés. Les mesures de l'expérience phénoménologique ont révélé que chaque groupe d'âge associait les erreurs sémantiques à de forts souvenirs détaillés mais aussi à des souvenirs plus flous, ainsi qu'à des jugements de hasard. De plus, en reconnaissance différée, des taux similaires de fausses reconnaissances étaient produits entre les groupes d'âge. Cependant, lorsque nous avons utilisé le modèle de reconnaissance conjointe simplifiée pour estimer les traces sous-tendant les faux souvenirs, les enfants plus âgés avaient plus de traces gist et verbatim en mémoire que les enfants plus jeunes. Nous avons également observé que les traces verbatim et les traces gist augmentaient chez les enfants plus âgés lorsque les mécanismes de maintien en mémoire de travail étaient supprimés. Ceci suggère que les mécanismes de maintien ont un effet sur les traces en mémoire pendant des délais plus longs que quelques secondes.

Chapter 7. The role of semantic similarity and working memory maintenance mechanisms in short-term false memories in children and adults.

7.1 Introduction

The topic of false memories has fascinated researchers over the past mid-century. Knowing that we could falsely remember an event or at least parts of an event can be quite intriguing indeed because it suggests that the human memory has flaws. These limitations have been shown to cause wrongful justice decisions when relying on erroneous eyewitness testimony (e.g., McMartin Preschool trial, Eberle & Eberle, 1993). In the literature, there has been evidence that false memories at long but also at short-term are facilitated by the presentation of items sharing a semantic similarity (e.g., McEvoy et al., 1999; Tehan, 2010). A recent study has shown that studying only four items with high semantic similarity fostered the processing of item meaning (i.e., gist memory), which caused the emergence of short-term false memories (Abadie & Camos, 2019). Additionally, there was evidence that short-term false memories could be decreased by the use of articulatory rehearsal, a maintenance mechanism of the WM, consisting of recirculation of memory traces in an articulatory loop to preserve its phonological form (Barrouillet & Camos, 2022). In the present study, we tested the effect of semantic similarity on short-term false memories in recognition and in recall as well as the potential moderator role of WM maintenance mechanisms. As LTM representations, such as gist memory, develop with age, and WM maintenance mechanisms become increasingly efficient between childhood and adulthood, we compared a group of older children with a group of young adults.

Semantic similarity in false memories

One of the most popular paradigms to detect false memories is the Deese-Roediger and McDermott, paradigm (Deese, 1959a; Roediger & McDermott, 1995). It consists of presenting participants lists of 15 words related in meaning to a critical item (e.g., ‘candy, sugar, sour, good, taste, honey’ are semantically related to ‘sweet’) for later recall and recognition. Results indicate that participants falsely recall or recognize ‘sweet’, or other semantically related words as part of the studied list, more than any unrelated word (Roediger & McDermott, 1995).

Interestingly, studies have shown that the more studied words were judged as semantically similar to the critical item, the more correct but also false memories were produced. Howe et al. (2009) indeed showed that the backward associative strength (BAS), a score reflective of the probability that an individual would produce the critical item (i.e., ‘sweet’ in the example above) as a first response when presented one of the words of the list, was a good predictor of false memory occurrences in children and in adults. The stronger the score was, the more false memories were produced. Brainerd et al. (2020) also showed that the gist strength score of DRM lists, which reflects the estimated relatedness strength between the words of a list, was an even better predictor of false memories. Hence, it seems that the chances of false memories are increased by the presentation of lists of words with strong semantic similarity.

According to the Fuzzy-Trace-Theory (FTT, Brainerd & Reyna, 2002a; Brainerd et al., 2008a; Chang & Brainerd, 2021), one of the most prominent theories in explaining false memories, this semantic similarity effect is underlined by the activation of memory traces in our LTM. Two types of traces are described by the FTT: gist and verbatim memory. Gist traces can be defined as representations of the semantic content of items and other relational information such as taxonomic, synonymous, or situational relations. This type of trace favors correct recall but also false memories, as illustrated by the relationship between gist scores and false memory rate in Brainerd et al.’s study (2020). Opposed to gist traces is verbatim memory. It represents detailed perceptive features of stimuli. Such type of trace promotes correct recall but contrary to gist traces, prevents false memories. The chances of false memories are therefore determined by the balance between gist and verbatim memory. Indeed, if gist traces are strong, like it is the case when lists of words with high semantic similarity are presented at study, verbatim memory might not counter the activation of semantically related distractors caused by gist memory. It would in turn cause higher rates of false memories.

False memories at short-term.

One particularity of verbatim traces is that it declines fast in time and in particular, faster than gist memory (Brainerd & Reyna, 2002b). Therefore, most studies measuring false memories used LTM tasks, such as the classical DRM task, in which recall or recognition occurs minutes after the study of the first stimuli, and for which the number of stimuli to learn, exceeds WM capacity. Under such circumstances, verbatim memory is not strong enough to counter gist memory, especially when semantically related words are presented at study. Some studies (Poirier & Saint-Aubin, 1995; Poirier et al., 2011; Saint-Aubin & Poirier, 1999a; Tehan,

2010), manipulated the semantic similarity of lists of words presented at study. Results of these studies indicate an increase of correct recall on lists of only 6 semantically related words, short-term correct memory is increased. A few other studies have shown that false memories occurred at short-term, even though the rate was lower than in LTM tasks (i.e., between 20-30%; Atkins & Reuter-Lorenz, 2008; Flegal et al. 2010, 2014; Olszewsha et al., 2015). Abadie and Camos (2019) investigated the semantic similarity effect in short-term false memories. In this study, participants were presented a WM task, in which they had to learn lists of four highly or weakly semantically related words. After a retention interval of 4 seconds, a recognition task was introduced. Results revealed that participants produced more false memories on related than unrelated word lists. Similarly, Atkins et al. (2011) showed that when the 4 lists-words were all related to a same thematic category, false memories increased in young adults compared to when half of the lists-words were related to one category (i.e., fruits) and the other half to another (i.e., country). In their study, Abadie and Camos (2019) used the simplified conjoint recognition model of the FTT (SCR, Sthal & Klauer, 2008) to get an objective estimation of the contribution of gist and verbatim traces to false memories in recognition. They found that gist memory estimates were higher for related than unrelated lists of words, which is additional evidence for the role of gist memory in short-term false memories. A developmental study using a similar paradigm to Abadie and Camos', with a measure of gist traces via SCR model, showed that children from age 4 also produced gist-based false memories (Rousselle et al, 2023). Moreover, other studies have measured the phenomenological experience associated to the short-term false memory illusion. Abadie et al., (in rev., Exp 4), have shown that adult participants associated semantic recall errors in WM tasks mainly to a statement attributable to gist memory ('I studied this word or a word related in meaning to this word'), but also to guessing or to a more detailed vivid memory statement (i.e., 'I studied this word'). Flegal et al. (2010) also showed that false recognitions of distractors thematically related to the studied words, were associated in one third of cases to 'remember' judgments, which corresponds to a detailed memory experience and in one third of cases to 'know' judgments which can be assimilated to experience based on fuzzier memory. The last third was assimilated to guessing experience. In a same vein, Rousselle et al. (in prep.) showed that one third of semantic recall errors in 8-year-old children were associated to the same 'gist' statement as in Abadie et al. (in rev.), one third to even more detailed remembrance ('I heard this word'), and the last third to a guessing statement.

The role of working memory maintenance mechanisms.

One particularity of false memory at short-term, is that it was found to be not only impacted by the activation of gist memory, but also by WM maintenance mechanisms. In particular, WM tasks, which simulate real-life situations (e.g., taking notes during a class, calculate the price of a basket at the grocery shop, etc.) require constant shifting between information maintenance and processing (Barrouillet & Camos, 2015). Verbal information is usually maintained through mechanisms when the processing of a concurrent stimuli is not ongoing or interfering with it. We can distinguish two types of maintenance mechanisms. The first one, described above, is articulatory rehearsal, which allows the maintenance of verbal information (e.g., Baddeley, 1986). The second maintenance mechanism relies on attentional resources and was called *attentional refreshing* in the Time-Based-Resource-Sharing model of WM (Barrouillet & Camos, 2021), which is one of the most influential models of WM. Attentional refreshing consists of briefly thinking back to recently active memory items (Camos, 2015, 2017). Several studies have shown that when adults had the opportunity to rehearse the words to learn, they produced fewer immediate false memories than when rehearsal was prevented by a concurrent articulation (Abadie & Camos, 2019; in rev.; Atkins et al. 2011; Macé & Caza, 2011). Abadie and Camos (2019) even showed that the use of rehearsal preserved verbatim memory using the SCR model, explaining in consequence the decrease in false memories in immediate recognition. For refreshing, Abadie and Camos (2019, in rev.) tested the specific effect of refreshing on short-term false memories and showed no evidence that being able to maintain memory items through refreshing during the 4 second retention interval following stimuli presentation impacted false memories. Turning to children, two studies have recently tested the impact of WM maintenance mechanisms on false memories in children. Rousselle et al. (2023) presented lists of 3 to 5 semantically related words to 4- and 8-year-olds, followed by a 6-8 second retention interval. During this interval, WM maintenance mechanisms could either be used freely, or were impaired by a concurrent task asking for a concurrent articulation, to block rehearsal and asking for attentional resources to block refreshing (Exp 2. and 3.). Results reveal that false memories in the immediate recognition test, were not impacted by the manipulation of maintenance mechanisms in any age groups, even though there is evidence that maintenance mechanisms develop between 5- and 7-years-old (see Camos & Barrouillet, 2018, for a review). Another study from Rousselle et al. (2023) showed however that short-term semantic errors in recall were increased in 8-year-olds, when WM maintenance

mechanisms were impaired by a concurrent task, which was introduced during the 5-second retention intervals between the presentation of each semantically related word.

Comparing false memories between children and adults.

Studies using the DRM paradigm, showed that false memories increased between childhood and adulthood (e.g., Anastasi & Rhodes, 2008; Brainerd et al., 2002c, 2018; Howe, 2005; Paz-Alonso et al., 2008). False recognition of semantic distractors is indeed produced at higher rates in adults than in children, as well as semantic errors in recall. The FTT explains this increase by a development of gist memory traces with age. As children grow, their ability to categorize information and make meaningful relationships between stimuli through experienced events develops (Brainerd & Reyna, 2002a). Some experimental manipulations revealed evidence for this. For instance, when gist processing is enhanced by lists adapted in meaning to age groups (e.g., Anastasi & Rhodes, 2008; Carneiro et al., 2007; Metzger et al., 2008) or by using visual stimuli (Ghetti et al., 2002; Howe, 2006; Howe et al., 2004; Khanna & Corteze, 2009) or using ecological paradigms (e.g., Howe & Wilkinson, 2011; Lyons et al., 2010; Odegard et al., 2009), the developmental trend on false memories was reduced. Howe et al. (2009) also showed a correlation between the increase of false memories with age and the BAS score associated to DRM lists. Moreover, verbatim memory also develops with age, but because gist memory increase is sharper than verbatim (Brainerd et al., 2002c; Reyna & Kiernan, 1994) and that gist memory is also more resistant to time decay than verbatim memory (Brainerd & Reyna, 2002b), most studies using the classical DRM paradigm observed an age increase in false memories from childhood to adulthood. Note however that in some studies using shorter lists of DRM words (i.e., 7), the developmental effect on false memories was attenuated or even suppressed, when comparing older children (i.e., between 8 and 11-year-olds) and young adults (Metzger et al., 2008; Ghetti, et al., 2002; Sugrue & Hayne, 2006; Sugrue, et al., 2009). Hence, the increase of short-term false memories with age might not be as strong as in LTM tasks.

Furthermore, WM maintenance mechanisms were shown to develop with age. There is indeed evidence in the literature that rehearsal starts to be used between 5- and 7-years-old (Elliott et al, 2021; Henry et al., 2012; Tam et al., 2010) and keeps on increasing in efficiency until adulthood (e.g. Cowan, Saults & Morey, 2006). Attentional refreshing starts to be efficient from age 7 and keeps on maturing until age 15 (Barrouillet et al., 2009, in prep.). A very recent study compared the production of short-term false memories of 8-10-year-old children and

young adults using a Brown-Peterson task (Abadie & Rousselle, in press.). Participants were shown lists of 8-semantically related words, followed by an 8-second aloud high attentionally demanding parity judgment task or a silent low attentionally demanding detection task. The first condition aimed at preventing the use of rehearsal and refreshing for maintenance of word-lists more than the second condition. In the following immediate recognition test, participants produced high rates of short-term false memories in both age groups (around 48% for adults, around 43% for children). Adults produced more false memories than children. The manipulation of WM maintenance mechanisms did not however impact false memory production in any age group. Most importantly, the increase of false recognition in adults compared to children was associated with an increase of phantom recollection, a phenomenon relating to strong detailed memories, based on gist memory and accompanied with recollective experience (Brainerd et al., 2001). It suggests differences of memory traces underlying short-term false memories between children and adults.

The present study.

It seems that previous studies have shown that semantic relatedness was a determinant in the occurrence of false memories in long-term but also in WM tasks. Moreover, it seems that maintenance mechanisms specific to WM tasks, might moderate the occurrence of false memories. In the present study, we assessed the extent to which semantic relatedness of studied items impacted short-term false memories in children and in adults. We also investigated the moderator role of WM maintenance mechanisms in short-term false memories. Young children aged 9 and young adults completed a complex span task. Lists of six words were presented. Three words were semantically related and the other three were unrelated. In between each word-list, a concurrent task was introduced for 5.4 seconds. Participants had to say aloud whether digits appearing sequentially on the screen were odd or even. The pace of digit presentation varied. It was either slow or fast. In the slow pace condition, refreshing and rehearsal opportunities were greater than in the fast pace condition, because participants had more time to reallocate their attentional resources on thinking back to the item (i.e., refreshing) and they also had more time to repeat the studied words. In Experiment 1, each list of words was followed by a recognition test composed of a target, a semantically related and an unrelated probe. We measured the underlying gist and verbatim traces to false recognition, using the SCR model. In Experiment 2, a free recall test replaced the recognition test. We assessed the phenomenological experience associated to responses in recall.

Because the FTT predicts an increase in false memories with age due to gist memory increase, we expected to find an increase of false memories in older children in both experiments. Furthermore, because studying semantically related words should increase gist memory, we expected to observe more false memories produced on these words than on unrelated words in both recognition and recall. This semantic similarity effect was expected to be bigger in adults than in children because gist memory is thought to be more developed in this age group. Regarding the effect of WM maintenance mechanisms suppression on false memories, we expected different outcomes in recognition than recall. Indeed, if studies using recall memory tests have shown that preventing WM maintenance mechanisms in immediate recall decreased false memories in adults (Abadie et al., in rev.) as well as in children (Rousselle et al., in prep.), studies using recognition tests showed more contrasted patterns of results. Some studies found that short-term false recognition was increased by the suppression of WM maintenance mechanisms (e.g., Abadie & Camos, 2019; Atkins et al., 2011; Macé & Caza, 2011), while others showed no effect in adults and children (e.g., Abadie & Rousselle, in press.; Rousselle et al., 2023). Hence, we expected that suppressing WM maintenance mechanisms would increase false memories in recall and that this effect might be attenuated in recognition tasks, which seems to be less sensitive to such manipulation and for which contrary to the patterns observed at long-term, the rates of short-term false memories are often lower compared to recall tasks.

7.2 Experiment 1

In this Experiment, lists of six words, with three semantically related and three semantically unrelated words, were presented to 9-year-old children and to young adults. In between each word, a 5.4 second concurrent task was introduced, during which participants had to say aloud and on the keyboard, whether numbers presented sequentially on screen were odd or even. The pace of this task was either ‘slow’ or two times faster, to vary the time available for rehearsal, and refreshing in between each word presentation. At the end of each list, participants were shown three probe types: a word from the studied list (i.e., target), a word related either to the three studied related words or to one of the unrelated studied words (i.e., related) or a word unrelated to any studied words and not from the studied list (i.e., unrelated). Predictions are described in the previous section.

Furthermore, in this Experiment, we used the SCR model of the FTT to estimate gist and verbatim memory processes underlying recognition. We expected that adults would have

more gist and verbatim memory than children. We expected related list-words to cause more gist traces than unrelated list-words. This effect was expected to be greater in adults for whom gist memory is more developed. Moreover, the use of WM maintenance mechanisms should increase both verbatim and gist memory in adults. Indeed, Abadie and Camos (2019) showed that rehearsal was responsible for verbatim increases and refreshing for gist memory increases in adults.

7.2.1 Method

Participants

Forty-six 4th grade children were recruited in an elementary French school (31 girls; *mean age* = 9.04 years; *SD* = .36, range 8.08-10.3 years). They were all native French speakers. Forty-nine young adults were recruited at Aix-Marseille University (35 girls; *mean age* = 19.88 years; *SD* = 1.75, range 18.16 – 25.25 years). They participated in the study in exchange for course credit. The study was conducted in accordance with the APA Ethics Code. For each participant, a written consent agreement was obtained from the participant (young adults) or from their parent (children). Ethic approval was obtained for both experiments from the institutional review board of Aix-Marseille University (“Working memory and false memories during childhood”, protocol number: 2019-12-12-003).

Material

Twelve lists of words containing three thematically related words and three unrelated words were created based on the French lists of words used by Abadie & Camos (2019). All words were recorded by a computerized female French voice on the website VoiceBooking.com (VoiceBooking Team, 2021). Out of the four words from Abadie & Camos’ lists, we kept three words to compose our lists (e.g. ‘rooster’, ‘chick’, ‘hen’ are related to the theme word chicken) excluding verbs and favoring the words with the smallest syllable number possible. Unrelated words composing the rest of the lists were chosen from other thematical lists in Abadie & Camos (2019). Each unrelated word from the list belonged to a different thematic list (e.g. ‘fire’, ‘pasta’, ‘palm’). Two extra lists were created for the training phase of the experiment. Lists can be found at (<https://osf.io/ebm7y/>). Among the 12 lists of words, 6 were associated with one experimental block condition which was determined by the concurrent task pace (see procedure) and the other 6 with the other block. This block-list association was balanced

between subjects such as one subject was presented with the first six lists in one block condition and another subject was presented with the last six lists in the same block condition.

Furthermore, three probe types were associated to each list for recognition: *targets*, *related* and *unrelated* probes. Target probes were words studied in the lists. For every subject, each list was associated with either a target probe that belonged to the three thematically related words for half of the lists (e.g., rooster) or that belonged to the unrelated words for the other half (e.g., fire). Between subjects, the target probe of a same list was either one of the related words or one of the unrelated words. For instance, the list ‘rooster’, ‘chick’, ‘hen’, ‘fire’, ‘pasta’, ‘palm’, was paired with the target probe ‘rooster’ for one subject or to the target probe ‘fire’ for another subject. Related probes were distractors that were not part of the lists of studied words. They were related to either the three semantically related words in the list (e.g., ‘chicken’ related to ‘rooster’, ‘chick’, ‘hen’), or to one of the three unrelated words in the list (e.g., ‘matches’ related to ‘fire’). These probes corresponded to the critical lure in Abadie & Camos’ lists (2019). Between subjects, the same list was shown half of the time with a probe related to the three thematic words or to one of the unrelated words. Unrelated probes were unstudied words that were not thematically related to any words of the presented list. There were two variants for a same list (‘sail’ or ‘key’, with the previous example).

Procedure

This experiment was displayed on a computer and built with E-Prime 3.0 software (version 3.0.3.82, 2018). Participants were asked to do a complex span task (Figure 14) in which they were presented lists of six spoken words that they were asked to retain (see material for list composition). Words in each list were presented randomly however, in order to increase the likelihood of semantic processing between the three thematically related words of the list, we made sure that the position of those three words were following each other (e.g., Lampinen et al. 2006). The position of those three words varied between lists and was fixed so that every position was used across lists. For instance, in the first list, the 3 related words were presented in position 1,2 and 3. In the second list, they were presented in the position 2,3,4, and so on. In between each word they were asked to perform a parity judgment task. In this task digits were displayed on the screen sequentially and participants had to say out loud and by pressing the matching keyboard button if the digit was odd or even among 9 different digits. This task lasted 5.4 seconds and was preceded by a 250 ms ‘bip’ to remind the participant of the task switch. In order to vary the availability of WM maintenance mechanisms during this task, the pace of digit

presentation during this interval varied. Digits could be presented at a ‘slow’ pace in half of the trials or a ‘fast’ pace in the other half. The ‘fast’ pace condition was meant to prevent the use of WM maintenance mechanisms during the concurrent task more than the ‘slow’ pace condition. Each condition was blocked and randomized in presentation order. Moreover, the pace of digit presentation was adapted to the age group so that the parity task would be equally as difficult for both targeted age groups. Based on Rosselet-Jordan et al. (2022), who used the same task with the same age groups, 2 stimuli were presented in the slow pace condition (2.4 second each) in the group of children and 3 stimuli were presented in the group of adults (1.5 second each) with an interval inter-stimuli (ISI) of 300 milliseconds in both age groups. In the fast pace condition, the number of presented stimuli was doubled (i.e. 4 for children, each stimulus lasting 1.05 second and 6 for adults, each lasting 600 ms), with the same interval inter-stimuli duration. For each list, after the last concurrent task was displayed, three recognition probes were presented sequentially in a random order: a target, a related and an unrelated probe. Participants were first asked for each probe whether they heard this word in the studied list. If they answered ‘yes’, it was recorded by the experimenter as ‘target’. If they answered ‘no’, they were asked: ‘could this word be related to one or some of the words of the study list?’. If they answered, ‘yes’ the experimenter noted their answer as ‘related’. If they answered ‘no’, it was noted as ‘unrelated’.

At the beginning of each block, participants were trained on the concurrent task first without time limit and then with time limit matching the experimental condition block that would follow. They were also trained on the complex span task with one list. At the end of the Experiment, children were rewarded with a diploma.

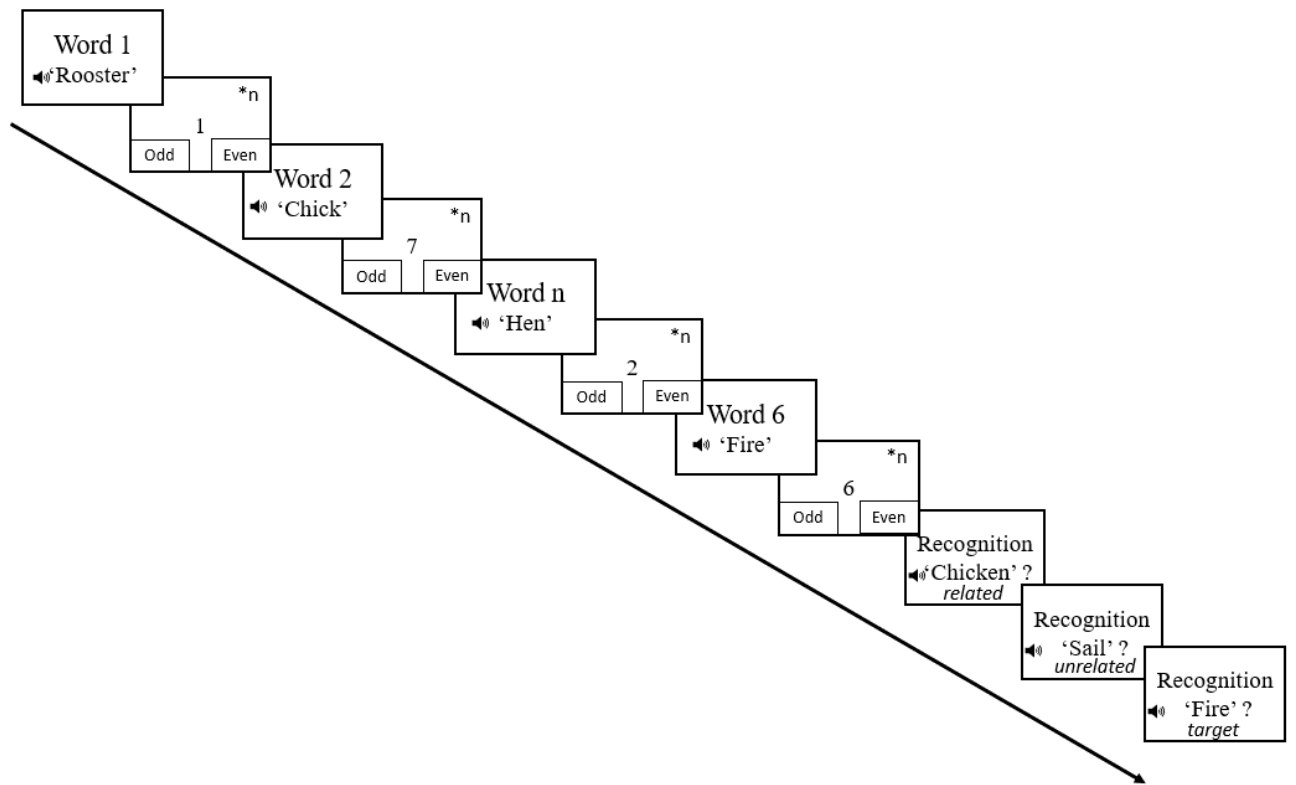


Figure 14. Illustration of Experiment 1’s procedure for one trial. $*n = 2 \times (2.4 \text{ seconds} + 300 \text{ ms ISI})$ for children and $3 \times (1.5 \text{ sec.} + 300 \text{ ms ISI})$ for adults in the ‘slow’ pace condition and $4 \times (1.06 \text{ sec.} + 300 \text{ ms ISI})$ for children and $6 \times (600 \text{ ms} + 300 \text{ ms ISI})$ for adults in the ‘fast’ pace condition; there is a bip of 250 ms between the end of a word presentation and the beginning of the parity concurrent task.

7.2.2 Results

Data for both Experiments can be found at <https://osf.io/ebm7y/>. We conducted Bayesian analysis with JASP Version 0.16.3 (Jasp Team, 2022), using default mode settings. All generated models were compared to the null. We reported the model with the largest Bayes Factor (BF_{10}). The value for each effect (i.e., main effect or interaction) included in the best model, indicated by $BF_{inclusion}$ was then reported. It indicates the likelihood of each model including a given effect compared to all models stripped of that effect. The strength of the Bayes Factors was interpreted based on Kass and Raftery (1995) norms.

First, we report analysis on the accuracy of the concurrent task, comparing the two age groups, and the pace of the task. Second, we report analysis on correct and false recognition, with age group, pace of the concurrent task and relatedness condition as tested factors. Finally, we report gist, verbatim estimates on both target and related probes, as well as guessing

parameters and describe analysis on each parameter with age, concurrent task and relatedness as tested factors.

Concurrent task

Two children were excluded from the analysis because their performances at the concurrent task in the fast pace condition were under 2.5 standard deviation from the mean group performance. A BANOVA was conducted on performance at the concurrent task with age group as between-subject factor (children vs. adults) and pace as within-subject factor (slow vs. fast). The best model was the full ($BF_{10} = 2.14 \times 10^{40}$). There was decisive evidence that performance increased in the slow pace condition compared to fast pace ($M_{slow} = 80.8\%$, $SD = 6.4$; $M_{fast} = 60.6\%$, $SD = 11.8$; $BF_{inclusion} = 1.08 \times 10^{38}$). There was only weak evidence for an effect of age ($BF_{inclusion} = .55$) but there was decisive evidence for an interaction between pace and age group ($BF_{inclusion} = 439.18$). T-test revealed that the pace effect was stronger in adults than children ($M_{slow} = 81.6\%$, $SD = 6.6$; $M_{fast} = 57.6\%$, $SD = 12.6$; $BF_{10} = 1.77 \times 10^{19}$; $M_{slow} = 80\%$, $SD = 6.1$; $M_{fast} = 63.9\%$, $SD = 9.9$; $BF_{10} = 1.47 \times 10^{17}$, respectively).

True and false recognition

Discriminability indexes (Pr) on correct and false recognition were computed following the guidelines provided by different studies comparing age groups regarding developmental differences in ‘yes-saying’ bias to recognition (e.g., Otgaar et al., 2014). For correct recall, ‘target’ responses to unrelated probes were subtracted from the proportion of ‘target’ responses to target probes. For false recognition, ‘target’ responses to unrelated probes were subtracted from ‘target’ responses to related probes. True and false recognition proportions are depicted in Figure 15.

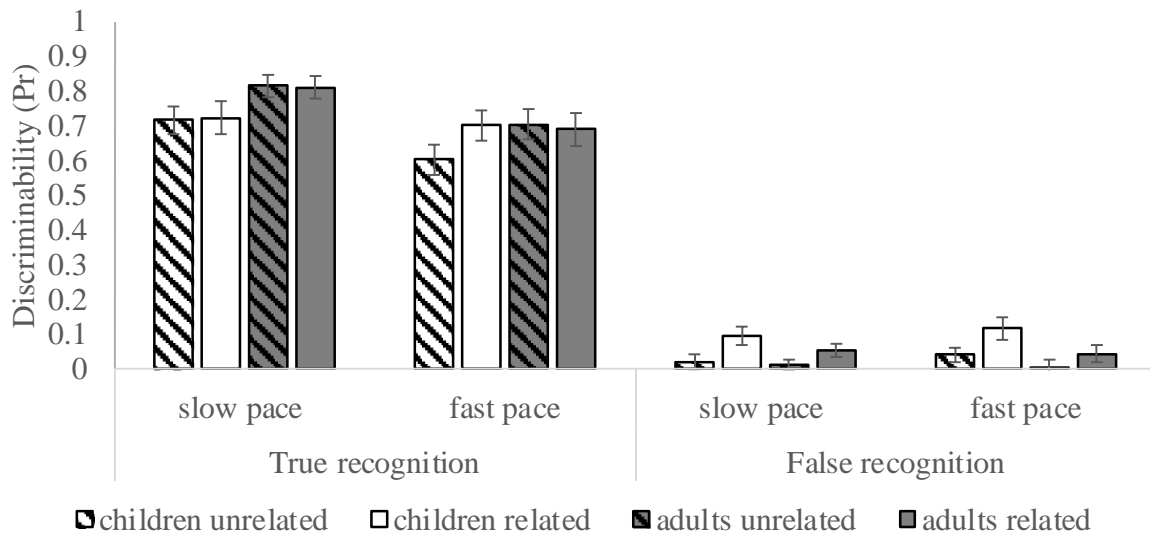


Figure 15. True and false recognition proportion as a function of age group and word-relatedness concurrent task pace in Experiment 1. Error bars represent standard errors.

Mixed-measures BANOVA were conducted on true and false recognition with age group as between subject-factor and words-relatedness (related vs. unrelated) and concurrent task pace as within subject-factor. For true recognition, the best model contained a main effect of pace ($BF_{10} = 44.54$). There was more true recognition in the slow than fast pace condition ($M_{slow} = .77, SD = .2; M_{fast} = .68, SD = .24$, respectively). For false recognition, the best model contained a main effect of word relatedness ($BF_{10} = 2.27 \times 10^3$). There was more false recognition on related words than unrelated words ($M_{related} = .09, SD = .13; M_{unrelated} = .03, SD = .09$, respectively).

Gist and verbatim memory

The multinomial model of the simplified conjoint recognition model (Stahl & Klauer, 2008) was used to compute parameter estimates for verbatim and gist memory for targets (Vt , Gt , respectively) and for related probes (Vr , Gr , respectively) and guessing parameters (b and a). The parameters Vt and Gt represent the probability of retrieving a verbatim (Vt) or a gist trace (Gt) of a target when a target probe is presented at test. The parameters Vr and Gr correspond to the probability of retrieving a verbatim (Vr) or a gist trace (Gr) when a related probe is presented at test. Verbatim retrieval is assumed to lead to correct identification of target and related probes. When verbatim retrieval fails, participants can retrieve gist memory. They were familiar with the meaning of the target or the related probe, but they cannot remember whether the probe itself or a related one with the same gist was presented at study. They then

guess whether the probe is a target (with probability a) or a related distractor (with probability $1 - a$). When neither verbatim nor gist memory is available, a participant can still guess that the probe meaning is old with the probability b . The decision between the target and related responses is again modelled by the parameter a . Otherwise, the participant guesses that the probe is new with the probability $1 - b$. Verbatim and gist traces do not intervene in the responses to unrelated distractors because these probes do not trigger the retrieval of verbatim or gist representations of the study phase. Therefore, the responses to unrelated distractors are based entirely on b .

To compute each parameter, hierarchical analyses were conducted using the latent-trait approach (Klauer, 2010) implemented in the TreeBUGS package used on R software (Heck et al., 2018). This method allowed individual parameters to be computed. Mean parameter estimates are presented in Table 8.

The model fit was assessed with T_1 which is the distance between the observed and the expected mean frequencies, and T_2 which is the summed distance between the observed and the expected covariance statistics. The data fitted the model well, as indicated by nonsignificant test results in the children's group in both the slow pace ($T_1: p = .44, T_2: p = .52; T_1: p = .51, T_2: p = .39$; for unrelated and related word-lists, respectively) and the fast pace condition ($T_1: p = .37, T_2: p = .27; T_1: p = .21, T_2: p = .43$; for unrelated and related word-lists, respectively) and in the adults group, both for slow ($T_1: p = .47, T_2: p = .58; T_1: p = .53, T_2: p = .67$; for unrelated and related word-lists, respectively) and fast pace ($T_1: p = .36, T_2: p = .54; T_1: p = .44, T_2: p = .51$; for unrelated and related word-lists, respectively).

We conducted a mixed-measures BANOVA on individual estimates of each parameter with age group as between- and concurrent task type and word-list relatedness as within-subject factors.

Table 8.

Mean estimates of each parameter to the SCR model depending on age, concurrent task pace and word-list relatedness.

Age		Children				Adults			
group	CT	slow pace	fast pace	slow pace	fast pace	slow pace	fast pace	slow pace	fast pace
Word-list	related-	Un		R		Un		R	
		ness	Un	R	Un	R	Un	R	Un
Vt		.73 (.045)	.67 (.183)	.60 (.079)	.66 (.078)	.81 (.029)	.78 (.033)	.71 (.136)	.66 (.204)
Gt		.11 (.045)	.59 (.109)	.06 (.012)	.34 (.155)	.28 (.109)	.63 (.178)	.07 (.02)	.65 (.068)
Vr		.17 (.069)	.27 (.075)	.10 (.015)	.26 (.267)	.43 (.061)	.59 (.203)	.23 (.097)	.40 (.151)
Gr		.12 (.026)	.43 (.092)	.18 (.101)	.43 (.187)	.23 (.050)	.71 (.124)	.26 (.197)	.72 (.108)
b		.10 (.07)	.11 (.099)	.09 (.032)	.10 (.023)	.07 (.02)	.07 (.017)	.14 (.06)	.16 (.066)
a		.41 (.087)	.39 (.146)	.29 (.052)	.41 (.057)	.25 (.054)	.22 (.026)	.18 (.05)	.21 (.025)

Note. Standard deviations are in brackets. CT = concurrent task, Un = unrelated, R = related.

For the *Vt*, the full model was the best ($BF_{10} = 1.25 \times 10^{16}$). Analysis of effect showed evidence for a main effects of age group ($BF_{Inclusion} = 1.04 \times 10^4$), of concurrent task pace ($BF_{Inclusion} = 5.19 \times 10^{11}$), and for a three-way interaction between pace, age group and word relatedness in lists ($BF_{Inclusion} = 28.99$). *Vt* was higher in adults than children and in the slow than the fast pace condition. To decompose the three-way interaction, we conducted two separate BANOVAs on unrelated and related word in lists, separately testing age and concurrent task pace effects. For unrelated words, the best model contained age and pace main effects ($BF_{10} = 8.45 \times 10^{21}$). Adults had more *Vt* than children ($BF_{Inclusion} = 1.65 \times 10^8$) and participants had more *Vt* in the slow pace condition ($BF_{Inclusion} = 1.77 \times 10^{16}$). There was however weak evidence for an interaction between both variables ($BF = .53$). By contrast, for related words, the best model was the full ($BF_{10} = 1.63 \times 10^3$). *Vt* was higher under slow than fast pace ($BF_{Inclusion} = 46.91$) and for adults than children ($BF_{Inclusion} = 3.15$), but most importantly, there was strong evidence for an interaction between age and concurrent task pace ($BF_{Inclusion} = 12.28$). Paired sample t-tests, revealed that the effect of pace was stronger in adults than in children ($BF_{10} = 5.72 \times 10^4$; $BF_{10} = 109.53$, respectively).

For *Gt*, the best model was the full ($BF_{10} = 8.46 \times 10^{146}$). Adults had more *Gt* than children ($BF_{Inclusion} = 4.06 \times 10^{13}$). There was more *Gt* in the slow pace condition than fast pace and for related studied words ($BF_{Inclusion} = 1.80 \times 10^{19}$; $BF_{Inclusion} = 2.53 \times 10^{119}$, respectively). Regarding interactions, evidence was above weak only for the interaction between age and relatedness ($BF_{Inclusion} = 59.23$) and for the three-way interaction ($BF_{Inclusion} = 3.13 \times 10^{19}$). We directly decomposed the three-way interaction, running two separate BANOVAs on related and unrelated studied words with age and concurrent task pace as tested factors. For unrelated words, the best model was the full ($BF_{10} = 1.18 \times 10^{44}$). *Gt* was higher for adults and under slow pace ($BF_{Inclusion} = 1.58 \times 10^{11}$, $BF_{Inclusion} = 1.82 \times 10^{24}$, respectively). There was decisive evidence for the interaction between age group and pace. T-tests showed that the pace effect was greater in adults than in children ($BF_{10} = 5.39 \times 10^{15}$, $BF_{10} = 2.20 \times 10^7$, respectively). For related studied words, the full model was the best ($BF_{10} = 4.56 \times 10^{22}$). Age, pace effects and the interaction between both were confirmed by the analysis of effects ($BF_{Inclusion} = 2.25 \times 10^9$, $BF_{Inclusion} = 2.42 \times 10^4$, $BF_{Inclusion} = 3.93 \times 10^{10}$, respectively). T-tests revealed that *Gt* memory dropped in children under fast pace ($BF_{10} = 5.27 \times 10^{12}$) but not in adults ($BF_{10} = .21$).

Regarding *Vr*, the best model contained all three main effects and an interaction effect between age group and pace ($BF_{10} = 6.02 \times 10^{48}$). *Vr* traces were higher in adults than children ($BF_{Inclusion} = 3.61 \times 10^{19}$), under slow than fast pace ($BF_{Inclusion} = 1.02 \times 10^{14}$) and for related

studied words than unrelated ones ($BF_{\text{Inclusion}} = 5.22 \times 10^{20}$). There was decisive evidence for the interaction between age and pace ($BF_{\text{Inclusion}} = 3.97 \times 10^5$). T-tests, revealed an effect of pace in adults ($BF_{10} = 2.77 \times 10^{10}$) but not in children ($BF_{10} = .97$).

For *Gr*, the best model contained the three tested main effects and an interaction between age and relatedness ($BF_{10} = 1.83 \times 10^{112}$). Adults had more *Gr* than children ($BF_{\text{Inclusion}} = 7.71 \times 10^{19}$) and there was more *Gr* on related studied words ($BF_{\text{Inclusion}} = 3.16 \times 10^{92}$). There was only weak evidence for an effect of pace ($BF_{\text{Inclusion}} = 1.39$). There was however evidence for the interaction between age and relatedness ($BF_{\text{Inclusion}} = 1.66 \times 10^{12}$). T-tests revealed that the relatedness effect was greater in adults than in children ($BF_{10} = 1.09 \times 10^{27}$; $BF_{10} = 8.35 \times 10^{16}$, respectively).

For the parameter *b*, the best model included main effects of age, pace and the interaction between the two ($BF_{10} = 1.43 \times 10^{26}$). Analysis of effects provided substantial evidence only for the effect of pace and the interaction between pace and age ($BF_{\text{Inclusion}} = 5.98 \times 10^{10}$; $BF_{\text{Inclusion}} = 8.81 \times 10^{17}$, respectively). *b* was larger under fast pace. T-tests revealed that the pace effect concerned only adults and not children ($BF_{10} = 4.67 \times 10^9$; $BF_{10} = .21$, respectively).

Finally, for the parameter *a*, the best model included the three tested main effects and an interaction effect between pace and relatedness and between age and relatedness ($BF_{10} = 2.53 \times 10^{45}$). Regarding main effect, *a* was larger in children, under slow pace and for related words ($BF_{\text{Inclusion}} = 6.69 \times 10^{27}$; $BF_{\text{Inclusion}} = 3.09 \times 10^8$; $BF_{\text{Inclusion}} = 101.09$, respectively). T-tests used to decompose the interaction between pace and age ($BF_{\text{Inclusion}} = 30.96$) revealed evidence for a relatedness effect only in children ($BF_{10} = 6.58 \times 10^3$). There was decisive evidence for the interaction between relatedness and pace ($BF_{\text{Inclusion}} = 9.48 \times 10^8$). T-tests revealed evidence for a relatedness effect on *a* only in the fast pace concurrent task condition ($BF_{10} = 4.41 \times 10^{14}$).

7.2.3 Discussion

As expected, false recognition was increased when the probe was related to three related studied words, in both age groups. It indicates that a stronger gist memory activation directly increases false memories in immediate recognition in WM tasks. These findings are supported by the estimation of underlying memory processes done with the SCR. Indeed, gist memory was shown to be increased by related word-lists in both children and adults.

Consistently with previous studies using similar manipulations (e.g., Abadie & Camos, 2019), when WM maintenance mechanisms were more strongly impeded due to the presentation of a faster aloud parity judgment task, true recognition decreased. This effect is mirrored by an increase of gist and verbatim memory on target probes under slow compared to fast pace. There was however no direct effect of concurrent task pace on false recognitions, which is consistent with some findings in the literature (Abadie & Rousselle, in rev.; Rousselle et al., 2023). This is mirrored by rivaling rates on gist memory on related probes in the slow and fast pace condition. This will be further discussed in the general discussion.

The absence of age and pace effect on false recognition might reflect the insensitivity of recognition tests found and discussed in other previous studies (e.g., Uittenhove et al., 2019). Indeed, even though false memory did not increase with age, adults had more gist and verbatim memory on related probes than children. Moreover, verbatim traces in adults were impacted by the suppression of maintenance mechanisms. This suggests that WM maintenance mechanisms might have a role in short-term false memories and that short-term false memories should increase with age. Note also that in our experiment, participants produced low rates of false recognition. As there was evidence that recall tests might be more sensitive than recognition tests in WM tasks (Uittenhove et al., 2019), we conducted a second experiment, in which we replaced the recognition test with a recall test.

7.3 Experiment 2

This Experiment was similar to Experiment 1 except that recognition was replaced by a recall test. Semantic recall errors were expected to be increased in adults compared to children. As in Experiment 1, semantic errors were expected to increase with the presentation of semantically related words. Moreover, we expected that semantic errors would increase under fast compared to slow pace. This effect was expected to be larger in adults.

We also assessed the phenomenological experience associated to recall answers. Participants were asked to say whether (a) they knew they heard the recalled word (i.e., precise strong memory), (b) the recalled word was a related word (i.e., recollection-rejection of the recalled word, which corresponds to the retrieval of the precise specific memory of the studied words, or some of them, allowing participants to know that they haven't heard that word), (c) they thought they heard this word or a word similar in meaning (i.e., fuzzier memory) or (d) if they did not know if they heard this word (i.e., guessing). Based on previous studies using

similar measures in adults (Flegal et al., 2010, 2014, Abadie & Camos, in rev.) and in children (Rousselle et al., in prep.), participants were expected to have associated semantic errors with strong detailed judgments (i.e., ‘heard’), with fuzzier gist-based judgements (i.e., ‘heard or related’) and with guessing judgements. By contrast, correct recall was predicted to be associated mainly with strong detailed judgements. Moreover, we expected that adults would produce more semantic errors associated to ‘Heard or related’ judgment than children, mirroring the development of gist memory with age. We also expected adults to produce more ‘heard’ judgement on correct recall, because of the increase of verbatim with age. Moreover, we expected more ‘Heard or related’ judgements in the related than unrelated word-list condition for semantic errors and correct recall. Finally, we expected that the slow pace condition would lead to more ‘Heard’ and ‘Heard or related’ judgments compared to the fast pace condition, in which the rate of ‘guess’ judgements should increase.

7.3.1 Method

Participants

Forty-nine 4th grade children were recruited in an elementary French school (25 girls; *mean age* = 9.12 years; *SD* = .33, range 9.17 – 10.16 years). They were all native French speakers and had no history of neurological disorder according to their teachers. Forty-nine young adults (between 18 and 25 years-old included) were recruited at Aix-Marseille University (38 girls; *mean age* = 21.2 years; *SD* = 1.99, range 21.16 – 25.58). They participated in the study in exchange for course credit. For each participant, a written consent agreement was obtained from the participant himself or from their legal tutor.

Material and Procedure

The material and procedure were similar to the one in Experiment 1 except that instead of presenting recognition probes at the end of each list trial, participants were asked to recall the studied words. When participants did not recall 6 words, they were informed how many words were missing. If they still did not recall all the words, they were encouraged to say the words they thought they heard. Next, a scale of subjective memory experience was introduced. The experimenter gave back the words children recalled, and they were asked to respond for each word whether: (a) it was a word they were asked to memorize (‘*heard*’ response), (b), it was a word related in meaning to one or several words they studied (‘*related*’ response), (c) it was

either a studied or a related word (*'heard or related'* response), (d) they guessed this word (*'guessing'* response).

7.3.2 Results

As in the first Experiment, we report analysis on the concurrent task accuracy. Then, we report analysis on correct recall, semantic and non-semantic errors, with age, concurrent task pace, and relatedness as tested factors (except for relatedness for non-semantic errors). Finally, we report analysis on each judgment associated to correct recall, semantic and then non-semantic errors with the same tested factors.

Concurrent task

Two children were excluded from the analysis because their performance on the concurrent task in the fast pace condition was less than 2.5 SD from the group mean performance. A BANOVA was conducted on concurrent task performance with age group and pace. The full model was the best ($BF_{10} = 6.47 \times 10^{44}$). There was decisive evidence that performance increased in the slow pace condition compared to fast pace ($M_{slow} = 92.4\%$, $SD = 7$; $M_{fast} = 65.4\%$, $SD = 16.7$; $BF_{inclusion} = 3.21 \times 10^{35}$). There was substantial evidence that children's performance rate was higher than adults ($M_{children} = 82.1\%$, $SD = 9.95$; $M_{adults} = 76.7$, $SD = 10.75$; $BF_{inclusion} = 6.46$) and there was decisive evidence for an interaction between pace and age group ($BF_{inclusion} = 3.28 \times 10^7$). T-test revealed that the pace effect was stronger in adults than children ($M_{slow} = 94.5\%$, $SD = 5.8$; $M_{fast} = 58.8\%$, $SD = 15.7$; $BF_{10} = 1.14 \times 10^{21}$; $M_{slow} = 90.3\%$, $SD = 7.5$; $M_{fast} = 73.9\%$, $SD = 12.4$; $BF_{10} = 8.33 \times 10^{12}$, respectively).

Recall task

Responses to recall were classified by two trained raters. They were regrouped under six response types (see Table 9 for the estimates). The first was correct recall, which was when the recalled word matched a studied word, no matter the position in which they were recalled. The second was semantic errors, which was the recall of a word non-presented at study but judged as semantically related to either one of the three independent presented words or to the three related studied words (Atkins & Reuter-Lorenz, 2008). The third error type was phonological errors, which were words sharing at least the same phoneme at the beginning, middle or end of the word (see Gupta et al. 2005, for classification of phonological errors). The fourth error type was intrusions from previous word lists. Words were considered as intrusions if they were heard

in previous lists, produced during previous recall or semantically related to one of the last two. The last two error types were other errors which had no link to the studied words and omissions. Interrater agreement was 95.8 %. After discussion, full agreement was obtained.

Table 9.

Proportion of correct recall and each error type (semantic, phonological, other, intrusions, omissions) as a function of age group, pace of the concurrent task and word-list relatedness for correct recall and semantic errors. +33 6 26 38 46 77

		slow pace		fast pace	
		unrelated	related	unrelated	related
Correct recall	children	0.57 (.138)	0.68 (.148)	0.49 (.107)	0.53 (.174)
	adults	0.80 (.124)	0.87 (.11)	0.65 (.128)	0.73 (.166)
Recall Error					
semantic	children	0.05 (.102)	0.09 (.116)	0.07 (.119)	0.12 (.160)
	adults	0.03 (.058)	0.06 (.073)	0.04 (.051)	0.12 (.109)
phonologic	children	0.02 (.024)		0.02 (.029)	
	adults	0.02 (.022)		0.01 (.019)	
other	children	0.04 (.067)		0.06 (.096)	
	adults	0.002 (.008)		0.02 (.036)	
intrusions	children	0.04 (.045)		0.04 (.06)	
	adults	0.01 (.048)		0.02 (.061)	
omission	children	0.2 (.02)		0.26 (.028)	
	adults	0.09 (.079)		0.18 (.015)	

Note. Standard deviation are in brackets.

Correct recall, semantic and non-semantic errors.

BANOVAs were conducted on both correct recall and semantic errors comparing age groups, pace of the concurrent task and word-list relatedness. For correct recall, the best model contained main effects of age, pace and word relatedness ($BF_{10} = 4.36 \times 10^{40}$). Adults performed better on correct recall than children ($BF_{inclusion} = 1.84 \times 10^{13}$). There was more correct recall of related than unrelated studied words ($BF_{inclusion} = 2.72 \times 10^8$). Finally, there was more correct recall in the slow than the fast pace condition ($BF_{inclusion} = 4.08 \times 10^{23}$). For semantic errors, the best model contained main effects of concurrent task, relatedness and an interaction between the two ($BF_{10} = 3.22 \times 10^8$). There was strongly more semantic errors in the fast than slow pace condition ($BF_{inclusion} = 37.7$). There was also decisive evidence that there were more semantic errors related to related studied words than unrelated ones ($BF_{inclusion} = 4.55 \times 10^5$). There was only weak evidence for an interaction between concurrent task and relatedness ($BF_{inclusion} = 2.17$).

Finally, we regrouped phonological intrusions and other errors as ‘non-semantic errors’ and conducted a BANOVA on it with age group, concurrent task pace as tested factors. The best model contained main effects of concurrent task and age group ($BF_{10} = 1.07 \times 10^5$). There were more non-semantic errors under fast than slow pace ($BF_{inclusion} = 18.8$). Children produced more non-semantic errors than adults ($BF_{inclusion} = 566.34$).

Phenomenological experience.

Subjective judgments associated to correct recall, semantic and non-semantic errors are presented in Table 10. A Bayesian ANOVA was conducted on correct recall testing the differences between each subjective judgement (heard, related, related or heard, guess). Then, we conducted a BANOVA on each distinct phenomenological experience, with age group, concurrent task pace and relatedness. We did the same for semantic and non-semantic errors, without testing the relatedness factor for non-semantic errors. Note that we did not split phonological error depending on the relatedness condition given the low rates of this error type.

Table 10.

Rate of subjective judgments associated to correct recall, semantic and non-semantic errors depending on age group, word-list relatedness and pace of the concurrent task.

Response type	Age group	CT pace	Word-lists	Judgment				
				Guessing	Related or heard	Realted	Heard	
Correct recall	children	slow	Unrelated	.005 (.016)	.008 (.026)	.004 (.014)	.558 (.144)	
			Related	.005 (.019)	.013 (.031)	.013 (.026)	.651 (.15)	
		fast	Unrelated	.002 (.011)	.008 (.023)	.005 (.019)	.474 (.113)	
			Related	.005 (.016)	.012 (.028)	.017 (.048)	.499 (.183)	
	adults	slow	Unrelated	.008 (.02)	.012 (.04)	.006 (.02)	.772 (.136)	
			Related	.003 (.018)	.039 (.053)	.022 (.044)	.807 (.13)	
		fast	Unrelated	.006 (.02)	.017 (.036)	0	.628 (.134)	
			Related	.002 (.011)	.045 (.063)	.026 (.051)	.652 (.169)	
	Semantic error	children	slow	Unrelated	.020 (.076)	.009 (.024)	.011 (.03)	.013 (.024)
				Related	.018 (.057)	.025 (.043)	.026 (.067)	.025 (.045)
			fast	Unrelated	.009 (.027)	.019 (.033)	.027 (.096)	.015 (.032)
				Related	.020 (.057)	.019 (.045)	.039 (.089)	.044 (.068)
adults		slow	Unrelated	.007 (.033)	.008 (.023)	.014 (.029)	.006 (.02)	
			Related	.003 (.013)	.016 (.032)	.02 (.043)	.020 (.039)	
		fast	Unrelated	.011 (.032)	.008 (.023)	.014 (.035)	.006 (.017)	
			Related	.023 (.049)	.027 (.044)	.041 (.061)	.026 (.032)	
Non-semantic error		children	slow		.031 (.056)	.021 (.038)	.011 (.021)	.034 (.043)
			fast		.045 (.076)	.026 (.045)	.019 (.038)	.036 (.058)
		adults	slow		.007 (.014)	.006 (.013)	.003 (.010)	.014 (.021)
			fast		.017 (.034)	.016 (.026)	.005 (.012)	.011 (.017)

Note. Standard error is in brackets. The sum of rates for each line equals to the rate of the response type in the corresponding condition (e.g., the addition of the rate of the first line

equals to .575, which is the rate of correct recall under slow pace condition in children for unrelated words). CT = concurrent task.

When comparing subjective judgment to correct recall, the BANOVA confirmed a main effect of subjective judgement as best model ($BF_{10} = 6.06 \times 10^{224}$). Post-hoc analysis revealed that there was mainly 'heard' judgements compared to 'related' ($BF_{10} = 2.93 \times 10^{61}$), 'related or heard' ($BF_{10} = 1.20 \times 10^{60}$) or 'guessing' judgments ($BF_{10} = 7.32 \times 10^{61}$). There were also more 'heard or related' judgments than 'guessing' judgments ($BF_{10} = 6.6 \times 10^4$). There were more 'related' than 'guessing' judgments ($BF_{10} = 8.94$).

The best model for judgment 'Heard' was the additive model containing the three main effects ($BF_{10} = 2.01 \times 10^{28}$). This judgement was more used in the slow than fast pace condition ($BF_{inclusion} = 2.66 \times 10^{18}$), when words were related ($BF_{inclusion} = 40.61$) and by adults than children ($BF_{inclusion} = 3.39 \times 10^9$). The best model for the judgment 'Related' was a main effect of relatedness ($BF_{10} = 1.05 \times 10^4$). This judgment was more used on related words. The best model for judgment 'Related or Heard' contained main effects of relatedness, age group and an interaction between the two ($BF_{10} = 2.42 \times 10^4$). There was evidence that this judgment was more used on related than unrelated words ($BF_{inclusion} = 90.016$) and in adults than in children ($BF_{inclusion} = 34.86$). The breakdown of the interaction between age and relatedness ($BF_{inclusion} = 6.92$) with independent sample t-tests, revealed that adults used this judgement more than children but only for related words ($BF_{10} = 176.49$; compared to $BF_{10} = .41$). The best model for 'Guessing' judgment was the null ($BF_{10} = 1.0$).

For semantic errors, when comparing subjective judgments, the best model was the null ($BF_{10} = 1.00$). It means that there were as many 'heard', 'related', 'related or heard' and guessing judgments.

The best model for judgment 'Heard' contained main effects of relatedness and age group ($BF_{10} = 5.53 \times 10^3$). This judgment was more frequent in the related word condition ($BF_{inclusion} = 5.04 \times 10^3$) There was weak evidence for a main effect of age ($BF_{inclusion} = 1.1$). The best model for the judgment 'Related' was a main effect of relatedness and concurrent task ($BF_{10} = 59.47$). This judgement was used more in the related than unrelated condition ($BF_{inclusion} = 24.01$). There was only weak evidence that this judgment was more frequent under fast pace condition ($BF_{inclusion} = 2.58$). The best model for judgment 'Related or Heard' was a main effect

of relatedness ($BF_{10} = 17.17$). This judgement was used more for related than unrelated words. The best model for ‘Guessing’ judgment was the null ($BF_{10} = 1.00$).

For non-semantic errors, when comparing subjective judgments, the best model was the main effect of subjective judgment ($BF_{10} = 13.05$). Post-hoc analysis revealed that ‘related’ judgement was used less than the other judgements ($BF_{10} = 6.59$; $BF_{10} = 9.27$; $BF_{10} = 4.53 \times 10^3$, for the comparison with ‘guess’, ‘heard or related’ and ‘heard’ judgments, respectively).

The best model for judgment ‘Heard’ was a main effect of age group ($BF_{10} = 66.54$). Children used this judgement more often than adults. The best model for the judgment ‘Related’ was the additive model ($BF_{10} = 7.2$). There was weak evidence for a main effect of concurrent task ($BF_{inclusion} = 1.12$). Children used this judgement more than adults ($BF_{inclusion} = 6.21$). The best model for judgment ‘Related or Heard’ was the additive model ($BF_{inclusion} = 4.76$). There was only weak evidence for a main effect of age and of concurrent task ($BF_{inclusion} = 1.64$; $BF_{inclusion} = 2.89$, respectively) The best model for ‘Guessing’ judgment was the additive model ($BF_{10} = 163.70$). This judgment was used more in the fast pace condition ($BF_{inclusion} = 29.21$) and in children than adults (6.61).

7.3.3 Discussion

As expected, semantically related words at study, which is thought to boost gist memory, increased semantic errors, compared to unrelated words. Moreover, semantic errors increased when WM maintenance mechanisms were impaired. It could suggest that these mechanisms decrease short-term false memories. Nonetheless, as they also decreased other error types, it seems that this impact is not specific of semantic errors. Note also that ‘heard’ judgments were increased under slow pace concurrent task condition. It suggests that those mechanisms preserve detailed strong memory traces.

Besides, even though both age groups produced as many semantic errors, children produced as many semantic as non-semantic errors, while adults produced mostly semantic errors ($BF_{10} = .30$; $BF_{10} = 27.92$, respectively). It shows a difference of error type dominance between age groups. These findings are in line with previous studies showing that semantic errors become the dominant produced error type with age. Finally, as expected, correct recall was mainly associated to ‘heard’ judgments, while semantic errors were distributed across precise detailed memories (i.e. ‘heard), post-hoc correct rejection (i.e., related), gist memories (i.e., ‘heard or related’) and guessing. Besides, the increase of correct recall and semantic errors on related

words seem to be associated with an increase of ‘heard’, ‘related’, and ‘heard or related’ judgments, but not with guessing judgments. It supports the idea that presenting related wordlists leads to semantic errors that are based on fuzzy to precise memory traces rather than on guessing.

7.4 General discussion

In this study, we assessed the role of semantic similarity and WM maintenance mechanisms in false memories in immediate recognition (Exp 1.) and recall (Exp.2) in 9- year-olds and young adults. In both experiments, participants had to complete a complex span task, in which they were shown lists composed equally of semantically related and unrelated words. During the 5.4 second interval between each word-list, participants had to complete a parity judgment task during which WM maintenance mechanisms were either more (fast pace) or less impaired (slow pace). Recognition test in the first experiment allowed the use of the SCR model of the FTT to assess underlying memory traces (i.e., gist and verbatim) to correct and false recognition. Recall in the second experiment was accompanied with a subjective experience scale, to get an appreciation of the phenomenological bases of the memory it was based on.

In the first Experiment, children and adults had similar rates of false recognitions. This rate was however not underpinned by similar memory traces. Indeed, adults had more gist and verbatim memory on related probes. In the second Experiment, adults tended to produce more semantic than non-semantic errors, whereas children produced as many semantic as non-semantic errors. Moreover, false memories were increased by the presentation of semantically related words in both age groups, confirming that being in conditions that foster gist processing favors the likelihood of false memories in WM tasks. Using the SCR model provided an objective measure of gist memory in Experiment 1 and confirmed that gist memory increased on the presentation of related list-words. Moreover, errors were increased by the suppression of maintenance mechanisms. False recognition rates were not impacted by the pace of the concurrent task. In the second experiment, subjective judgment showed that false memories were based equally as much on vivid remembering as on a fuzzy trace and on guessing. One fourth of semantic error recall was also rejected during subjective memory judgment as being a ‘true’ memory. It indicates that all semantic errors are not necessarily based on similar traces. Some are experienced as stronger memories than others. Besides, we found that all three factors: age, WM maintenance mechanisms and word-list relatedness impacted phenomenological experience. Indeed, the percentage of strong remembering in semantic error was higher in adults

than children. There was however, no age difference in fuzzy gist-based judgements. Moreover, we found that the manipulation of gist activation via the relatedness of list words, increased fuzzy gist-based phenomenology. This is consistent with the idea that studying related words increases gist processing, as defended by the FTT. Finally, we found that when WM maintenance mechanisms could be used to maintain information, correct recall was more often associated to strong detailed memories. The rate of semantic errors associated to strong illusory memories, by contrast, decreased. This last point highlights the role of WM maintenance mechanisms in preventing false memories in WM tasks.

The role of gist memory in short-term false memory

One main objective of this study was to test the impact of gist memory activation on short-term false memories via an experimental manipulation. We did so by varying the semantic relatedness of words within studied lists. As expected, semantic similarity increased false memories of both adults and children in both Experiments. It suggests that gist processing increases short-term false memory like it does in LTM tasks (e.g., Brainerd et al. 2020; Howe et al., 2009; McEvoy et al., 1999; Montefinese et al., 2015). By extension, it indicates that when in situations where there is a strong gist memory activation, false memories are more likely to occur, and so, from the first few seconds after stimuli exposure. These findings were supported by the estimation of gist memory with the SCR in Experiment 1, which showed higher rates of gist memory on related probes for semantically similar than dissimilar words. This estimation was more fine grained than raw false recognition scores and allowed us to highlight differences between age groups. Indeed, the increase in gist memory on related probes between semantic related and unrelated words was greater in adults than in children. This difference may be explained by the fact that gist memory is more developed in adults than children. Moreover, in Experiment 2, semantic errors were more often attributed to ‘heard’, ‘related’ and ‘heard or related’ but not to ‘guessing’ judgements in the related than unrelated condition. It means that the increased semantic error rate in the related condition compared to the unrelated condition, might be based on detailed to fuzzier gist-based memory traces rather than on guessing. This is an indirect argument that presenting semantically related words increases gist memory. In sum, our study seems to indicate that short-term false memories are dependent on gist memory activation.

The role of maintenance mechanisms in false memories

In our study, the suppression of WM maintenance mechanisms via a concurrent task increased semantic error in recall (Exp. 2). It suggests that when they can be used, these mechanisms could prevent false memories in immediate recall in WM tasks in children like in adults. Note however, that the suppression of maintenance mechanisms also increased non-semantic errors. In consequence, the increase of semantic errors by the suppression of WM maintenance mechanisms might only be the mirror of the decrease of correct recall when WM maintenance mechanisms are prevented. Therefore, the conclusion that WM maintenance mechanisms have a specific role in false memories could be premature. In fact, in this study, given that refreshing was manipulated through the variation of the pace of the concurrent task, and that in both conditions, the concurrent task asked for a concurrent articulation, it is possible that we mainly measured the impact of refreshing on recall errors. If so, it may be possible that our study, in line with previous other studies (Abadie & Camos, 2019; Abadie et al., in rev.), revealed the nonspecific role of refreshing in short-term memories. Regarding false recognitions (Exp 1.), we found no effect of maintenance mechanisms on false memories. One thought could be that these mechanisms were not used by participants, because recognition tests may be perceived as an effortless type of task, that often relies on passive LTM storage (Uittenhove et al. 2019). However, in the first Experiment, we found that correct recognition was decreased by the introduction of a faster concurrent task. Therefore, it seems that participants did engage in active maintenance during the retention intervals of the task. It suggests that even though participants used maintenance mechanisms, either it did not affect false recognition, or using a recognition test might not be ideal to measure the effect of WM maintenance mechanisms on short-term false memories. In the first case, it would support the existence of a dissociation between recognition and recall. Such distinction has indeed been reported in the past (for a review see Brainerd & Reyna, 2005). However, some studies which assessed the role of WM maintenance mechanisms on false recognition, have shown an increase of false memories when maintenance mechanisms were prevented (Abadie & Camos, 2019; Macé & Caza, 2011). In those studies, the rate of short-term false recognition was higher than in Experiment 1 (around 20-30% under articulatory suppression, against 10%). Therefore, it is likely that in Experiment 1, the rate of false memories might have been too low to reveal the effect of WM maintenance mechanism on false recognition. Or it could also be that we mainly measured the effect of refreshing on false recognition, which may have no specific role in short-term false recognition. Further studies might be necessary to bring further evidence for or against this point and help to settle the matter.

7.5 Conclusion

Overall, we showed that the gist memory activation was responsible for short-term false memories both in adults and in children. Indeed, when gist processing was fostered by the presentation of semantically related words, false memories increased compared to the presentation of unrelated words. The use of the SCR for gist trace estimation in Experiment 1 and the measure of phenomenological experience associated to recall in Experiment 2, provide supplementary arguments for this conclusion. Indeed, gist memory was higher on semantically related than unrelated studied words. The increase of semantic errors on related studied words was associated to strong detailed and fuzzier gist-based judgements rather than to guessing. Moreover, it seems that even though WM maintenance mechanisms may prevent semantic errors, this effect is not specific to this error type, as it also prevents non-semantic errors. Further studies might be needed to bring clarification on the matter.

Summary

This study assesses the role of both LTM traces and maintenance mechanisms in WM in short-term false memories in children and adults. In a complex span task, the activation of LTM gist traces was manipulated via the semantic relatedness of the 6 words presented within lists (i.e., 3 related, 3 unrelated). WM maintenance mechanisms were manipulated via the variation of the pace parity judgment concurrent task (i.e., fast or slow), introduced during the 5 seconds retention intervals, occurring between each word-list presentation. Experiment 1 measured false memories through an immediate recognition test and Experiment 2 through immediate recall. In both tests, false memories were increased by the presentation of semantically related studied words in children and adults, suggesting that gist memory is responsible for false memories in WM tasks. These findings were backed up by gist estimates which were higher in related than unrelated studied words and by phenomenological experience associated to semantic errors which were associated to strong detailed (i.e., 'I heard this word') but also fuzzier memory judgments (i.e., 'I heard this word or a related one'). Moreover, errors in recall were increased by the suppression of WM maintenance mechanisms in both age groups, suggesting that the use of these mechanisms can prevent both semantic and non-semantic errors both in children and adults. In the first Experiment, false memory rates were similar across age groups, but an objective measure of gist and verbatim memory showed that adults had more gist memory than children. The absence of false recognition increases with age could be explained by higher rivaling rates of verbatim memory in older children. In Experiment 2, adults produced predominantly semantic recall errors, whereas children produced as many semantic as non-semantic errors. These findings are in line with the Fuzzy-Trace Theory, showing an increase of gist memory with age.

Résumé

Cette étude s'intéresse au rôle des traces en mémoire à long-terme et des mécanismes de maintien en mémoire de travail dans les faux souvenirs à court-terme chez l'enfant et l'adulte. Nous avons manipulé les traces gist via la manipulation du lien sémantique entre les 6 mots d'une liste, présentés dans une tâche d'empan complexe (i.e., 3 mots reliés, 3 mots non-reliés). Les mécanismes de maintien en mémoire de travail étaient manipulés via la variation de la vitesse de présentation de chiffres dans une tâche de jugement de parité (i.e., rapide ou lente). Celle-ci se déroulait durant l'intervalle de rétention de 5 secondes présenté entre chaque mot à étudier. Dans l'expérience 1, les faux souvenirs étaient mesurés à travers une tâche de reconnaissance immédiate et dans l'Expérience 2 à travers une tâche de rappel immédiat. Dans chacun des tests, les faux souvenirs augmentaient suite à la présentation de mots sémantiquement reliés, chez les enfants et les adultes, suggérant que les traces gist sont responsables des faux souvenirs en tâche de mémoire de travail. Ces résultats ont été appuyés par l'estimation des traces gist qui était plus élevée pour les mots étudiés reliés que non-reliés ainsi que par l'expérience phénoménologique associée aux erreurs sémantiques. En effet, les erreurs sémantiques étaient associées à de forts souvenirs détaillés, mais aussi à des souvenirs plus flous. De plus, la suppression des mécanismes de maintien en mémoire de travail a augmenté les erreurs en rappel. Ceci suggère que l'utilisation des mécanismes de maintien peut prévenir des erreurs chez les enfants et les adultes, sans être spécifique aux erreurs sémantiques. Dans la première expérience, le taux de faux souvenirs était similaire entre les groupes d'âges mais une mesure objective des traces gist ou verbatim a permis de montrer que les adultes avec plus de traces gist que les enfants. L'absence d'augmentation de fausse reconnaissance avec l'âge, pourrait être expliquée par des traces verbatim augmentées également chez les adultes, ce qui s'opposerait à l'augmentation des traces gist. Enfin, dans la deuxième expérience, les adultes produisent des erreurs sémantiques de façon prédominante, alors que les enfants ont produit autant d'erreurs sémantiques que non-sémantiques. Ces résultats s'inscrivent dans la continuité de la théorie des traces floues qui montre une augmentation des traces gist avec l'âge.

General Discussion

Chapter 8. General discussion and perspectives.

8.1 Synthesis and discussion

8.1.1 Synthesis of Results

This thesis addresses several important questions regarding the production of false memories. First, it aimed at assessing whether false memories occurred in WM tasks in children of different ages. Indeed, in classical LTM tasks, self-generated false memories were shown to increase between early childhood and adulthood (Brainerd & Reyna, 2012, for a review). However, we did not know whether such rapid illusion could occur in children, and if the developmental trend observed in LTM tasks would also be observed in WM tasks. Our three studies showed that false memories did occur in children's WM in immediate and in delayed tests. Indeed, children as young as 4 years old produced around 10% of false recognition in the immediate test and 17% in the delayed test (Study 1). In recall tests, 5-year-olds also produced 9% of semantic errors in immediate recall tests and 8-year-olds produced up to 17%. Regarding age differences, our research showed that false recognition did not increase between 4-year-olds and 8-year-olds both in immediate and delayed tests. In Study 3 we also observed no increase in false recognition between 9-year-olds and young adults. However, Study 2 showed an increase of semantic errors in recall between 5- and 8-year-olds in immediate tests. Study 3 also demonstrated that semantic errors were the dominant error type in adults, whereas 9-year-olds in comparison made an equal number of semantic and non-semantic recall errors. Our results suggest that in WM tasks like in LTM tasks, there is an increase of false memories with age, at least in recall tasks, even though this increase is smaller than in LTM tasks.

Second, this thesis aimed at assessing the role of WM maintenance mechanisms in false memories in immediate and delayed tests, in different age groups. These mechanisms develop with age between 5 and 7 years old, and keep on maturing until adulthood (Camos & Barrouillet, 2018, for a review). Hence there is a real interest in comparing different age groups with different abilities in terms of information maintenance in WM to fully understand the implication of these mechanisms in false memories in immediate and delayed tests. In our first study, we showed no impact of our manipulations of WM maintenance mechanisms on false recognition in 4- and 8-year-olds children. Based on these findings, we could have concluded that WM maintenance mechanisms do not impact false memories in children as they do with adults (e.g. Abadie & Camos, 2019). However, there was evidence that the paradigm used in

the first study (i.e., Brown-Peterson with a recognition test) might not have favored the use of WM maintenance mechanisms in children. In the second study, when using a complex span task with an immediate recall test, we found that semantic errors were decreased in older children when WM maintenance mechanisms were prevented by an attentionally demanding concurrent task. More precisely, hindering maintenance mechanisms increased the dominant error type of each age group- i.e.- semantic for 8-year-olds and non-semantic for 4-year-olds. Finally, in our third study, when comparing older children and young adults in recognition and recall tests, we found that WM maintenance mechanisms decreased false memories in both age groups in immediate recall but not in immediate recognition. Besides, this effect was not specific of semantic errors. Non-semantic errors also increased with the suppression of WM maintenance mechanisms. Given that in this study, articulatory suppression was implemented in both conditions (even though it was implemented at a faster pace in one of the two conditions), Study 3 may reflect a non-specific impact of refreshing on recall errors.

Finally, this thesis aimed at understanding the type of memory representations on which false memories in WM tasks are based. False memories in LTM tasks are known to be based on memory traces relating to the content of items (i.e., gist memory). WM is considered by some as a distinct system, separate from LTM. Even though it relies on LTM, we wonder whether similar memory representations underly false memories in WM compared to LTM tasks. To assess this theory, we relied on several indicators. One of them was an experimental manipulation of the semantic relatedness between words in lists studied by participants. This manipulation can be considered as an indirect manipulation of gist memory activation, a memory trace shown to be responsible for false memories in LTM tasks. In Study 3, we showed that short-term false memories both in recall and in recognition were increased in children and in adults when they had to study semantically related words (see Abadie & Camos, 2019 for similar findings in adults). This is the first evidence that false memories in WM tasks may be underpinned by gist memory traces. Furthermore, in all studies, we used the SCR model to obtain an objective measure of gist and verbatim memory traces underlying recognition. We found that false recognitions were based on gist memory from early childhood, and showed that gist memory between young and older children may be the source of qualitative differences, resulting in a different proportion of correct and false recognition. Older children might indeed be able to correctly reject a semantically related distractor more often than younger children, based on gist memory. We also found an increase in gist and verbatim memory with age, between young and older children (Study 2; Study 1, Exp. 2, for gist traces) but also between

older children and young adults (Study 3, Exp. 1). These findings are in line with the FTT. Finally, we collected phenomenological experience associated with recall. We found that phenomenological experience associated with semantic errors was heterogeneous. Indeed, some semantic errors were associated with strong detailed judgements, but also with fuzzier judgements and guessing judgements. These findings were true for young, older children and adults. It suggests that not all false memories are associated with the same memory traces. False memories seem to be mainly underpinned by memory traces (gist and verbatim) rather than guessing. In line with this interpretation, we found that the increase of semantic errors in adults compared to children in Study 3 was associated with an increase of both detailed and fuzzy phenomenological experience and not with guessing. Specific points will now be discussed.

8.1.2 Comparing the production of false memory in different age groups.

In classical LTM tasks using the DRM paradigm, false recognition was shown to increase with age (Brainerd & Reyna, 2012). Hence, we expected that false recognition would increase with age in WM tasks as they would in LTM tasks. In our studies, recall errors became increasingly semantic with age, but we found no increase of false recognition between the young and older children, nor between older children and adults. With post-hoc reflection, this does not seem so surprising after all, given that some studies comparing false recognition in children of different age groups and adults found that the age increase obtained in classical 15 word- DRM list became null when the lists were shortened to 7-8 words (Metzger et al, 2008; Ghetti, et al., 2002; Sugrue & Hayne, 2006; Sugrue, et al., 2009). In these studies, the authors propose a common interpretation to such findings, based on the FTT. According to them, shorter lists of words would create simultaneously smaller gist memory and greater verbatim memory. Indeed, the more words about a common theme there are, the bigger the gist strength score is likely to be. By contrast, as verbatim represents detailed memory traces, and this type of trace is susceptible to interference, the fewer words there are to study, the most chances there are that the verbatim memory of a given word will be preserved. In consequence, the balance of gist and verbatim memory on short lists might be different than that of long-lists. The classical increase of false recognition in adults in 15-word lists likely disappears because even though adults might have more gist than children on short lists, they also likely have more verbatim traces and fewer gist traces on short than long lists. Stronger verbatim memory for the older compared age group would therefore allow for correct rejection of semantic distractors more often on short lists.

Our findings are in line with this interpretation. Indeed, by using the SCR model, we could obtain estimates of gist and verbatim memory. These findings showed that even though rates of false recognition were similar between the compared age groups, we found differences in gist and verbatim memory traces between age groups in Study 2 and 3. Indeed, gist and verbatim memory on related probes were increased in the older age group compared to the younger one in both studies. We also found qualitative differences in the use of gist memory in Study 1 and 2, where gist memory was used most often by 8-year-olds to produce correct rejection of related items, whereas younger children produced proportionally more false recognitions based on gist traces. These qualitative differences likely contributed to the equalization of false memory rate between age groups in our studies. Besides, it seems that when looking at the estimates of verbatim memory on related probes in Studies 2 and 3, they are descriptively greater than verbatim memory estimates found in studies using protocols with longer lists of words, which are usually around 0 (Sthal & Klauer, 2008). A future study could test varying the length of DRM-like lists within a same protocol, and use the SCR model to estimate gist and verbatim memory associated with recognition. It would allow for a direct comparison between short and longer lists and provide an objective measure of gist and verbatim memory with word-list length.

Nonetheless, in recall tests, we found that the older age group (8-year-olds compared to 5-year-olds and adults compared to 9-year-olds) produced an increasing proportion of semantic errors. This seems to indicate that false memories in WM tasks might, like false memories in LTM tasks, develop and increase with age. Several indirect elements also point to the fact that this increase may be underpinned by the development of gist memory. One of them is the measure of gist memory estimates in recognition test. The increase of gist memory in the older age group (study 2 and 3) in these memory tests supports and indirectly suggests that the increase of semantic errors with age is likely due to the development of gist memory with age. Another indicator was the phenomenological experience associated with semantic errors. In Study 3, the increase of semantic errors with age seem to be associated with an increase in detailed and fuzzier memory judgments, and not with guessing. Hence, it seems that the development of semantic errors in WM tasks may be underlied by a development of gist memory traces.

8.1.3 The role of WM maintenance mechanisms in false memories.

WM maintenance mechanism in false recognition.

If we summarize our findings on recognition tests, there is striking evidence that the suppression of WM maintenance mechanisms did not impact false recognition. Starting from Study 1, in which we hypothesized that the suppression of WM maintenance mechanisms would impact false recognition for 8-year-olds, as they were shown to spontaneously use them (e.g., Oftinger & Camos, 2016, 2018). In Study 3 (Exp 1.), we also expected to see the effect of the suppression of maintenance mechanisms on false recognition in adults and in children. Instead, we observed no effect of WM maintenance mechanisms manipulation on false recognition. The associated reason might be different for Study 1 versus Study 2 and 3. For the first study, it was argued in the discussion (Chapter 5) that the manipulation of WM maintenance mechanisms might not have impacted short- and long-term false recognition because children might simply have not engaged in active maintenance of information (see also Abadie & Rousselle, 2023). In Studies 2 and 3, it seems that participants did engage in active maintenance, as shown by the decrease of correct recognition when WM maintenance mechanisms were impaired. However, in Study 2, the manipulation of maintenance mechanisms did not impact delayed recognition in either age group, despite evidence that rehearsal promoted delayed false recognition in participants using such mechanisms (Abadie & Camos, 2019). In Study 3 (Exp. 1), short-term false recognition was also not impacted by our manipulation of WM maintenance mechanisms, despite evidence that rehearsal prevented the occurrence of false recognition in adults (Abadie & Camos, 2019; Atkins et al., 2011; Macé & Caza, 2011).

One argument defended in the discussion of Study 3 was that false recognition rates might have been too low in our study to reveal effects of WM maintenance mechanisms manipulation on false recognition, or even age effects by contrast with recall tasks. This argument added to the one from the first study, lead to think that recognition tests might not be the best suited for measuring the effect of WM maintenance mechanisms on false memories in WM tasks. Indeed, in some cases it does not seem to encourage engaging in active maintenance, and in other cases it might just not be sensitive enough to capture the effect of our manipulation on false recognition. This conclusion is supported by previous work showing that the manipulation of WM maintenance mechanisms leads to smaller effects in recognition than in recall tasks (Allen et al., 2018; Uittenhove et al., 2019). One theoretical explanation is that recognition relies mainly on familiarity (Oberauer, 2008), a process shown to be not or almost not impacted by the manipulation of WM maintenance mechanisms. By contrast, recall would rely mostly on recollection (Malmberg, 2008), a process shown to be impacted by the implementation of a concurrent task (Gardiner & Parkin, 1990; Gruppuso, et al., 1997; Jacoby & Kelley, 1992). A

recent study from Rosselet-Jordan et al. (2022) discusses the implication of such findings and evokes a potential dissociation regarding the retrieval processes underpinning false recognition and false recall.

Nonetheless, even though recognition might not be sensitive enough to capture the effect of WM maintenance mechanisms on false memories, the underlying memory processes were impacted by the manipulation of such mechanisms. In particular, in immediate recognition (Study 3, Exp. 1), verbatim memory on related probes was shown to decrease in adults when maintenance mechanisms were suppressed. In delayed recognition, verbatim memory of related probes decreased under the suppression of maintenance mechanisms, and gist memory increased under similar conditions in 8-year-olds. These findings are in line with Abadie and Camos's (2019) who showed an increase of verbatim memory at short-term when rehearsal could be used, and an increase of gist memory at long-term when refreshing was used. Besides, the fact that these effects were found only in the older tested age group for each study (i.e., 8-year-olds for Study 2 and adults for Study 3) suggests the existence of an interaction between age group and WM maintenance mechanisms. This point will be further discussed in a following section.

The non-specific role of WM maintenance mechanisms in semantic errors.

By contrast with false recognition, semantic errors in immediate recall were shown to be impacted by our manipulation of WM maintenance mechanisms. Indeed, in Study 2, preventing WM maintenance mechanisms decreased correct recall both in young and older children and it increased errors, mostly non-semantic in younger children. In Study 3 (Exp. 2), correct recall also decreased, and semantic error increased by the suppression of WM maintenance mechanisms both in children and in adults. However, non-semantic errors also increased under WM maintenance suppression. It suggests that WM maintenance mechanisms might not have a specific role in false memories but might increase all error types. Nonetheless, in Study 2, we showed that WM maintenance mechanisms impacted mostly the dominant error type of each age group. Hence, 8-year-olds who produced more semantic errors mostly had an increase of this specific error type when WM maintenance mechanisms were hindered. In contrast to Study 3, the second study allowed the use of both rehearsal and refreshing during retention intervals in one condition. In Study 3, even with a slow pace, participants had to identify the parity of the digits aloud. Therefore, this study may have favored the use of refreshing compared to Study 2, which may have favored the use of rehearsal. It might therefore explain why in Study 3 we

did not observe a stronger effect of WM maintenance mechanisms on semantic errors compared to non-semantic errors. Indeed, refreshing was shown to have no specific role in short-term false memories in previous studies (i.e., Abadie & Camos, 2019). Results from Study 3 might therefore only reflect the non-specific impact of refreshing on recall errors.

Nonetheless, it is not the only difference between Study 2 and 3 and our interpretation deserves further investigation. It could be worthwhile to investigate the specific role of rehearsal and refreshing, using distinct experimental manipulations. Indeed, in studies with adult participants, rehearsal was found to decrease short-term false memories (Abadie & Camos, 2019; Abadie et al., in rev.; Atkins et al., 2011; Macé & Caza, 2011), whereas refreshing was shown to increase long-term false memories (Abadie & Camos, 2019). Moreover, rehearsal was found to preserve verbatim traces whereas refreshing was found to preserve gist traces (Abadie & Camos, 2019). This dissociative effect is not so surprising given that the two mechanisms are shown to work independently and to rely on different resources (Camos et al., 2009). Refreshing relies on domain general attentional resources, and rehearsal on a verbal articulatory loop (Barrouillet & Camos, 2021). Additional research could assess the isolated role of each of these mechanisms on false memories, which is what we intended to do in our first study. Indeed, in Experiment 1, we compared a condition preventing articulatory rehearsal with a condition allowing it. In Experiment 2, both maintenance mechanisms were blocked in one condition or were by contrast allowed in another condition. In Experiment 3, the impact of refreshing alone was assessed.

Unfortunately, we did not observe the expected effects of our manipulations on false recognitions. In consequence, in the second and third study, we did a coarse manipulation of WM maintenance mechanisms, to see if they would impact false memories in children. Indeed, before conducting Studies 2 and 3, and based only on Study 1's findings, there was a possibility that WM maintenance mechanisms did not impact false memories in children. However, now that our findings do suggest that WM maintenance mechanisms may impact false memories in children, at least from age 8, it would be interesting to disentangle the role of rehearsal and refreshing by manipulating both mechanisms orthogonally. This would bring further precision regarding the specific or unspecific role of each mechanism in short-term false memories. It would also test the role of each mechanism in immediate and delayed tests. In particular, given that we observed that in the delayed test of Study 2 there was an increase of gist memory under WM maintenance suppression in 8-year-olds. It would be worthwhile to know whether this effect is based on the use of refreshing for maintenance as shown by Abadie and Camos (2019)

in young adults. In immediate recognition (Study 3), we also found an increase of verbatim memory when maintenance mechanisms were not hindered. We should test whether this effect is caused by the use of rehearsal alone.

An interaction between maintenance in working memory and age group.

Given that WM maintenance mechanisms develop with age, we expected that WM maintenance mechanisms would more strongly impact correct recall for the older tested age group in Studies 2 and 3, and semantic errors for the older tested age group in Study 3. One point that was not raised before and that could explain the absence of interaction between age group and the manipulation of WM maintenance mechanisms, could be that in both studies, we adapted the difficulty of the task to the targeted age group. Indeed, even though there is evidence that equalizing the cognitive load of the concurrent task between two age groups does not remove developmental differences in memory performance (e.g., Barrouillet et al., 2009; Gaillard et al., 2011; Gavens & Barrouillet, 2004), it remains possible that by adapting the difficulty of the concurrent task processing, it also balanced the impact of WM maintenance mechanisms on correct and false recall between the compared age groups. In Study 2, we can observe when comparing Experiment 1 and 2 that the suppression of WM maintenance mechanisms impacted semantic errors for both age groups in Experiment 1, whereas it impacted only older children in Experiment 2. Now looking at the performance at the concurrent task between age groups in Experiment 1 and 2, we found that in the first Experiment, there were no differences between age groups, suggesting that the task was perfectly adapted in difficulty to the targeted age group. By contrast, in Experiment 2, the performance at the concurrent task preventing WM maintenance was lower in older than younger children. Similarly, in Study 3 (Exp. 2), adults' performance at the concurrent task was lower than children's. It suggests that when the concurrent task is a little bit harder for the older age group, the expected increased effect of WM maintenance mechanisms in the older age group might be prevented.

Of course, this is post-hoc discussion and there could be different explanations and interpretations of these findings. One explanation could be that lower performance at the concurrent task in the older age group might not be indicative of a harder task, but may simply reflect that the older age group prioritizes the memory task over the concurrent task when compared to the younger age group. Another argument is that even though the expected effect might not have been observed on immediate recall, there is indirect evidence that the impact of WM maintenance mechanisms on underlying LTM traces might be stronger in the older age

groups. Indeed, when looking at gist and verbatim estimates of immediate recognition in Experiment 1 of Study 3, we obtained an increase of verbatim memory when WM maintenance mechanisms can be used, in adults but not in children. We could therefore believe that similar effects would occur in recall on such fine grained processes.

Nonetheless, it would be interesting for future studies to use a WM paradigm in which the difficulty of the concurrent task is not adapted to the tested age group, but is the same for everyone. It would more accurately reflect real life situations, as we are all exposed to the same environment no matter our age. We could compare this condition with a condition in which the concurrent task is perfectly adapted in difficulty to each tested individual. To do so, a titration procedure assessing the processing speed of each participant could precede the experimental procedure.

8.2 Contribution of our findings to general questions

8.2.1 A distinction between false memories in immediate and delayed tests.

In Studies 2 and 3, we used immediate and delayed tests to measure the impact of WM maintenance mechanisms on false memories for two distinct age groups with short and longer delays in WM tasks. One reason for doing this was that previous studies measuring false memories in WM tasks had shown distinctive effects of some factors on immediate and delayed tests. Indeed, Flegal et al. (2014) showed an effect of the LOP during encoding on delayed but not immediate recognition. Deeper semantic processing increased false recognition compared to shallow processing in the delayed test but not in the immediate test. Olszewska et al. (2015) have shown lower rates of immediate false recognition on the presentation of auditive stimuli in contrast with visual presentation; the opposite was found in delayed recognition. Finally, Abadie and Camos (2019) showed that articulatory rehearsal preserved short-term false recognition, whereas attentional refreshing promoted long-term false recognition. These differences were interpreted (Flegal et al., 2014; Olszewska et al., 2015) and were found (Abadie & Camos, 2019) to be associated with a difference in verbatim and gist memory traces available in the immediate and delayed test, depending on the manipulated variable. For the LOP effect, verbatim memory was indeed likely more available at short-term on all study items, no matter the encoding condition. Therefore, it may have prevented false recognition of all studied items by contrast with long-term recognition, for which stronger gist memory created by a deeper LOP may have increased false recognition, as verbatim memory was lower and

could not counter gist traces' impact. For the modality effect, the authors suggested that auditory stimuli created more verbatim memory traces, but that those traces would fade away quickly in time, letting the visual modality take the advantage with time. Finally and most importantly, Abadie and Camos (2019) showed that rehearsal was associated with an increase of verbatim memory at short-term, and that refreshing produced increased rates of gist memory which impacted long-term false memories. Indeed, the rate of verbatim memory being lower at longer delays, it increased the likelihood of delayed false recognition when gist memory was boosted. How do our findings contribute to this line of questioning ?

First, we found that false memories in children as young as 4 years of age were underpinned by gist memory both in immediate and delayed recognition. Study 1 was the only direct comparison between immediate and delayed tests as it included a recognition test at both testing times. In this study, we observed a similar age pattern between both types of tests; both age groups produced similar rates of false recognition. Moreover, the underlying gist and verbatim memory estimates on related probes showed consistent patterns between short- and long-term delays. Indeed, gist memory was higher for older than for younger children in both tests. Unfortunately, in this study, we did not observe the effects of our manipulation of WM maintenance mechanisms response to recognition, as discussed in a previous section, making it difficult to compare to the findings of Abadie and Camos (2019). However, in Study 2, even though the type of test was different between immediate and delayed retrieval, we found that preventing WM maintenance decreased short-term false recall, whereas it did not impact long-term false recognition. More interestingly, our findings on delayed recognition showed an increase of gist memory when WM maintenance was prevented in 8-year-olds. These findings are in line with Abadie & Camos' and suggest indirectly that memory traces underlying false memories in delayed tests might be impacted differently by WM maintenance mechanisms than false memories in immediate tests.

8.2.2 An interaction between LTM and WM?

One big debate in the literature concerns the relationship between WM and LTM. Most models acknowledge that WM and LTM are used in WM tasks (e.g., Logie et al., 2021), and some studies tried to define how both systems influenced each other in WM tasks. In particular, studies tested the interaction between WM maintenance mechanisms and LTM semantic effects. For instance, Abadie and Camos (2018) reported a smaller frequency effect in a WM task favoring the use of attentional refreshing. In the same vein, Loaiza et al. (2015) reported a

smaller lexicality effect when a concurrent task preventing refreshing was introduced. Rose et al. (2014, 2015) reported that when using articulatory rehearsal, the LOP effect did not occur. Campoy et al. (2015) by contrast did not show that increasing the demand of the concurrent task impacted the concreteness effect. The word frequency effect was also found to be independent from the manipulation of articulatory rehearsal (Tehan & Humphreys, 1988), and Camos et al.' study (2019) corroborated these findings by showing that neither the frequency nor the lexicality effect were moderated by the variation of the attentional demand of the concurrent task, nor by the implementation of a concurrent articulation.

Hence there seem to be contrasted evidence as to whether WM maintenance mechanisms interact with LTM effects. In this thesis, the LTM effect of interest was the semantic similarity effect. There are only few studies that looked at whether this semantic similarity effect interacted with WM maintenance mechanisms. One study from Rosselet-Jordan et al. (2022), has shown that the semantic similarity effect was independent from the effect of the variation of the attentional demand of the concurrent task on performance. Performance was decreased by a high attentionally demanding task and by the presentation of semantically unrelated words, but these effects did not interact. This suggested that refreshing is independent from the semantic similarity effect. Abadie and Camos (2019) showed however that using rehearsal suppressed the increase of short-term false recognition on semantically related compared to unrelated lists, observed when rehearsal was hindered by a concurrent task and suggesting that this mechanism moderated LTM activations. In the third study of this thesis, we found that the semantic similarity effect was independent of the effect of WM maintenance mechanism suppression in adults and children. Note however that there are some methodological differences between Abadie and Camos' experiment and Study 3. First, we manipulated the semantic relatedness of words within lists, whereas Abadie and Camos compared pure lists of semantically related and unrelated words. Second, in our study, we did not use a concurrent task that specifically isolated the impact of rehearsal on short-term false memories like Abadie and Camos did. Therefore, it is possible that participants used a combination of rehearsal and refreshing and other strategies such as elaboration on related words to try to maintain the six words per list presented in our study. This could have masked a potential interaction effect between rehearsal and the semantic similarity effect.

Nonetheless, by comparing age groups in which LTM traces and WM maintenance mechanisms are not developed similarly, we showed in Study 2 (Exp 2.), that semantic errors in 5-year-olds were not impacted by the suppression of WM maintenance, whereas they were

in 8-year-olds. This suggests indirectly the existence of an interaction between LTM and maintenance mechanisms in WM, in that gist memory is less developed in younger children, hence it may not be impacted by the manipulation of maintenance mechanisms. The estimation of underlying memory traces to short-term false recognition in Study 3 and to long-term false recognition in Study 2 also provides evidence in favor of this interpretation. Indeed, in Study 2, the effect of WM maintenance mechanisms suppression on semantically related probes occurred only in older children, for which both verbatim and gist memory are more developed. Indeed, verbatim memory decreases under maintenance suppression, and gist memory increases under the same condition in 8-year-olds. Note that similar effects on gist and verbatim memory have been obtained in Abadie and Camos (2019) in adults.

One possible criticism of this argument is that we found no interaction between WM maintenance mechanisms and LTM traces in young children because they simply do not use the mechanisms. However, as we found an effect of WM maintenance mechanisms manipulation on correct recall and on non-semantic errors in this age group, this argument seems appropriate. Moreover, when comparing older children who were capable of using WM maintenance mechanisms to adults in Study 3, we also found that verbatim memory on related probes in immediate recognition was impacted by the manipulation of WM maintenance mechanisms but in the older age group only. Therefore, we could interpret the increasing modulation of LTM traces by maintenance mechanisms manipulation with age as evidence for an interaction between LTM and WM maintenance mechanisms. I remain however cautious about any firm conclusion on the matter, because comparing age groups is not synonymous with comparing differences in LTM activations. Indeed, other processes develop with age, starting from WM maintenance mechanisms. Therefore, comparing younger with older age groups would be as if we manipulated both LTM and WM maintenance mechanisms at the same time.

8.2.3 Which test is best suited to measure false memories in working memory tasks?

In the studies conducted within this thesis, two types of WM tasks were used: Brown-Peterson and Complex span tasks. Moreover, two types of memory tests were also used: recognition and recall. These protocol differences led to differences in the impact of our tested factors on false memories. Indeed, it seems that using recognition tasks did not reveal age differences in false memories, whereas using recall showed an increase of the proportion of

semantic errors with age. Moreover, WM maintenance mechanisms did not impact false recognition, whereas they did impact semantic errors. These findings suggest that recall tasks might be more adapted to reveal the factors involved in false memories in WM tasks, especially in children. This could seem surprising at first given that in LTM tasks, the rate of false memories is bigger in recognition than in recall tasks. In particular in young children for whom the rate of semantic recall error was described by Brainerd and Reyna (2012) at near-floor. In our studies, the rate of semantic recall error in children rivals false recall rates in LTM tasks. By contrast, taking our findings as well as those of other studies measuring false memories in WM tasks, it seems undeniable that false recognition rates are lower in WM tasks than in LTM tasks, especially in children for whom the rate of false recognition is around 10% in our studies. Moreover, semantic recall error rates in WM tasks were not always higher than false recognition rates, as they were found to be in LTM tasks. It suggests that recognition tests might not be sensitive to false memories in WM tasks, like they are in LTM tasks, as discussed in section 8.1.3. One plausible cause could be that, if verbatim memory is more available in WM tasks, when a probe is presented at recognition, it would be easier to reject it or confirm that it was part of the studied list, compared to a recall task where no cue is given at retrieval. By contrast, if LTM tasks foster gist memory and produce lower levels of verbatim, false memories might be facilitated when a related distractor is presented at recognition.

Moreover, in the first study of this thesis, we used a Brown-Peterson paradigm. In the second and third study, we used instead a complex span task. Even though each study had its own specificity and we cannot make a direct comparison between studies, they provide evidence that complex span tasks might be best suited to reveal the impact of WM maintenance mechanisms on false memories in WM tasks. Indeed, in the first study, in which a Brown-Peterson task was used, we found no impact of these mechanisms on false recognition. In the third study, in which a complex span task was used, with a recognition test like in Study 1, we found no impact of WM maintenance mechanisms on false recognition but there was an impact of these mechanisms on correct recognition. This difference with Study 1 is quite important because it suggests very different participant behaviors. Indeed, it suggests that in the first study, children may have not engaged in active maintenance of information, relying on passive LTM (Uittenhove et al., 2019). On the contrary, in Study 3, children seem to have engaged in active maintenance, but it did not impact false recognition. It suggests that complex span tasks might be more adapted to capture the impact of WM maintenance mechanisms on false memories. A recent study also corroborates this point (Langerock et al., 2023). In conclusion,

studies conducted in this thesis seem to indicate that using a combination of complex span task with a recall test might be the most adapted protocol to use to reveal the role of specific factors involved in false memories in WM tasks.

8.3 Perspectives

8.3.1 Generalization of our findings to ecological situations

One question that was addressed regarding false memories using the DRM paradigm was whether they could predict real life false memories. One could wonder whether producing a majority of semantic errors when studying lists of words related in meaning could be representative of the types of false memories that we would produce in ecological situations. Even though studying a list of words comes with a context, could isolated words be representative of a complete contextual scene? In the literature, there is evidence that false memories in the DRM were predictive of real-life false memories. Indeed, Gallo (2010) reviewed several studies showing that DRM false memories were predictive of autobiographical false memories. Moreover, Brainerd and Reyna (2012) reminded us that:

Laws that are routinely applied in courtrooms, clinics, and classrooms were discovered with word-list tasks (e.g., the forgetting function, reminiscence, encoding specificity, massed versus distributed practice, proactive and retroactive interference, short-term memory capacity, serial position curves). If such tasks are irrelevant to memory for complex real-life events, why have they generated general laws of human memory? (p.241).

Therefore, it seems that using word-list memory tasks may be appropriate to predict real life situations. One point to be raised however is that semantic errors in memory tests are more likely to occur when the words in the list are semantically related. If we consider, as the spreading activation theories does, that information in memory is primarily organized in associations, or as FTT does, that memory traces relating to the meaning of a lived event is stored in memory, then we could conclude that the type of false memories in real-life that are most likely to occur would be semantic. Besides, information processed in our environment is contextualized. Items from the same environment are usually related categorically or thematically. Hence, presenting lists of words related in meaning is a way of recreating an artificial environment. Therefore, it seems that using DRM-like lists to measure false memories, could reflect real-life false memories. By extension, this conclusion could be applied to the findings of this thesis. One important aspect of this thesis though, was to test the differences of

false memory production between age group. In LTM DRM tasks, there was evidence that the usual developmental trend occurred when using more ecological material such as visual scenes (Hannigan & Reintz, 2001; Otgaar et al., 2014), sentences (e.g., Paris & Carter, 1973), real lived events (Odegard et al., 2009), and contextual stories (e.g., Dewhurst et al., 2007; Howe & Wilkinson, 2011). However, compared to the classical DRM tasks, this trend was sometimes reduced. Therefore, we should be cautious about generalizing our findings to real-life situations. It could be interesting to conduct future studies using more ecological material in a WM task, to test the effect of age group and of WM maintenance mechanisms on false memories.

One of the interests of generalizing our findings to real-life situations, is to draw conclusions about the behaviors we should adopt in some situations to prevent the occurrence of false memories. In particular in the theoretical part of this thesis, we mentioned two fields in which findings on false memories in WM tasks could be relevant: the field of justice and the field of education.

One societal question was indeed to know whether reports given at different ages during childhood and in adulthood were reliable. Given that semantic errors increased with age, it seems that semantic false memories are more likely to be produced by older than younger children and by adults than children, from the first few seconds after an event occurred. Nonetheless, it does not mean that younger children's testimony would be more reliable. Indeed, younger children compared to older ones and older ones compared to adults produce fewer correct memories and are more prone to other types of errors, such as intrusions and other unrelated errors. One reason for this, evoked in Chapter 2, might be that the immaturity of their executive system could lead to lower source monitoring abilities (Cycowicz et al., 2001; Ruffman et al., 2001; Rybash & Colilla, 1994). In consequence, young children might not produce more reliable testimonies. However they might produce different error types than older children or adults (i.e., intrusions or other unrelated errors). Errors in adults might however be more difficult to identify than in children, because they would be semantically close to a true experience and in consequence be accepted more easily as plausible. By contrast, it might be more easily identifiable if children produce errors because they are more likely to be out of context. This should be taken under consideration when collecting testimonies but also in classroom situations. Indeed, if teachers are aware of the predominant error type of the age group they are teaching to, they might adapt their pedagogic strategies. Teachers may also want to use tasks that favor WM maintenance, so that children can be efficient at maintaining information. Using tasks that encourage or at least grant the use of maintenance mechanisms

would allow to create a precise memory trace and limit erroneous short-term information retrieval. One way to do that would be to avoid double tasking and to give a limited quantity of information at once so it could be maintained. Children would have to be actively engaged in the task and feel the need to maintain information. For instance, if they had to take notes about what is being said, they would have to engage in active maintenance, because they would have to recall what was being said just before.

8.3.2 Are false memories in WM tasks limited to semantic errors?

The evolution of error type with age found in our studies questions the need for investigation of other error types than semantic, especially in young children. As described before, we found that semantic errors became increasingly semantic with age. Younger children relative to older ones and older children relative to adults had more intrusion of words from previous lists. In our studies, low rates of phonological errors were produced. Nonetheless, there is evidence that phonological effects can occur in WM tasks, in particular the phonological similarity effect, which was found to emerge around age 6 (e.g., Allik & Siegel, 1976; Henry et al., 2012, but see Conrad, 1971). Importantly, there was evidence that articulatory rehearsal reinforces phonological characteristics of memory words in WM (Camos et al., 2011; Mora & Camos, 2013). Indeed, Camos et al., (2011) and Mora and Camos, (2013) showed that the phonological similarity effect appeared only when rehearsal could be used to maintain the word list to be learned, but not when rehearsal was prevented during retention times. In LTM tasks, the general dominant error type might be semantic, but in WM tasks the pattern might be different. A recent study from McBride et al. (2019), compared the rate of short-term false recognitions on semantically and phonologically related lists of 6 words. They found that false alarms of phonologically related critical items on phonological lists were greater than those of semantically related critical items on semantic lists. They also found in a second experiment, that when inserting one or two semantically related words into phonologically related lists, the rate of phonological false memories was not impacted. By contrast, the rate of semantic false memory was increased when one or two phonological associates were introduced into the semantically related lists. Moreover, an unpublished study conducted by Coane et al. (in prep.) has done a direct comparison between short- and long-term false memories for both types of lists. Results indicate higher phonological than semantic false memory rates in an immediate test, whereas the opposite was found in a delayed test. These findings support a dissociation view between short and long-term false memories and highlights the importance of phonological false memories at short-term.

In the developmental field, phonological false memories, have not yet been measured in WM tasks, comparing different age groups. The only studies comparing phonological and semantic lists in children of different age groups were conducted using LTM tasks. Most of them have shown that in contrast with semantic false memories, phonological false memories, following the presentation of phonologically related lists of words did not increase with age (Brainerd & Reyna, 2007; Dewhurst et al., 2012; Holliday & Weekes, 2006). Swannell and Dewhurst, (2012), who used a recall task instead of recognition, showed however an increase of phonological errors with age (but see Dewhurst and Robinson, 2004). Therefore, semantic and phonological false memories might elicit different developmental patterns. One perspective for future research would be to assess the occurrence of phonological false memories in WM tasks in different age groups in children, as this error type seem to be of particular saliency at short-term. Children could be presented phonologically similar lists of words. In such task, we could also assess the impact of WM maintenance mechanisms and in particular of articulatory rehearsal on the production of such illusion. Given that the phonological similarity effect occurs from age 6, we expect that the manipulation of rehearsal would impact children beyond that age.

Turning to intrusion error, Atkins et al., (2011) showed evidence that intrusions from previous lists occurred at-short term. They also showed that this error type was facilitated when articulatory rehearsal was prevented. Given that we found in our studies that young children produced many intrusions compared to older children or to young adults, another future study could aim at assessing the evolution of intrusion errors in WM tasks. We could use a paradigm similar to Atkins et al.'s (2011), where recognition probes from previous lists are presented at test. Younger children might produce more of this error type. Adding a manipulation of WM maintenance mechanisms, we could observe an increase of intrusions when rehearsal is prevented, for children using such mechanism.

Conclusion

This thesis contributed to our knowledge on false memories in WM tasks, by adopting a developmental perspective. Three main findings should be remembered from this work. First, it seems that false memories in WM tasks are based on gist memory, like they are in LTM tasks. This finding is true for young, older children and young adults. The absence of age increase in false recognition in WM tasks, in contrast with LTM tasks, is likely explained by a difference of ratio between gist and verbatim memory between the two types of tasks. It may also emerge from qualitative differences in gist memory between age groups, resulting in a greater capacity to rely on gist memory to produce correct rejection of distractors in older children. In recall tests, there seem to be however an increasing dominance of semantic errors with age, which may be underpinned by the development of gist memory.

Second, it seems that WM maintenance mechanisms play a role in false memories at short-term and so at least from age 8. Indeed, it seems when rehearsal and refreshing can be used, immediate semantic errors decrease. However, the question whether this effect is specific to false memories was raised by the findings from Study 3, showing an impact of WM maintenance manipulation on non-semantic errors. We discussed the necessity to disentangle the role of each maintenance mechanisms in children as they may produce different impacts on false memories.

Finally, it seems that gist and verbatim memory might be impacted by the suppression of maintenance mechanisms. Indeed, gist memory was increase in delayed tests under maintenance suppression in Study 2. Verbatim memory was decreased under WM maintenance suppression in Study 3. These findings are in line with Abadie & Camos' (2019) and suggest a dissociation between factors involved in immediate and delayed memory tests. Besides, these effects were found only in the older age group of each study, suggesting that WM maintenance mechanisms impact more older than younger children and adults than children. These findings also imply an interaction between WM and LTM.

Furture studies should aim at testing the generalizibility of our findings in ecological situation and should test the evolution with age of two other error types: intrusions from previous lists, which were mainly found to be produced in younger children, and phonological errors, which were recently shown to be a critical error type in WM tasks.

References

- Abadie, M., & Camos, V. (2018). Attentional refreshing moderates the word frequency effect in immediate and delayed recall tasks: Refreshing moderates the word frequency effect. *Annals of the New York Academy of Sciences*, 1424(1), 127-136. <https://doi.org/10.1111/nyas.13847>
- Abadie, M., & Camos, V. (2019). False memory at short and long term. *Journal of Experimental Psychology: General*, 148(8), 1312-1334. <https://doi.org/10.1037/xge0000526>
- Abadie, M., Gavard, E., & Guillaume, F. (2021). Verbatim and gist memory in aging. *Psychology and Aging*, 36(8), 891. <https://doi.org/10.1037/pag0000635>
- Abadie, M., Guette, C., Troubat, A., & Camos, V. (in rev.). False recall in working memory.
- Abadie, M., & Rousselle, M. (2023). Short-Term Phantom Recollection in 8–10-Year-Olds and Young Adults. *Journal of Intelligence*, 11(4), 67. <https://doi.org/10.3390/jintelligence11040067>
- Abadie, M., Waroquier, L. (2020). On the memory processes underlying conscious deliberation in complex decision making: the role of verbatim and gist memory. *Psychological Research* 84, 1714–1722. <https://doi.org/10.1007/s00426-019-01180-8>
- Abadie, M., Waroquier, L., & Terrier, P. (2013). Gist memory in the unconscious thought effect. *Psychological Science*, 24(7), 1253-1259. <https://doi.org/10.1177/0956797612470958>
- Abadie, M., Waroquier, L., & Terrier, P. (2017). The role of gist and verbatim memory in complex decision making: Explaining the unconscious-thought effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(5), 694-705. <https://doi.org/10.1037/xlm0000336>
- Acheson, D. J., Postle, B. R., & MacDonald, M. C. (2010). The interaction of concreteness and phonological similarity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1), 17. <https://doi.org/10.1037/a0017679>
- Ackil, J. K., & Zaragoza, M. S. (1995). Developmental differences in eyewitness suggestibility and memory for source. *Journal of experimental child psychology*, 60(1), 57-83. <https://doi.org/10.1006/jecp.1995.1031>

- Alexander, K. W., Goodman, G. S., Schaaf, J. M., Edelstein, R. S., Quas, J. A., & Shaver, P. R. (2002). The role of attachment and cognitive inhibition in children's memory and suggestibility for a stressful event. *Journal of Experimental Child Psychology*, 83(4), 262-290. [https://doi.org/10.1016/S0022-0965\(02\)00149-2](https://doi.org/10.1016/S0022-0965(02)00149-2)
- Allen, R. J., Hitch, G. J., & Baddeley, A. D. (2018). Exploring the sentence advantage in working memory: Insights from serial recall and recognition. *Quarterly Journal of Experimental Psychology*, 71(12), 2571-2585. <https://doi.org/10.1177/1747021817746929>
- Allik, J. P., & Siegel, A. W. (1976). The use of the cumulative rehearsal strategy: A developmental study. *Journal of Experimental Child Psychology*, 21(2), 316-327. [https://doi.org/10.1016/0022-0965\(76\)90045-X](https://doi.org/10.1016/0022-0965(76)90045-X)
- Anastasi, J.S., & Rhodes, M.G. (2008). Examining differences in the levels of false memories in children and adults using child-normed lists. *Developmental Psychology*, 44, 889-894. <https://doi.org/10.1037/0012-1649.44.3.889>
- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, 22(3), 261-295. [https://doi.org/10.1016/S0022-5371\(83\)90201-3](https://doi.org/10.1016/S0022-5371(83)90201-3)
- Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, 411,305-309. <http://dx.doi.org/10.1038/35077083>
- Anderson, J. R., & Pirolli, P. L. (1984). Spread of activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(4), 791. <https://doi.org/10.1037/0278-7393.10.4.791>
- Arndt, J. (2006). Distinctive information and false recognition: The contribution of encoding and retrieval factors. *Journal of Memory and Language*, 54(1), 113-130. <https://doi.org/10.1016/j.jml.2005.08.003>
- Arndt, J. (2010). The role of memory activation in creating false memories of encoding context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36 (1), 66-79. <https://doi.org/10.1037/a0017394>
- Arndt, J. (2015). The influence of forward and backward associative strength on false memories for encoding context. *Memory*, 23(7), 1093-1111. <https://doi.org/10.1080/09658211.2014.959527>
- Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global-matching perspective. *Journal of Memory and Language*, 39, 371-391. <https://doi.org/10.1006/jmla.1998.2581>

- Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language*, 48(1), 1-15. [https://doi.org/10.1016/S0749-596X\(02\)00518-1](https://doi.org/10.1016/S0749-596X(02)00518-1)
- Atkins, A. S., Berman, M. G., Reuter-Lorenz, P. A., Lewis, R. L., & Jonides, J. (2011). Resolving semantic and proactive interference in memory over the short-term. *Memory & Cognition*, 39, 806-817. <https://doi.org/10.3758/s13421-011-0072-5>
- Atkins, A., S., & Reuter-Lorenz, P. A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Memory and cognition*, 36(1), 74-81. <https://doi.org/10.3758/mc.36.1.74>
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89-195). Academic press. [https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Audacity team (2016). Audacity (version 2.4.2) [Computer software]. Retrieved from <https://www.audacityteam.org/>
- Baddeley, A.D. (1986). *Working memory*. Oxford, UK: Clarendon Press.
- Baddeley, A. D. (2000). The episodic buffer: a new component of working memory ? *Trends in Cognitive Science*, 4(11), 417-423. [https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/10.1016/S1364-6613(00)01538-2)
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Baddeley, A. D., Hitch, G., & Allen, R. (2021). A multicomponent Model of Working Memory. In R. Logie, V. Camos, & N. Cowan (Eds.), *Working Memory*. Oxford University Press.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior*, 14(6), 575-589. [https://doi.org/10.1016/S0022-5371\(75\)80045-4](https://doi.org/10.1016/S0022-5371(75)80045-4)
- Bardikoff, N., & Sabbagh, M. (2017). The differentiation of executive functioning across development: Insights from developmental cognitive neuroscience. *New perspectives on human development*, 47, 66. In N. Budwig, E. Turiel, & P. D. Zelazo (Eds.) *New perspective on human development* (pp. 27-46). Cambridge UK: Cambridge University Press.
- Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, 133(1), 83-100. <https://doi.org/10.1037/0096-3445.133.1.83>

- Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 570–585. <https://doi.org/10.1037/0278-7393.33.3.570>
- Barrouillet, P., & Camos, V. (2001). Developmental increase in working memory span: Resource sharing or temporal decay?. *Journal of Memory and Language*, *45*(1), 1-20. <https://doi.org/10.1006/jmla.2001.2767>
- Barrouillet, P., & Camos, V. (2015). *Working memory: Loss and reconstruction*.
- Barrouillet, P., & Camos, V. (2021). The time-based resource-sharing model of working memory. In R. H. Logie, V. Camos, & N. Cowan (Eds.), *Working memory: State of the science* (pp. 85–115). Oxford University Press. <https://doi.org/10.1093/oso/9780198842286.003.0004>
- Barrouillet, P., & Camos, V. (2022). La mémoire de travail : Théories, développement et pathologies. Mardaga. ISBN 9782804720926.
- Barrouillet, P., Camos, V., Perruchet, P., & Seron, X. (2004). ADAPT: A Developmental, Asemantic, and Procedural Model for Transcoding From Verbal to Arabic Numerals. *Psychological Review*, *111*(2), 368–394. <https://doi.org/10.1037/0033-295X.111.2.368>
- Barrouillet, P., Dagry, I., & Gauffroy, C. (en prep.). On the sources of working memory development: Manipulating temporal factors (almost) abolishes developmental differences.
- Barrouillet, P., De Paepe, A., & Langerock, N. (2012). Time causes forgetting from working memory. *Psychonomic Bulletin & Review*, *19*, 87-92. <https://doi.org/10.3758/s13423-011-0192-8>
- Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology*, *45*(2), 477-490. <https://psycnet.apa.org/doi/10.1037/a0014615>
- Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review*, *118*(2), 175-192. <https://doi.org/10.1037/a0022324>
- Barrouillet, P., Uittenhove, K., Lucidi, A., & Langerock, N. (2018). On the sources of forgetting in working memory: The test of competing hypotheses. *Quarterly Journal of Experimental Psychology*, *71*(8), 1714-1733. <https://doi.org/10.1080/17470218.2017.1358293>

- Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.
- Bartlett, F. C., & Bartlett, F. C. (1995). *Remembering: A study in experimental and social psychology*. Cambridge university press.
- Baugerud, G. A., Howe, M. L., Magnussen, S., & Melinder, A. (2016). Maltreated and non-maltreated children's true and false memories of neutral and emotional word lists in the Deese/Roediger- McDermott task. *Journal of Experimental Child Psychology, 143*, 102–110. <https://doi.org/10.1016/j.jecp.2015.10.007>
- Beers, S. R., & De Bellis, M. D. (2002). Neuropsychological function in children with maltreatment-related posttraumatic stress disorder. *American journal of Psychiatry, 159*(3), 483-486. <https://doi.org/10.1176/appi.ajp.159.3.483>
- Bekerian, D. A., & Bowers, J. M. (1983). Eyewitness testimony: Were we misled?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(1), 139. <https://doi.org/10.1037/0278-7393.9.1.139>
- Belletier, C., Doherty, J. M., Graham, A. J., Rhodes, S., Cowan, N., Naveh-Benjamin, M., Barrouillet, P., Camos, V., & Logie, R. H. (2022). Strategic adaptation to dual-task in verbal working memory: Potential routes for theory integration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0001106>
- Bergman, E. T., & Roediger, H. L. (1999). Can Bartlett's repeated reproduction experiments be replicated?. *Memory & cognition, 27*(6), 937-947. <https://doi.org/10.3758/BF03201224>
- Bertrand, R., & Camos, V. (2015). The role of attention in preschoolers' working memory. *Cognitive Development, 33*, 14-27. <https://doi.org/10.1016/j.cogdev.2014.10.002>
- Bigelow, J., & Poremba, A. (2014). Achilles' ear? Inferior human short- term and recognition memory in the auditory modality. *PLoS ONE, 9*(2), e89914. <http://dx.doi.org/10.1371/journal.pone.0089914>
- Binet, A. (1900). La suggestibilité. Paris: Schleicher Freres. Véase también: Binet, A.(1905), La science du temoignage. *L'Annee Psychologique, 11*, 128-137.
- Bixter, M. T., & Daniel, F. (2013). Working memory differences in illusory recollection of critical lures. *Memory and Cognition, 41*(5), 716–725. <https://doi.org/10.3758/s13421-013-0293-x>
- Bjorklund, D. F. (2000). *False-memory creation in children and adults: Theory, research, and implications*. Psychology Press.

- Bjorklund, D. F., Bjorklund, B. R., Brown, R. D., & Cassel, W. S. (1998). Children's susceptibility to repeated questions: How misinformation changes children's answers and their minds. *Applied Developmental Science*, 2(2), 99-111. https://doi.org/10.1207/s1532480xads0202_4
- Bjorklund, D. F., Cassel, W. S., Bjorklund, B. R., Brown, R. D., Park, C. L., Ernst, K., & Owen, F. A. (2000). Social demand characteristics in children's and adults' eyewitness memory and suggestibility: The effect of different interviewers on free recall and recognition. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 14(5), 421-433.
- Bjorklund, D. F., & Harnishfeger, K. K. (1990). The resources construct in cognitive development: Diverse sources of evidence and a theory of inefficient inhibition. *Developmental Review*, 10(1), 48-71. [https://doi.org/10.1016/0273-2297\(90\)90004-N](https://doi.org/10.1016/0273-2297(90)90004-N)
- Bjorklund, D. F., & Hock, H. S. (1982). Age differences in the temporal locus of memory organization in children's recall. *Journal of Experimental Child Psychology*, 33(2), 347-362. [https://doi.org/10.1016/0022-0965\(82\)90025-X](https://doi.org/10.1016/0022-0965(82)90025-X)
- Bjorklund, D. F., & Jacobs III, J. W. (1985). Associative and categorical processes in children's memory: The role of automaticity in the development of organization in free recall. *Journal of Experimental Child Psychology*, 39(3), 599-617. [https://doi.org/10.1016/0022-0965\(85\)90059-1](https://doi.org/10.1016/0022-0965(85)90059-1)
- Bjorklund, D. F., & Muir, J. E. (1988). Children's development of free recall memory: Remembering on their own. *Annals of Child Development*, 5(79), 123.
- Bland, C. E., Howe, M. L., & Knott, L. (2016). Discrete emotion-congruent false memories in the DRM paradigm. *Emotion*, 16, 611–619. <http://dx.doi.org/10.1037/emo0000153>
- Blaye, A. (2022). Development of Cognitive Flexibility. *Cognitive Flexibility: The Cornerstone of Learning*, 23-52.
- Blaye, A., & Chevalier, N. (2011). The role of goal representation in preschoolers' flexibility and inhibition. *Journal of Experimental Child Psychology*, 108(3), 469–483. <https://doi.org/10.1016/j.jecp.2010.09.006>
- Bookbinder, S. H., & Brainerd, C. J. (2016). Emotion and false memory: The context–content paradox. *Psychological Bulletin*, 142(12), 1315. <https://doi.org/10.1037/bul0000077>
- Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review*, 1(1), 122-125. <https://doi.org/10.3758/BF03200768>

- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11*, 177–220.
- Brainerd, C. J. (2013). Developmental reversals in false memory: A new look at the reliability of children's evidence. *Current Directions in Psychological Science, 22*(5), 335-341. <https://doi.org/10.1177/0963721413484468>
- Brainerd, C. J., Bialer, D. M., & Chang, M. (2021). Fuzzy-trace theory and false memory: Meta-analysis of conjoint recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. <https://doi.org/10.1037/xlm0001040>
- Brainerd, C. J., Chang, M., & Bialer, D. M. (2020). From association to gist. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(11), 2106. <https://doi.org/10.1037/xlm0000938>
- Brainerd, C. J., Forrest, T. J., Karibian, D., & Reyna, V. F. (2006). Development of the false-memory illusion. *Developmental Psychology, 42*(5), 962–979. <https://doi.org/10.1037/0012-1649.42.5.962>.
- Brainerd, C. J., Holliday, R. E., & Reyna, V. F. (2004). Behavioral measurement of remembering phenomenologies: So simple a child can do it. *Child development, 75*(2), 505-522. <https://doi.org/10.1111/j.1467-8624.2004.00689.x>
- Brainerd, C. J., Holliday, R. E., Reyna, V. F., Yang, Y., & Toglia, M. P. (2010). Developmental reversals in false memory: Effects of emotional valence and arousal. *Journal of Experimental Child Psychology, 107*, 137–154. <http://dx.doi.org/10.1016/j.jecp.2010.04.013>
- Brainerd, C.J., Nakamura, K., Chang, M., & Bialer, D.M. (2019). Verbatim editing: A general model of recollection rejection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(10), 1776–1790. <https://doi.org/10.1037/xlm0000684>
- Brainerd, C. J., Payne, D. G., Wright, R., & Reyna, V. F. (2003). Phantom recall. *Journal of Memory and Language, 48*(3), 445-467. [https://doi.org/10.1016/S0749-596X\(02\)00501-6](https://doi.org/10.1016/S0749-596X(02)00501-6)
- Brainerd, C. J., & Poole, D. A. (1997). Long-term survival of children's false memories: A review. *Learning and Individual Differences, 9*(2), 125-151. [https://doi.org/10.1016/S1041-6080\(97\)90003-0](https://doi.org/10.1016/S1041-6080(97)90003-0)
- Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy-trace theory and children's false memories. *Journal of experimental child psychology, 71*(2), 81-129. <https://doi.org/10.1006/jecp.1998.2464>

- Brainerd, C.J., & Reyna, V.F. (2002a). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*(5), 164–169. <https://doi.org/10.1111/1467-8721.00192>
- Brainerd, C. J., & Reyna, V. F. (2002b). Recollection rejection: How children edit their false memories. *Developmental Psychology*, *38*(1), 156-172. <https://doi.org/10.1037/0012-1649.38.1.156>
- Brainerd, C.J., & Reyna, V. F. (2005). *The science of false memory*. New York, NY: Oxford University Press.
- Brainerd, C.J., & Reyna, V. F. (2007). Explaining developmental reversals in false memory. *Psychological Science*, *18*(5), 442-448. <https://doi.org/10.1016/j.lindif.2007.03.010>
- Brainerd, C.J., & Reyna, V.F. (2012). Reliability of children’s testimony in the era of developmental reversals. *Developmental Review*, *32*(3), 224–267. <https://doi.org/10.1016/j.dr.2012.06.008>
- Brainerd, C.J., & Reyna, V. F. (2015). Fuzzy-trace theory and lifespan cognitive development. *Developmental Review*, *38*, 89-121. <https://doi.org/10.1016/j.dr.2015.07.006>
- Brainerd, C.J., Reyna, V. F., & Ceci, S.J. (2008a). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin*, *134*, 343–382. <https://doi.org/10.1037/0033-2909.134.3.343>
- Brainerd, C.J., Reyna, V.F., Ceci, S.J., & Holliday, R.E. (2008c). Understanding developmental reversals in false memory: Reply to Ghetti (2008) and Howe (2008). *Psychological Bulletin*, *134*(5), 773–777. <https://doi.org/10.1037/0033-2909.134.5.773>
- Brainerd, C.J., Reyna, V.F., & Forrest, T.J. (2002c). Are Young Children Susceptible to the False–Memory Illusion? *Child Development*, *73*(5), 1363–1377. <https://doi.org/10.1111/1467-8624.00477>
- Brainerd, C. J., Reyna, V. F., & Holliday, R.E. (2018). Developmental reversals in false memory: Development is complementary, not compensatory. *Developmental Psychology*, *54*(9), 1773-1784. <https://doi.org/10.1037/dev0000554>
- Brainerd, C. J., Reyna, V. F., & Mojardin, A.H. (1999). Conjoint Recognition. *Psychological Review*, *106*(1), 160-179
- Brainerd, C.J., Reyna, V. F., Wright, R., & Mojardin, A.H. (2003). Recollection Rejection: False-memory editing in children and adults. *Psychological Review*, *110*(4), 762-784. <https://doi.org/10.1037/0033-295X.110.4.762>

- Brainerd, C. J., Stein, L. M., Silveira, R. A., Rohenkohl, G., & Reyna, V. F. (2008b). How does negative emotion cause false memories? *Psychological Science, 19*, 919–925. <http://dx.doi.org/10.1111/j.1467-9280.2008.02177.x>
- Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(2), 307. <https://doi.org/10.1037/0278-7393.27.2.307>
- Brainerd, C. J., Wright, R., Reyna, V. F., & Payne, D. G. (2002d). Dual-retrieval processes in free and associative recall. *Journal of Memory and Language, 46*(1), 120-152. <https://doi.org/10.1006/jmla.2001.2796>
- Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008d). Semantic processing in “associative” false memory. *Psychonomic bulletin & review, 15*, 1035-1053. <https://doi.org/10.3758/PBR.15.6.1035>
- Branch, D. L., Hilgert, L. D., Browne, B. L., & Monetti, D. M. (2007). False recall of critical lures in students with diagnosed learning disabilities. *Psychological Reports, 100*(3), 713–720. <https://doi.org/10.2466/pr0.100.3.713-720>.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of verbal learning and verbal behavior, 11*(6), 717-726. [https://doi.org/10.1016/S0022-5371\(72\)80006-9](https://doi.org/10.1016/S0022-5371(72)80006-9)
- Braver, T.S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences, 16*, 106–113. <https://doi.org/10.1016/j.tics.2011.12.010>
- Braun, K. A., Ellis, R., & Loftus, E. F. (2002). Make my memory: How advertising can change our memories of the past. *Psychology & Marketing, 19*(1), 1-23. <https://doi.org/10.1002/mar.1000>
- Bremner, J. D., & Narayan, M. (1998). The effects of stress on memory and the hippocampus throughout the life cycle: Implications for childhood development and aging. *Development and psychopathology, 10*(4), 871-885. <https://doi.org/10.1017/S0954579498001916>
- Bremner, J. D., Vythilingam, M., Vermetten, E., Southwick, S. M., McGlashan, T., Nazeer, A., ... & Charney, D. S. (2003). MRI and PET study of deficits in hippocampal structure and function in women with childhood sexual abuse and posttraumatic stress disorder. *American journal of psychiatry, 160*(5), 924-932. <https://doi.org/10.1176/appi.ajp.160.5.924>
- Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive psychology, 13*(2), 207-230. [https://doi.org/10.1016/0010-0285\(81\)90008-6](https://doi.org/10.1016/0010-0285(81)90008-6)

- Bright-Paul, A., Jarrold, C., & Wright, D. B. (2008). Theory-of-mind development influences suggestibility and source monitoring. *Developmental Psychology, 44*(4), 1055. <https://doi.org/10.1037/0012-1649.44.4.1055>
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10*, 12-21. <https://doi.org/10.1080/17470215808416249>
- Browning, M. D., Hoffer, B. J., & Dunwiddie, T. V. (1992). Alcohol, memory, and molecules. *Alcohol Research and Health, 16*(4), 280.
- Bruck, M., & Ceci, S. J. (1999). The suggestibility of children's memory. *Annual review of psychology, 50*(1), 419-439.
- Buckey v. County of Los Angeles, 968 F.2d 791 (9th Cir. 1992), 90-56333.
- Budson, A. E., Todman, R. W., Chong, H., Adams, E. H., Kensinger, E. A., Krangel, T. S., & Wright, C. I. (2006). False recognition of emotional word lists in aging and Alzheimer disease. *Cognitive and Behavioral Neurology, 19*, 71–78. <http://dx.doi.org/10.1097/01.wnn.0000213905.49525.d0>
- Camos, V. (2015). Storing verbal information in working memory. *Current directions in Psychological Science, 24*(6), 440-445. <https://doi.org/10.1177/0963721415606630>
- Camos, V. (2017). Domain-specific vs. domain-general maintenance in working memory: reconciliation within the time-based resource sharing model. In H. Ross (Ed.) *Psychology of Learning and Motivation, Vol. 67* (pp. 135–171). Cambridge, MA: Academic Press.
- Camos, V., & Barrouillet, P. (2011). Developmental change in working memory strategies: from passive maintenance to active refreshing. *Developmental psychology, 47*(3), 898. <https://doi.org/10.1037/a0023193>
- Camos, V., & Barrouillet, P. (2018). *Working Memory in Development* (1st ed.). Routledge.
- Camos, V., Barrouillet, P., & Fayol, M. (2001). Does the coordination of verbal and motor information explain the development of counting in children?. *Journal of Experimental Child Psychology, 78*(3), 240-262. <https://doi.org/10.1006/jecp.2000.2570>
- Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language, 61*(3), 457-469. <https://doi.org/10.1016/j.jml.2009.06.002>
- Camos, V., Mora, G., & Oberauer, K. (2011). Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition, 39*, 231–244. <https://doi.org/10.3758/s13421-010-0011-x>

- Camos, V., Mora, G., Oftinger, A., Mariz Elsig, S., Schneider, P., & Vergauwe, E. (2019). Does semantic long-term memory impact refreshing in verbal working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 1664-1682. <https://doi.org/10.1037/xlm0000657>
- Camos, V., & Portrat, S. (2015). The impact of cognitive load on delayed recall. *Psychonomic Bulletin & Review*, *22*, 1029-1034. <https://doi.org/10.3758/s13423-014-0772-5>
- Campoy, G., Castellà, J., Provencio, V., Hitch, G. J., & Baddeley, A. D. (2015). Automatic semantic encoding in verbal short-term memory: Evidence from the concreteness effect. *Quarterly Journal of Experimental Psychology*, *68*(4), 759-778. <https://doi.org/10.1080/17470218.2014.966248>
- Carneiro, P., Albuquerque, P., Fernandez, A. & Esteves, F. (2007). Analyzing false memories in children with associative lists specific for their age. *Child Development*, *78*(4), 1171-1185. <https://doi.org/10.1111/j.1467-8624.2007.01059.x>
- Carneiro, P., & Fernandez, A. (2010). Age differences in the rejection of false memories: The effects of giving warning instructions and slowing the presentation rate. *Journal of Experimental Child Psychology*, *105*(1-2), 81-97. <https://doi.org/10.1016/j.jecp.2009.09.004>
- Carneiro, P., Fernandez, A., & Dias, A. R. (2009). The influence of theme identifiability. *Memory & Cognition*, *37*(2), 115-129. <https://doi.org/10.3758/MC.37.2.115>
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision research*, *51*(13), 1484-1525. <https://doi.org/10.1016/j.visres.2011.04.012>
- Carrasco, M. (2014). Spatial covert attention: Perceptual modulation. In A. C. Nobre & S. Kastner (Eds.), *The Oxford handbook of attention* (pp. 183–230). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199675111.013.004>
- Carrion, V. G., Haas, B. W., Garrett, A., Song, S., & Reiss, A. L. (2010). Reduced hippocampal activity in youth with posttraumatic stress symptoms: An fMRI study. *Journal of Pediatric Psychology*, *41*, 1–11. <https://doi.org/10.1093/jpepsy/jsp112>
- Carrion, V. G., Weems, C. F., & Reiss, A. L. (2007). Stress predicts brain changes in children: A pilot longitudinal study on youth stress, posttraumatic stress, and the hippocampus. *Pediatrics*, *119*, 509–516. <https://doi.org/10.1542/peds.2006-2028>
- Case, R. (1985). A developmentally based approach to the problem of instructional design In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and Learning Skills* (Vol. 2: Research and Open Questions). Routledge.

- Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, 33(3), 386-404. [https://doi.org/10.1016/0022-0965\(82\)90054-6](https://doi.org/10.1016/0022-0965(82)90054-6)
- Cassel, W. S., & Bjorklund, D. F. (1995). Developmental patterns of eyewitness memory and suggestibility. *Law and Human Behavior*, 19(5), 507-532.
- Cassel, W. S., Roebbers, C. E., & Bjorklund, D. F. (1996). Developmental patterns of eyewitness responses to repeated and increasingly suggestive questions. *Journal of Experimental Child Psychology*, 61(2), 116-133.
- Ceci, S. J. & Bruck, M. (1993). Suggestibility of the child witness: a historical review and synthesis. *Psychological Bulletin*, 113(3), 403-439.
- Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony*. American Psychological Association. <https://doi.org/10.1037/10180-000>
- Ceci, S.J., & Bruck, M. (1998). The ontogeny and durability of true and false memories: A fuzzy trace account. *Journal of Experimental Child Psychology*, 71(2), 165-169. <https://doi.org/10.1006/jecp.1998.2468>
- Ceci, S. J., Papierno, P. B., & Kulkofsky, S. (2007). Representational constraints on children's suggestibility. *Psychological Science*, 18(6), 503-509. <https://doi.org/10.1111/j.1467-9280.2007.01930.x>
- Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987). Suggestibility of children's memory: Psycholegal implications. *Journal of Experimental Psychology: General*, 116(1), 38–49. <https://doi.org/10.1037/0096-3445.116.1.38>
- Chae, Y., Goodman, G. S., Eisen, M. L., & Qin, J. (2011). Event memory and suggestibility in abused and neglected children: Trauma-related psychopathology and cognitive functioning. *Journal of Experimental Child Psychology*, 110, 520–538. <https://doi.org/10.1016/j.jecp.2011.05.006>
- Chang, M., & Brainerd, C.J. (2021). Semantic and phonological false memory: A review of theory and data. *Journal of Memory and Language*, 119, 104210. <https://doi.org/10.1016/j.jml.2020.104210>
- Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: The limit in words retained without covert articulation. *The Quarterly Journal of Experimental Psychology*, 62(7), 1420-1429. <https://doi.org/10.1080/17470210802453977>
- Chevalier, N., & Blaye, A. (2008). Cognitive flexibility in preschoolers: The role of representation activation and maintenance: Paper. *Developmental Science*, 11(3), 339–353. <https://doi.org/10.1111/j.1467-7687.2008.00679.x>

- Chevalier, N., & Blaye, A. (2009). Setting goals to switch between tasks: effect of cue transparency on children's cognitive flexibility. *Developmental Psychology, 45*(3), 782–797. <https://doi.org/10.1037/a0015409>
- Chevalier, N., James, T.D., Wiebe, S.A., Nelson, J.M., & Espy, K.A. (2014). Contribution of reactive and proactive control to children's working memory performance: Insight from item recall durations in response sequence planning. *Developmental Psychology, 50*, 1999–2008. <https://doi.org/10.1037/a0036644>
- Choi, H. Y., Kensinger, E. A., & Rajaram, S. (2013). Emotional content enhances true but not false memory for categorized stimuli. *Memory & Cognition, 41*, 403–415. <http://dx.doi.org/10.3758/s13421-012-0269-2>
- Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic bulletin & review, 3*(1), 37-60. <https://doi.org/10.3758/BF03210740>
- Coane, J. H., Huff, M. J., & Hutchison, K. A. (2016). The ironic effect of guessing: Increased false memory for mediated lists in younger and older adults. *Aging, Neuropsychology, and Cognition, 23*(3), 282-303. <https://doi.org/10.1080/13825585.2015.1088506>
- Coane, J. H., McBride, D. M., Huff, M. J., Chang, K., Marsh, E. M., & Smith, K. A. (2021). Manipulations of list type in the DRM paradigm: a review of how structural and conceptual similarity affect false memory. *Frontiers in Psychology, 12*, 668550. <https://doi.org/10.3389/fpsyg.2021.668550>
- Coane, J.H., McBride, D.M., Raulerson, B.A., III, & Jordan, J.S. (2007). False memory in a short-term memory task. *Experimental Psychology, 54*, 62–70. <https://doi.org/10.1027/1618-3169.54.1.62>
- Coane, J. H., McBride, D. M., & Xu, S. (2020). The feature boost in false memory: the roles of monitoring and critical item identifiability. *Memory, 28*(4), 481-493. <https://doi.org/10.1080/09658211.2020.1735445>
- Coane, J. H., Yu, Z., Xu, S., Walsh, C., Feng, Y., and Gray, S. (manuscript in preparation). False memory for associatively and categorically related lists in younger and older adults. In Coane, J. H., McBride, D. M., Huff, M. J., Chang, K., Marsh, E. M., & Smith, K. A. (2021). Manipulations of list type in the DRM paradigm: a review of how structural and conceptual similarity affect false memory. *Frontiers in Psychology, 12*, 668550.
- Cohen, R. L., & Harnick, M. A. (1980). The susceptibility of child witnesses to suggestion. *Law and Human Behavior, 4*(3), 201-210.

- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6), 407. <https://doi.org/10.1037/0033-295X.82.6.407>
- Connolly, D. A., & Price, H. L. (2006). Children's suggestibility for an instance of a repeated event versus a unique event: The effect of degree of association between variable details. *Journal of Experimental Child Psychology*, 93(3), 207-223. <https://doi.org/10.1016/j.jecp.2005.06.004>
- Conrad, R. (1971). The chronology of the development of covert speech in children. *Developmental Psychology*, 5(3), 398. <https://doi.org/10.1037/h0031595>
- Cowan, N., Saults, J. S., & Morey, C. C. (2006). Development of working memory for verbal-spatial associations. *Journal of memory and language*, 55, 274-289.
- Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological review*, 107(2), 261. <https://doi.org/10.1037/0033-295X.107.2.261>
- Cooke, S. F., & Bliss, T. V. (2006). Plasticity in the human central nervous system. *Brain*, 129(7), 1659-1673. <https://doi.org/10.1093/brain/awl082>
- Corson, Y., & Verrier, N. (2007). Emotions and false memories: Valence or arousal? *Psychological Science*, 18, 208-211. <http://dx.doi.org/10.1111/j.1467-9280.2007.01874.x>
- Cotel, S. C., Gallo, D. A., & Seamon, J. G. (2008). Evidence that nonconscious processes are sufficient to produce false memories. *Consciousness and Cognition*, 17(1), 210-218. <https://doi.org/10.1016/j.concog.2007.01.009>
- Coughlin, C., Hembacher, E., Lyons, K. E., & Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Developmental Science* 18(6), 957-971. <https://doi.org/10.1111/desc.12271>
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, 104(2), 163-191. <https://doi.org/10.1037/0033-2909.96.2.341>
- Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language*, 31(5), 668-884. [https://doi.org/10.1016/0749-596X\(92\)90034-U](https://doi.org/10.1016/0749-596X(92)90034-U)
- Cowan, N. (1995). *Attention and memory: An integrated framework*. New York, NY : Oxford University Press.

- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87-114. <https://doi.org/10.1017/S0140525X01003922>
- Cowan, N. (2014). Short-term and working memory in childhood. In P.J. Bauer & R. Fivush (Eds.), *The Wiley handbook on the development of children's memory* (p. 202–229). Wiley Blackwell.
- Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic bulletin & review*, 24, 1158-1170. <https://doi.org/10.3758/s13423-016-1191-6>
- Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological Bulletin*, 145(8), 822–847. <https://doi.org/10.1037/bul0000199>
- Cowan, N., & Alloway, T. (2009). Development of working memory in childhood. In M. L. Courage & N. Cowan (Eds.), *The development of memory in infancy and childhood* (pp. 303–342). Hove, UK: Psychology Press
- Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). *Psychological Science*, 15(9), 634-640. <https://doi.org/10.1111/j.0956-7976.2004.00732.x>
- Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, 51(1), 42-100. <https://doi.org/10.1016/j.cogpsych.2004.12.001>
- Cowan, N., Morey, C. C., & Naveh-Benjamin, M. (2021). An Embedded-Processes approach to Working Memory. In R. Logie, V. Camos, & N. Cowan (Eds.), *Working Memory*. Oxford University Press.
- Cowan, N., Nugent, L. D., Elliott, E. M., Ponomarev, I., & Saults, J. S. (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development*, 70(5), 1082-1097. <https://doi.org/10.1111/1467-8624.00080>
- Cowan, N., Ricker, T. J., Clark, K. M., Hinrichs, G. A., & Glass, B. A. (2015). Knowledge cannot explain the developmental growth of working memory capacity. *Developmental Science*, 18(1), 132-145. <https://doi.org/10.1111/desc.12197>
- Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work?. *Psychological Review*, 119(3), 480. <https://doi.org/10.1037/a0027791>

- Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*(2), 159. <https://doi.org/10.1037/0096-3445.125.2.159>
- Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268. <https://doi.org/10.1037/0096-3445.104.3.268>
- Cunningham, S. J., Brebner, J. L., Quinn, F., & Turk, D. J. (2014). The self-reference effect on memory in early childhood. *Child Development*, *85*(2), 808-823. <https://doi.org/10.1111/cdev.12144>
- Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, Familiarity, and Visual complexity. *Journal of Experimental Child Psychology*, *65*(2), 171-237. <https://doi.org/10.1006/jecp.1996.2356>
- Cycowicz, Y. M., Friedman, D., Snodgrass, J. G., & Duff, M. (2001). Recognition and source memory for pictures in children and adults. *Neuropsychologia*, *39*(3), 255-267. [https://doi.org/10.1016/S0028-3932\(00\)00108-1](https://doi.org/10.1016/S0028-3932(00)00108-1)
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450-466. [https://doi.org/10.1016/S0022-5371\(80\)90312-6](https://doi.org/10.1016/S0022-5371(80)90312-6)
- Davis, & Squire, L. R. (1984). Protein synthesis and memory: a review. *Psychological Bulletin*, *96*(3), 518-559. <https://doi.org/10.1037/0033-2909.96.3.518>
- De Bellis, M. D., Hall, J., Boring, A. M., Frustaci, K., & Moritz, G. (2001). A pilot longitudinal study of hippocampal volume in pediatric maltreatment-related posttraumatic stress disorder. *Biological Psychiatry*, *50*, 305-309. [https://doi.org/10.1016/S0006-3223\(01\)01105-2](https://doi.org/10.1016/S0006-3223(01)01105-2)
- Deese, J. (1959a). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, *5*(3), 305-312. <https://doi.org/10.2466/pr0.1959.5.3.305>
- Deese, J. (1959b). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17-22. <https://doi.org/10.1037/h0046671>
- Dehon, H., Larøi, F., & Van der Linden, M. (2010). Affective valence influences participant's susceptibility to false memories and illusory recollection. *Emotion*, *10*, 627-639. <http://dx.doi.org/10.1037/a0019595>

- Del Prete, F., Mirandola, C., Konishi, M., Cornoldi, C., & Ghetti, S. (2014). Paradoxical effects of warning in the production of children's false memories. *Journal of Cognition and Development, 15*(1), 94–109. <https://doi.org/10.1080/15248372.2012.721036>
- Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89*(1), 63. <https://doi.org/10.1037/0033-2909.89.1.63>
- Dempster, F. N. (1985). Short-term memory development in childhood and adolescence. *Basic Processes in Memory Development: Progress in Cognitive Development Research, 209*–248. https://doi.org/10.1007/978-1-4613-9541-6_5
- Dewhurst, S. A., Howe, M. L., Berry, D. M., & Knott, L. M. (2012). Test-induced priming increases false recognition in older but not younger children. *Journal of Experimental Child Psychology, 111*, 101–107. <http://dx.doi.org/10.1016/j.jecp.2011.08.006>.
- Dewhurst, S.A., Pursglove, R.C., Lewis, C. (2007). Story contexts increase susceptibility to the DRM illusion in 5-years-old. *Developmental Science, 10*, 374 –378. <https://doi.org/10.1111/j.1467-7687.2007.00592.x>
- Dewhurst, S. A., & Robinson, C. A. (2004). False memories in children: Evidence for a shift from phonological to semantic associations. *Psychological Science, 15*(11), 782–786. <https://doi.org/10.1111/j.0956-7976.2004.00756.x>
- DeZazzo, J., & Tully, T. (1995). Dissection of memory formation: from behavioral pharmacology to molecular genetics. *Trends in Neurosciences, 18*(5), 212–218. [https://doi.org/10.1016/0166-2236\(95\)93905-D](https://doi.org/10.1016/0166-2236(95)93905-D)
- Dodson, C. S., Darragh, J., & Williams, A. (2008). Stereotypes and retrieval-provoked illusory source recollections. *Journal of Experimental Psychology: Learning, Memory & Cognition, 34*, 460–477. <https://doi.org/10.1037/0278-7393.34.3.460>
- Dodson, C. S., & Hege, A. C. (2005). Speeded retrieval abolishes the false-memory suppression effect: Evidence for the distinctiveness heuristic. *Psychonomic Bulletin & Review, 12*(4), 726–731.
- Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 71*, 808 – 816.
- Eberle, P., & Eberle, S. (1993). *The Abuse of Innocence: The McMartin Preschool Trial* (p. 243-256). Prometheus Books.

- Eisen, M. L., Goodman, G. S., Qin, J., Davis, S., & Crayton, J. (2007). Maltreated children's memory: Accuracy, suggestibility, and psychopathology. *Developmental Psychology, 43*, 1275–1294. <https://doi.org/10.1037/0012-1649.43.6.1275>
- Eisen, M. L., Qin, J., Goodman, G. S., & Davis, S. L. (2002). Memory and suggestibility in maltreated children: Age, stress arousal, dissociation, and psychopathology. *Journal of Experimental Child Psychology, 83*(3), 167-212. [https://doi.org/10.1016/S0022-0965\(02\)00126-1](https://doi.org/10.1016/S0022-0965(02)00126-1)
- Elliott, E. M., Morey, C. C., AuBuchon, A. M., Cowan, N., Jarrold, C., Adams, E. J., ... & Voracek, M. (2021). Multilab direct replication of Flavell, Beach, and Chinsky (1966): Spontaneous verbal rehearsal in a memory task as a function of age. *Advances in Methods and Practices in Psychological Science, 4*(2), 1-20. <https://doi.org/10.1177/25152459211018187>
- El Sharkawy, J., Groth, K., Vetter, C., Beraldi, A., & Fast, K. (2008). False memories of emotional and neutral words. *Behavioural Neurology, 19*, 7–11. <http://dx.doi.org/10.1155/2008/587239>
- Emery, L., Hess, T. M., & Elliot, T. (2012). The illusion of the positive: The impact of natural and induced mood on older adults' false recall. *Aging, Neuropsychology, and Cognition, 19*, 677– 698. <http://dx.doi.org/10.1080/13825585.2011.645012>
- E-Prime team (2016). E-Prime (version 3.0) [Computer software]. Retrieved from <https://pstnet.com/>
- E-Prime 3.0 software (version 3.0.3.82, 2018). Retrieved from: <https://pstnet.com/>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fazio, L. K., & Marsh, E. J. (2008). Older, not younger, children learn more false facts from stories. *Cognition, 106*(2), 1081-1089. <https://doi.org/10.1016/j.cognition.2007.04.012>
- Ferguson, A. N., Bowey, J. A., & Tilley, A. (2002). The association between auditory memory span and speech rate in children from kindergarten to sixth grade. *Journal of Experimental Child Psychology, 81*(2), 141-156. <https://doi.org/10.1006/jecp.2001.2646>
- Fernández-Dols, J. M., Carrera, P., Barchard, K. A., & Gacitua, M. (2008). False recognition of facial expressions of emotion: causes and implications. *Emotion, 8*(4), 530. <https://doi.org/10.1037/a0012724>
- Finley, J. R., Sungkhasettee, V. W., Roediger III, H. L., & Balota, D. A. (2017). Relative contributions of semantic and phonological associates to over-additive false recall in

- hybrid DRM lists. *Journal of Memory and Language*, 93, 154-168.
<https://doi.org/10.1016/j.jml.2016.07.006>
- Fisher, A. V., & Sloutsky, V. M. (2005). When induction meets memory: Evidence for gradual transition from similarity-based to category-based induction. *Child development*, 76(3), 583-597. <https://doi.org/10.1111/j.1467-8624.2005.00865.x>
- Fitamen, C., Blaye, A., & Camos, V. (2019). The role of goal cueing in kindergarteners' working memory. *Journal of Experimental Child Psychology*, 187, 104666. <https://doi.org/10.1016/j.jecp.2019.104666>
- Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development*, 283-299. <https://doi.org/10.2307/1126804>
- Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2010). False memories seconds later: The rapid and compelling onset of illusory recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1331–1338. <https://doi.org/10.1037/a0019903>
- Flegal, K. E., and Reuter-Lorenz, P. A. (2014). Get the gist? The effects of processing depth on false recognition in short-term and long-term memory. *Memory and Cognition*, 42(5), 701-711. <https://doi.org/10.3758/s13421-013-0391-9>
- Flin, R., Boon, J., Knox, A., & Bull, R. (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology*, 83(3), 323-336. <https://doi.org/10.1111/j.2044-8295.1992.tb02444.x>
- Foley, M. A., Durso, F. T., Wilder, A., & Friedman, R. (1991). Developmental comparisons of explicit versus implicit imagery and reality monitoring. *Journal of Experimental Child Psychology*, 51(1), 1-13. [https://doi.org/10.1016/0022-0965\(91\)90074-3](https://doi.org/10.1016/0022-0965(91)90074-3)
- Foley, M. A., & Johnson, M. K. (1985). Confusions between memories for performed and imagined actions: A developmental comparison. *Child development*, 1145-1155. <https://doi.org/10.2307/1130229>
- Foley, M. A., Johnson, M. K., & Raye, C. L. (1983). Age-related changes in confusion between memories for thoughts and memories for speech. *Child development*, 51-60. <https://doi.org/10.2307/1129860>
- Foley, M. A., Passalacqua, C., & Ratner, H. H. (1993). Appropriating the actions of another: Implications for children's memory and learning. *Cognitive Development*, 8(4), 373-401. [https://doi.org/10.1016/S0885-2014\(05\)80001-2](https://doi.org/10.1016/S0885-2014(05)80001-2)
- Frenda, S. J., Knowles, E. D., Saletan, W., & Loftus, E. F. (2013). False memories of fabricated political events. *Journal of Experimental Social Psychology*, 49(2), 280-286. <https://doi.org/10.1016/j.jesp.2012.10.013>

- Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current Issues and Advances in Misinformation Research. *Current Directions in Psychological Science*, 20(1), 20–23. <https://doi.org/10.1177/0963721410396620>
- Frick, R. W. (1988). Issues of representation and limited capacity in the visuospatial sketchpad. *British Journal of Psychology*, 79(3), 289–308. <https://doi.org/10.1111/j.2044-8295.1988.tb02289.x>
- Frick, R. W. (1988a). Issues of representation and limited capacity in the auditory short-term store. *British Journal of Psychology*, 79(2), 213–240. <https://doi.org/10.1111/j.2044-8295.1988.tb02284.x>
- Frick, R. W. (1988b). Issues of representation and limited capacity in the visuospatial sketchpad. *British Journal of Psychology*, 79(3), 289–308. <https://doi.org/10.1111/j.2044-8295.1988.tb02289.x>
- Gaillard, V., Barrouillet, P., Jarrold, C., & Camos, V. (2011). Developmental differences in working memory: Where do they come from?. *Journal of experimental child psychology*, 110(3), 469–479. <https://doi.org/10.1016/j.jecp.2011.05.004>
- Gallo, D. A. (2006). Associative illusions of memory. New York, NY: Psychology Press.
- Gerkens, D. R., & Smith, S. M. (2004). Effects of perceptual modality on verbatim and gist memory. *Psychonomic Bulletin & Review*, 11(1), 143–140.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & cognition*, 38, 833–848. <https://doi.org/10.3758/MC.38.7.833>
- Gallo, D. A., Foster, K. T., & Johnson, E. L. (2009). Elevated false recollection of emotional pictures in young and older adults. *Psychology and Aging*, 24, 981–988. <http://dx.doi.org/10.1037/a0017545>
- Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review*, 40, 271–276. <https://doi.org/10.3758/BF03209405>
- Gallo, D. A., & Roediger, H. L. (2003). The effects of associations and aging on illusory recollection. *Memory & Cognition*, 31(7), 1036–1044. <https://doi.org/10.3758/BF03196124>
- Gallo, D. A., & Seamon, J. G. (2004). Are nonconscious processes sufficient to produce false memories?. *Consciousness and Cognition*, 13(1), 158–168. <https://doi.org/10.1016/j.concog.2003.09.001>
- Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition*, 18, 579–583. <http://dx.doi.org/10.3758/BF03197100>

- Gathercole, S. E., & Adams, A. M. (1993). Phonological working memory in very young children. *Developmental psychology*, 29(4), 770. <https://doi.org/10.1037/0012-1649.29.4.770>
- Gathercole, S., Pickering, S., & Ambridge, B. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40, 177–190. <https://doi.org/10.1037/0012-1649.40.2.177>
- Gavens, N., & Barrouillet, P. (2004). Delays of retention, processing efficiency, and attentional resources in working memory span development. *Journal of Memory and Language*, 51(4), 644-657. <https://doi.org/10.1016/j.jml.2004.06.009>
- Geraci, L., & McCabe, D. P. (2006). Examining the basis for illusory recollection: The role of remember/know instructions. *Psychonomic Bulletin & Review*, 13(3), 466-473.
- Ghetti, S. (2008). Rejection of false events in childhood: A metamemory account. *Current Directions in Psychological Science*, 17(1), 16-20. <https://doi.org/10.1111/j.1467-8721.2008.00540.x>
- Ghetti, S., & Alexander, K. W. (2004). “If it happened, I would remember it”: Strategic use of event memorability in the rejection of false autobiographical events. *Child development*, 75(2), 542-561. <https://doi.org/10.1111/j.1467-8624.2004.00692.x>
- Ghetti, S., & Angelini, L. (2008). The Development of Recollection and Familiarity in Childhood and Adolescence: Evidence From the Dual-Process Signal Detection Model. *Child Development*, 79(2), 339-358. <https://doi.org/10.1111/j.1467-8624.2007.01129.x>
- Ghetti, S., Lyons, K. E., Lazzarin, F., & Cornoldi, C. (2008). The development of metamemory monitoring during retrieval: The case of memory strength and memory absence. *Journal of experimental child psychology*, 99(3), 157-181. <https://doi.org/10.1016/j.jecp.2007.11.001>
- Ghetti, S., Qin, J., & Goodman, G.S. (2002). False memories in children and adults: age, distinctiveness, and subjective experience. *Developmental Psychology*, 38(5), 705-718. <https://doi.org/10.1037//0012-1649.38.5.705>
- Goodman, G. S., Ogle, C. M., Block, S. D., Harris, L. S., Larson, R. P., Augusti, E. M., . . . Urquiza, A. (2011). False memory for trauma-related Deese-Roediger-McDermott lists in adolescents and adults with histories of child sexual abuse. *Development and Psychopathology*, 23, 423–438. <http://dx.doi.org/10.1017/S0954579411000150>
- Goodman, G. S., Quas, J. A., & Ogle, C. M. (2010). Child maltreatment and memory. *Annual review of psychology*, 61, 325-351.

- Greene, N. R., & Naveh-Benjamin, M. (2020). A specificity principle of memory: Evidence from aging and associative memory. *Psychological Science, 31*(3), 316–331. <https://doi.org/10.1177/0956797620901760>
- Gruber, O. (2001). Effects of domain-specific interference on brain activation associated with verbal working memory task performance. *Cerebral Cortex, 11*(11), 1047–1055. <https://doi.org/10.1093/cercor/11.11.1047>
- Gruppuso, V., Lindsay, D. S., & Kelley, C. M. (1997). The process- dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 259–278. <http://dx.doi.org/10.1037/0278-7393.23.2.259>
- Halford, G. S., Wilson, W. H., Andrews, G., & Phillips, S. (2014). *Categorizing cognition: Toward conceptual coherence in the foundations of psychology*. MIT Press.
- Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*(6), 803–831. <https://doi.org/10.1017/S0140525X98001769>
- Hall, J. W. (1969). Errors in word recognition and discrimination by children of two age levels. *Journal of Educational Psychology, 60*(2), 144. <https://doi.org/10.1037/h0027008>
- Hannigan, S. L., & Tippens Reinitz, M. (2001). A demonstration and comparison of two types of inference-based memory errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(4), 931–940. <https://doi.org/10.1037/0278-7393.27.4.931>
- Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional responses: effects of a neocortical network on the limbic system. *Neuroreport, 11*(1), 43–48.
- Hebb, D. O. (1949). The first stage of perception: growth of the assembly. *The Organization of Behavior, 4*(60), 78–60.
- Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods, 50*(1), 264–284. <https://doi.org/10.3758/s13428-017-0869-7>
- Hedger, N., Adams, W. J., & Garner, M. (2015). Fearful faces have a sensory advantage in the competition for awareness. *Journal of Experimental Psychology: Human Perception and Performance, 41*(6), 1748.

- Hembacher, E., & Ghetti, S. (2014). Don't look at my answer: subjective uncertainty underlies preschoolers' exclusion of their least accurate memories. *Psychological Science*, 25(9), 1768-1776. <https://doi.org/10.1177/0956797614542273>
- Henry, L. A. (2012). *The development of working memory in children*. London: Sage.
- Henry, L.A., Messer, D., Luger-Klein, S., & Crane, L. (2012). Phonological, visual and semantic coding strategies and children's short-term picture memory span. *Quarterly Journal of Experimental Psychology*, 65, 2033-2053. <https://doi.org/10.1080/17470218.2012.672997>
- Hicks, J. L., & Starns, J. J. (2006). The roles of associative strength and source memorability in the contextualization of false memory. *Journal of Memory and Language*, 54(1), 39-53. <https://doi.org/10.1016/j.jml.2005.09.004>
- Hintzman, D. L. (1976). Repetition and memory. *Psychology of learning and motivation*, 10, 47-91. [https://doi.org/10.1016/S0079-7421\(08\)60464-8](https://doi.org/10.1016/S0079-7421(08)60464-8)
- Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, 16(2), 96-101. <https://doi.org/10.3758/BF03202365>
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological review*, 95(4), 528. <https://doi.org/10.1037/0033-295X.95.4.528>
- Hitch, G. J., Halliday, M. S., Dodd, A., & Littler, J. E. (1989). Development of rehearsal in short-term memory: Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology*, 7(4), 347-362. <https://doi.org/10.1111/j.2044-835X.1989.tb00811.x>
- Hitch, G. J., Halliday, S., Schaafstal, A. M., & Schraagen, J. M. C. (1988). Visual working memory in young children. *Memory & cognition*, 16, 120-132. <https://doi.org/10.3758/BF03213479>
- Hitch, G. J., Woodin, M. E., & Baker, S. (1989). Visual and phonological components of working memory in children. *Memory & cognition*, 17, 175-185. <https://doi.org/10.3758/BF03197067>
- Holliday, R.E., Brainerd, C.J., & Reyna, V.F. (2011). Developmental reversals in false memory: Now you see them, now you don't! *Developmental Psychology*, 47, 442-449. <https://doi.org/10.1037/a0021058>
- Holliday, R.E. Reyna, V.F., & Brainerd, C.J. (2008). Recall of details never experienced: Effects of age, repetition, and semantic cues. *Cognitive Development*, 23, 67-78.

- Holliday, R. E., & Weekes, B. S. (2006). Dissociated developmental trajectories for semantic and phonological false memories. *Memory, 14*(5), 624-636. <https://doi.org/10.1080/09658210600736525>
- Howe, M. L. (2000). *The fate of early memories: Developmental science and the retention of childhood experiences*. American Psychological Association. <https://doi.org/10.1037/10369-000>
- Howe, M. L. (2005). Children (but not adults) can inhibit false memories. *Psychological Science, 16*, 927-931. <http://dx.doi.org/10.1111/j.1467-9280.2005.01638.x>.
- Howe, M. L. (2006). Developmentally invariant dissociations in children's true and false memories: Not all relatedness is created equal. *Child Development, 77*(4), 1112-1123. <https://doi.org/10.1111/j.1467-8624.2006.00922.x>
- Howe, M. L. (2007). Children's emotional false memories. *Psychological Science, 18*, 856 – 860. <http://dx.doi.org/10.1111/j.1467-9280.2007.01991.x>
- Howe, M. L. (2008). Visual distinctiveness and the development of children's false memories. *Child Development, 79*(1), 65-79. <https://doi.org/10.1111/j.1467-8624.2007.01111.x>
- Howe, M. L., Candel, I., Otgaar, H., Malone, C., & Wimmer, M. C. (2010). Valence and the development of immediate and long-term false memory illusions. *Memory, 18*, 58 –75. <http://dx.doi.org/10.1080/09658210903476514>
- Howe, M. L., Cicchetti, D., Toth, S. L., & Cerrito, B. M. (2004). True and false memories in maltreated children. *Child development, 75*(5), 1402-1417. <https://doi.org/10.1111/j.1467-8624.2004.00748.x>
- Howe, M. L., Gagnon, N., & Thouas, L. (2008). Development of false memories in bilingual children and adults. *Journal of Memory and Language, 58*(3), 669-681. <https://doi.org/10.1016/j.jml.2007.09.001>
- Howe, M. L., Toth, S., & Cicchetti, D. (2006). Memory and developmental psychopathology. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology. Vol. 2: Developmental neuroscience* (2nd ed., pp. 629-655). New York, NY: Wiley.
- Howe, M. L., Toth, S. L., & Cicchetti, D. (2011). Can maltreated children inhibit true and false memories for emotional information? *Child Development, 82*, 967-981. <https://doi.org/10.1111/j.1467-8624.2011.01585.x>
- Howe, M.L. & Wilkinson, S. (2011). Using story contexts to bias children's true and false memories. *Journal of Experimental Child Psychology, 108*(1), 77-95. <https://doi.org/10.1016/j.jecp.2010.06.009>

- Howe, M. L., Wimmer, M. C., Gagnon, N., & Plumpton, S. (2009). An associative-activation theory of children's and adults' memory illusions. *Journal of Memory and Language*, *60*, 229-251. <https://doi.org/10.1016/j.jml.2008.10.002>
- Huff, M. J., Bodner, G. E., & Fawcett, J. M. (2015). Effects of distinctive encoding on correct and false memory: A meta-analytic review of costs and benefits and their origins in the DRM paradigm. *Psychonomic Bulletin & Review*, *22*, 349-365. <https://doi.org/10.3758/s13423-014-0648-8>
- Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*(6), 685-701. [https://doi.org/10.1016/0749-596X\(91\)90032-F](https://doi.org/10.1016/0749-596X(91)90032-F)
- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1217. <https://doi.org/10.1037/0278-7393.23.5.1217>
- Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*, 785-813. <https://doi.org/10.3758/BF03196544>
- Hutchison, K. A., & Balota, D. A. (2005). Decoupling semantic and associative information in false memories: Explorations with semantically ambiguous and unambiguous critical lures. *Journal of Memory and Language*, *52*(1), 1-28. <https://doi.org/10.1016/j.jml.2004.08.003>
- Huttenlocher, P. R. (1979). Synaptic density in human frontal cortex-developmental changes and effects of aging. *Brain Research*, *163*(2), 195-205. [https://doi.org/10.1016/0006-8993\(79\)90349-4](https://doi.org/10.1016/0006-8993(79)90349-4)
- Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic bulletin & review*, *4*(4), 577-581.
- Jacoby, L. L., & Kelley, C. (1992). Unconscious influences of memory: Dissociations and automaticity. In D. Milner & M. Rugg (Eds.), *The neuropsychology of consciousness* (pp. 201-233). San Diego, CA: Academic Press. <http://dx.doi.org/10.1016/B978-0-12-498045-7.50015-3>
- Jarrold, C., Cowan, N., Hewes, A. K., & Riby, D. M. (2004). Speech timing and verbal short-term memory: Evidence for contrasting deficits in Down syndrome and Williams syndrome. *Journal of Memory and Language*, *51*(3), 365-380. <https://doi.org/10.1016/j.jml.2004.06.007>

- JASP Team (2022). JASP (Version 0.16.4) [Computer software].
- Johnson, M. K., Bransford, J. D., & Solomon, S. K. (1973). Memory for tacit implications of sentences. *Journal of experimental psychology*, 98(1), 203. <https://doi.org/10.1037/h0034290>
- Johnson, M. K., & Foley, M. A. (1984). Differentiating fact from fantasy: The reliability of children's memory. *Journal of social issues*, 40(2), 33-50. <https://doi.org/10.1111/j.1540-4560.1984.tb01092.x>
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological bulletin*, 114(1), 3.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological review*, 88(1), 67. <https://doi.org/10.1037/0033-295X.88.1.67>
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122-149. <https://doi.org/10.1037/0033-295X.99.1.122>
- Kail, R. (2002). Developmental change in proactive interference. *Child Development*, 73(6), 1703-1714. <https://doi.org/10.1111/1467-8624.00500>
- Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90, 773–795. <http://dx.doi.org/10.1080/01621459.1995.10476572>
- 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 And Human Behavior*, 34(1), 3–38. <https://doi.org/10.1007/s10979-009-9188-6>
- Khanna, M. M., & Cortese, M. J. (2009). Children and adults are differentially affected by presentation modality in the DRM paradigm. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 23(6), 859-877. <https://doi.org/10.1002/acp.1519>
- Klein, S. B. (2012). Self, memory, and the self-Reference effect: An examination of conceptual and methodological issues. *Personality and Social Psychology Review*, 16, 283–300. <https://doi.org/10.1177/1088868311434214>
- Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, 75(1), 70-98. <https://doi.org/10.1007/s11336-009-9141-0>
- Knott, L. M., Howe, M. L., Wimmer, M. C., & Dewhurst, S. A. (2011). The development of automatic and controlled inhibitory retrieval processes in true and false recall. *Journal of Experimental Child Psychology*, 109(1), 91-108. <https://doi.org/10.1016/j.jecp.2011.01.001>

- Knott, L. M., & Thorley, C. (2014). Mood-congruent false memories persist over time. *Cognition and Emotion, 28*, 903–912. <http://dx.doi.org/10.1080/02699931.2013.860016>
- Knott, L. M., Threadgold, E., & Howe, M. L. (2014). Negative mood state impairs false memory priming when problem-solving. *Journal of Cognitive Psychology, 26*, 580–587. <http://dx.doi.org/10.1080/20445911.2014.922091>
- Koustaal, W., Reddy, C., Jackson, E. M., Prince, S., Cendan, D. L., & Schacter, D. L. (2003). False recognition of abstract versus common objects in older and younger adults: testing the semantic categorization account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 499. <https://doi.org/10.1037/0278-7393.29.4.499>
- Kowialiewski, B., & Majerus, S. (2018). The non-strategic nature of linguistic long-term memory effects in verbal short-term memory. *Journal of Memory and Language, 101*, 64-83. <https://doi.org/10.1016/j.jml.2018.03.005>
- Kulkofsky, S., & Klemfuss, J. Z. (2008). What the stories children tell can tell about their memory: narrative skill and young children's suggestibility. *Developmental Psychology, 44*(5), 1442. <https://doi.org/10.1037/a0012849>
- Lampinen, J. M., Leding, J. K., Reed, K. B., & Odegard, T. N. (2006). Global gist extraction in children and adults. *Memory, 14*, 952–964. <http://dx.doi.org/10.1080/09658210601008957>.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review, 104*(2), 211. <https://doi.org/10.1037/0033-295X.104.2.211>
- Langerock, N., Oberauer, K., Throm, E. V., & Vergauwe, E. (2023). The cognitive load effect in working memory: Refreshing the empirical landscape, removing outdated explanations.
- Langerock N., & Vergauwe, E. (2022). A robust cognitive load effect in 8 to 12 year old children? [Poster presentation]. https://www.researchgate.net/publication/359210154_A_robust_cognitive_load_effect_in_8_to_12_year_old_children
- Langerock, N., Vergauwe, E., & Barrouillet, P. (2014). The maintenance of cross-domain associations in the episodic buffer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(4), 1096. <https://doi.org/10.1037/a0035783>
- Langerock, N., Vergauwe, E., Throm, E., & Oberauer, K. (2020, September 1-3). The effect of cognitive load in working memory tasks: Limited to the complex span paradigm? [Conference session]. 10th European Working Memory Symposium, Cardiff, Wales.

https://www.youtube.com/watch?v=4gwzJE5qZsw&list=PLbongzUA9knd6D4yfpa4_11gtKNfXQ4Oa&index=4

- Lépine, R., Bernardin, S., & Barrouillet, P. (2005). Attention switching and working memory spans. *European Journal of Cognitive Psychology, 17*(3), 329-345. <https://doi.org/10.1080/09541440440000014>
- Lindsay, D. S., & Johnson, M. K. (1987). Reality monitoring and suggestibility: Children's ability to discriminate among memories from different sources. *Children's eyewitness memory, 92-121*.
- Lindsay, D. S., & Johnson, M. K. (1991). Recognition memory and source monitoring. *Bulletin of the Psychonomic Society, 29*(2), 203-205. <https://doi.org/10.3758/BF03335235>
- Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology, 52*(3), 297-318. [https://doi.org/10.1016/0022-0965\(91\)90065-Z](https://doi.org/10.1016/0022-0965(91)90065-Z)
- Lindsay, D. S., & Read, J. D. (2006). Adults' memories of long-past events. In G.-G. Nilson, & N. Ohta (Eds.), *Memory and Society: Psychological Perspectives* (pp. 51–72). Hove, UK: Psychology Press.
- Loaiza, V. M., & Camos, V. (2018). The role of semantic representations in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 863–881. <http://dx.doi.org/10.1037/xlm0000475>
- Loaiza, V. M., Duperreault, K. A., Rhodes, M. G., & McCabe, D. P. (2015). Long-term semantic representations moderate the effect of attentional refreshing on episodic memory. *Psychonomic Bulletin & Review, 22*, 274–280. <https://doi.org/10.3758/s13423-014-0673-7>
- Loaiza, V. M., & McCabe, D. P. (2013). The influence of aging on attentional refreshing and articulatory rehearsal during working memory on later episodic memory performance. *Aging, Neuropsychology and Cognition, 20*, 37–41. <https://doi.org/10.1080/13825585.2012.738289>
- Loftus, E. F., & Loftus, G. R. (1980). On the permanence of stored information in the human brain. *American Psychologist, 35*(5), 409. <https://doi.org/10.1037/0003-066X.35.5.409>
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of experimental psychology: Human learning and memory, 4*(1), 19. <https://doi.org/10.1037/0278-7393.4.1.19>

- 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(5), 585-589. [https://doi.org/10.1016/S0022-5371\(74\)80011-3](https://doi.org/10.1016/S0022-5371(74)80011-3)
- Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric annals*, 25(12), 720-725. <https://doi.org/10.3928/0048-5713-19951201-07>
- Loftus, E. F., Schooler, J. W., & Wagenaar, W. A. (1985). The fate of memory: comment on McCloskey and Zaragoza. *Journal of Experimental Psychology: General*, 114(3), 375–380. <https://doi.org/10.1037/0096-3445.114.3.375>
- Logie R. H., Camos V., & Cowan N. (2020). *Working memory: State of the science*. Oxford, UK: Oxford University Press.
- Logie, R. H., Camos, V., & Cowan, N. (2021). *Working memory: State of the science*. Oxford University Press.
- Logie, R. H., Della Sala, S., Wynn, V., & Baddeley, A. D. (2000). Visual similarity effects in immediate verbal serial recall. *The Quarterly Journal of Experimental Psychology Section A*, 53(3), 626-646. <https://doi.org/10.1080/713755916>
- Longoni, A. M., & Scalisi, T. G. (1994). Developmental aspects of phonemic and visual similarity effects: Further evidence in Italian children. *International Journal of Behavioral Development*, 17(1), 57-71. <https://doi.org/10.1177/016502549401700104>
- Lucenet, J., & Blaye, A. (2014). Age-related changes in the temporal dynamics of executive control: A study in 5-and 6-year-old children. *Frontiers in Psychology*, 5, 1–11. <https://doi.org/10.3389/fpsyg.2014.00831>
- Lyons, K.E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development*, 82(6), 1778–1787. <https://doi.org/10.1111/j.1467-8624.2011.01649.x>
- Lyons, K.E., & Ghetti, S. (2013). I don't want to pick! Introspection on uncertainty supports early strategic behavior. *Child Development*, 84(2), 726-736. <https://doi.org/10.1111/cdev.12004>
- Lyons, K. E., Ghetti, S., & Cornoldi, C. (2010). Age differences in the contribution of recollection and familiarity to false-memory formation: a new paradigm to examine developmental reversals. *Developmental Science*, 13(2), 355-362. <https://doi.org/10.1111/j.1467-7687.2009.00889.x>
- Maccoby, E. E., & Hagen, J. W. (1965). Effects of distraction upon central versus incidental recall: Developmental trends. *Journal of Experimental Child Psychology*, 2(3), 280-289. [https://doi.org/10.1016/0022-0965\(65\)90031-7](https://doi.org/10.1016/0022-0965(65)90031-7)

- Macé, A. L., & Caza, N. (2011). The role of articulatory suppression in immediate false recognition. *Memory, 19*(8), 891–900. <https://doi.org/10.1080/09658211.2011.613844>
- Magimairaj, B. M., & Montgomery, J. W. (2012). Children's verbal working memory: Relative importance of storage, general processing speed, and domain-general controlled attention. *Acta Psychologica, 140*, 196–207. <https://doi.org/10.1016/j.actpsy.2012.05.004>
- Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive psychology, 57*(4), 335-384. <https://doi.org/10.1016/j.cogpsych.2008.02.004>
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological review, 87*(3), 252. <https://doi.org/10.1037/0033-295X.87.3.252>
- Maren, S., & Holt, W. (2000). The hippocampus and contextual memory retrieval in Pavlovian conditioning. *Behavioural brain research, 110*(1-2), 97-108. [https://doi.org/10.1016/S0166-4328\(99\)00188-6](https://doi.org/10.1016/S0166-4328(99)00188-6)
- Marin, B. V., Holmes, D. L., Guth, M., & Kovac, P. (1979). The potential of children as eyewitnesses. *Law and Human behavior, 3*(4), 295-306.
- Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition, 25*(6), 826-837. <https://doi.org/10.3758/BF03211327>
- McBride, D. M., Coane, J. H., Xu, S., Feng, Y., & Yu, Z. (2019). Short-term false memories vary as a function of list type. *Quarterly Journal of Experimental Psychology, 72*(12), 2726-2741. <https://doi.org/10.1177/1747021819859880>
- McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General, 114*(1), 1.
- McCormack, T., Brown, G. D., Vousden, J. I., & Henson, R. N. (2000). Children's serial recall errors: Implications for theories of short-term memory development. *Journal of Experimental Child Psychology, 76*(3), 222-252. <https://doi.org/10.1006/jecp.1999.2550>
- McEvoy, C. L., Nelson, D. L., & Komatsu, T. (1999). What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(5), 1177. <https://doi.org/10.1037/0278-7393.25.5.1177>
- McGaugh, J. L. (2000). Memory--a century of consolidation. *Science, 287*(5451), 248-251. <https://doi.org/10.1126/science.287.5451.248>

- McKone, E., & Murphy, B. (2000). Implicit false memory: Effects of modality and multiple study presentations on long-lived semantic priming. *Journal of Memory and Language*, 43(1), 89-109. <https://doi.org/10.1006/jmla.1999.2702>
- McWilliams, K., Harris, L. S., & Goodman, G. S. (2014). Child maltreatment, trauma-related psychopathology, and eyewitness memory in children and adolescents. *Behavioral Sciences & the Law*, 32, 702–717. <https://doi.org/10.1002/bsl.2143>
- Metzger, R. L., Warren, A. R., Shelton, J. T., Price, J., Reed, A. W., & William, D. (2008). Do children 'DRM' like adults ? False memory production in children. *Developmental Psychology*, 44 (1), 169-181. <https://doi.org/10.1037/0012-1649.44.1.169>
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81. <https://doi.org/10.1037/h0043158>
- Miller, S., McCulloch, S., & Jarrold, C. (2015). The development of memory maintenance strategies: Training cumulative rehearsal and interactive imagery in children aged between 5 and 9. *Frontiers in psychology*, 6, 524. <https://doi.org/10.3389/fpsyg.2015.00524>
- Miller, L. M., & Roodenrys, S. (2009). The interaction of word frequency and concreteness in immediate serial recall. *Memory & Cognition*, 37(6), 850-865. <https://doi.org/10.3758/MC.37.6.850>
- Mirandola, C., Toffalini, E., Grassano, M., Cornoldi, C., & Melinder, A. (2014). Inferential false memories of events: Negative consequences protect from distortions when the events are free from further elaboration. *Memory*, 22(5), 451-461. <https://doi.org/10.1080/09658211.2013.795976>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive psychology*, 41(1), 49-100. <https://doi.org/10.1006/cogp.1999.0734>
- Montefinese, M., Zannino, G. D., & Ambrosini, E. (2015). Semantic similarity between old and new items produces false alarms in recognition memory. *Psychological research*, 79, 785-794. <https://doi.org/10.1007/s00426-014-0615-z>
- Mora, G., & Camos, V. (2013). Two systems of maintenance in verbal working memory: Evidence from the word length effect. *PLoS ONE*, 8, e70026. <http://dx.doi.org/10.1371/journal.pone.0070026>

- Mora, G., & Camos, V. (2015). Dissociating rehearsal and refreshing in the maintenance of verbal information in 8-year-old children. *Frontiers in Psychology*, 6, 11. <https://doi.org/10.3389/fpsyg.2015.00011>
- Moradi, A. R., Doost, H. T. N., Taghavi, M. R., Yule, W., & Dalglish, T. (1999). Everyday memory deficits in children and adolescents with PTSD: Performance on the Rivermead Behavioural Memory Test. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40(3), 357-361. <https://doi.org/10.1111/1469-7610.00453>
- Morton, J., Hammersley, R. H., & Bekerian, D. A. (1985). Headed records: A model for memory and its failures. *Cognition*, 20(1), 1-23. [https://doi.org/10.1016/0010-0277\(85\)90002-2](https://doi.org/10.1016/0010-0277(85)90002-2)
- Moshagen, M. (2010). MultiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, 42(1), 42-54. <https://doi.org/10.3758/BRM.42.1.42>
- Munakata, Y., Snyder, H.R., & Chatham, C.H. (2012). Developing cognitive control: Three key transitions. *Current Directions in Psychological Science*, 21, 71-77. <https://doi.org/10.1177/0963721412436807>
- Murphy, G., Loftus, E. F., Grady, R. H., Levine, L. J., & Greene, C. M. (2019). False memories for fake news during Ireland's abortion referendum. *Psychological science*, 30(10), 1449-1459. <https://doi.org/10.1177/0956797619864887>
- Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, 53, 53-81. <http://dx.doi.org/10.1146/annurev.psych.53.100901.135131>
- Naus, M. J., Ornstein, P. A., & Aivano, S. (1977). Developmental changes in memory: The effects of processing time and rehearsal instructions. *Journal of Experimental Child Psychology*, 23(2), 237-251. [https://doi.org/10.1016/0022-0965\(77\)90102-3](https://doi.org/10.1016/0022-0965(77)90102-3)
- Neath, I., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review*, 2, 429-441. <http://dx.doi.org/10.3758/BF03210981>
- Neuschatz, J. S., Benoit, G. E., & Payne, D. G. (2003). Effective warnings in the Deese-Roediger-McDermott false-memory paradigm: The role of identifiability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(1), 35-41. <https://doi.org/10.1037/0278-7393.29.1.35>

- Newman, E. J., & Lindsay, D. S. (2009). False memories: What the hell are they for?. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 23(8), 1105-1121. <https://doi.org/10.1002/acp.1613>
- Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & cognition*, 25(6), 838-848.
- Norris, D. (2019). Even an activated long-term memory system still needs a separate short-term store: A reply to Cowan (2019). *Psychological Bulletin*, 145(8), 848–853. <https://doi.org/10.1037/bul0000204>
- Oates, K., & Shrimpton, S. (1991). Children's memories for stressful and non-stressful events. *Medicine, Science and the Law*, 31(1), 4-10. <https://doi.org/10.1177/002580249103100102>
- Oberauer, K. (2003). Selective attention to elements in working memory. *Experimental Psychology*, 50(4), 257-269. <https://doi.org/10.1027//1618-3169.50.4.257>
- Oberauer, K. (2021). Toward a Theory of Working Memory: From Metaphors to Mechanisms. In R. Logie, V. Camos, & N. Cowan (Eds.), *Working Memory*. Oxford University Press.
- Odegard, T. N., Cooper, C. M., Lampinen, J. M., Reyna, V. F., & Brainerd, C. J. (2009). Children's eyewitness memory for multiple real-life events. *Child Development*, 80(6), 1877-1890. <https://doi.org/10.1111/j.1467-8624.2009.01373.x>
- Odegard, T.N., Holliday, R.E., Brainerd, C.J., Reyna, V.F. (2008). Attention to global gist processing eliminates age effects in false memories. *Journal of Experimental Child Psychology*, 99, 96-113. <https://doi.org/10.1016/j.jecp.2007.08.007>
- Oftinger, A., & Camos, V. (2016). Maintenance mechanisms in children's verbal working memory. *Journal of Educational and Developmental Psychology*, 6 (1), 16-28. <https://doi.org/10.5539/jedp.v6n1p16>
- Oftinger, A. L., & Camos, V. (2017). Phonological similarity effect in children's working memory: Do maintenance mechanisms matter. *Journal of Child Psychology*, 1(1), 5-11.
- Oftinger, A., & Camos, V. (2018). Developmental improvement in strategies to maintain verbal information in working memory. *International Journal of Behavioral Development*, 42(2), 1-10. <https://doi.org/10.1177/0165025416679741>
- Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology*, 30(10), 953–958.
- Olszewska, J. M., Reuter-Lorenz, P. A., Munier, E., & Bendler, S. A. (2015). Misremembering what you see or hear: Dissociable effects of modality on short- and long-term false

- recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(5), 1316–1325. <https://doi.org/10.1037/xlm0000115>
- Ornstein, P. A., Gordon, B. N., & Larus, D. M. (1992). Children's memory for a personally experienced event: Implications for testimony. *Applied Cognitive Psychology*, 6(1), 49-60. <https://doi.org/10.1002/acp.2350060103>
- Ornstein, P. A., Naus, M. J., & Liberty, C. (1975). Rehearsal and organizational processes in children's memory. *Child Development*, 46(4), 818-830. <https://doi.org/10.2307/1128385>
- 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(1), 76-85. <https://doi.org/10.1080/09658210701723323>
- Otgaar, H., & Candel, I. (2011). Children's false memories: Different false memory paradigms reveal different results. *Psychology, Crime & Law*, 17(6), 513-528. <https://doi.org/10.1080/10683160903373392>
- Otgaar, H., Candel, I., Merckelbach, H., & Wade, K. A. (2009). Abducted by a UFO: Prevalence information affects young children's false memories for an implausible event. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 23(1), 115-125. <https://doi.org/10.1002/acp.1445>
- Otgaar, H., Candel, I., Smeets, T., & Merckelbach, H. (2010). ‘You didn't take Lucy's skirt off’: The effect of misleading information on omissions and commissions in children's memory reports. *Legal and Criminological Psychology*, 15(2), 229-241. <https://doi.org/10.1348/135532509X471951>
- Otgaar, H., Howe, M. L., Brackmann, N., & Smeets, T. (2016). The malleability of developmental trends in neutral and negative memory illusions. *Journal of Experimental Psychology: General*, 145(1), 31. <https://doi.org/10.1037/xge0000127>
- Otgaar, H., Howe, M. L., Merckelbach, H., & Muris, P. (2018). Who is the better eyewitness? Sometimes adults but at other times children. *Current Directions in Psychological Science*, 27(5), 378-385. <https://doi.org/10.1177/0963721418770998>
- Otgaar, H., Howe, M. L., & Muris, P. (2017). Maltreatment increases spontaneous false memories but decreases suggestion-induced false memories in children. *British journal of developmental psychology*, 35(3), 376-391. <https://doi.org/10.1111/bjdp.12177>
- Otgaar, H., Howe, M. L., Muris, P. & Merckelbach, H. (2019a). Associative Activation as a Mechanism Underlying False Memory Formation. *Clinical Psychological Science*, 7(2), pp. 191-195. <https://doi.org/10.1177/2167702618807189>

- Otgaar, H., Howe, M. L., Muris, P., & Merckelbach, H. (2019b). Dealing with false memories in children and adults: Recommendations for the legal arena. *Policy Insights from the Behavioral and Brain Sciences*, 6(1), 87-93. <https://doi.org/10.1177/2372732218818584>
- Otgaar H., Howe M. L., Peters M., Sauerland M., Raymaekers L. (2013). Developmental trends in different types of spontaneous false memories: Implications for the legal field. *Behavioral Sciences & the Law*, 31, 666–682. <https://doi.org/10.1002/bsl.2076>
- Otgaar, H., Howe, M.L., Peters, M., Smeets, T., & Moritz, S. (2014). The production of spontaneous false memories across childhood. *Journal of Experimental Child Psychology*, 121, 28-41. <https://doi.org/10.1016/j.jecp.2013.11.019>
- Otgaar, H., Peters, M., & Howe, M. L. (2012a). Dividing attention lowers children's but increases adults' false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 204–210. <http://dx.doi.org/10.1037/a0025160>.
- Otgaar, H., & Smeets, T. (2010). Adaptive memory: Survival processing increases both true and false memory in adults and children. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1010–1016. <http://dx.doi.org/10.1037/a0019402>.
- Otgaar, H., Verschuere, B., Meijer, E. H., & van Oorsouw, K. (2012b). The origin of children's implanted false memories: Memory traces or compliance. *Acta Psychologica*, 139, 397–403. <https://doi.org/10.1016/j.actpsy.2012.01.002>
- Ozdes, A., Bagci, E., Burhan-Cavusoglu, P., & Ulusoy-Kok, N. (2021). Self-referential processing in false recognition and source monitoring: Self-other differences. *Journal of Cognitive Psychology*, 33(8), 823-836. <https://doi.org/10.1080/20445911.2021.1977816>
- Palmer, S. (2000). Working memory: A developmental study of phonological recoding. *Memory*, 8(3), 179-193. <https://doi.org/10.1080/096582100387597>
- Palmer, J. E., & Dodson, C. S. (2009). Investigating the mechanisms fuelling reduced false recall of emotional material. *Cognition and Emotion*, 23(2), 238-259. <https://doi.org/10.1080/02699930801976663>
- Paris, S. G., & Carter, A. Y. (1973). Semantic and constructive aspects of sentence memory in children. *Developmental Psychology*, 9(1), 109. <https://doi.org/10.1037/h0035071>
- Patihis, L., Frenda, S. J., & Loftus, E. F. (2018). False memory tasks do not reliably predict other false memories. *Psychology of Consciousness: Theory, Research, and Practice*, 5(2), 140. <https://doi.org/10.1037/cns0000147>
- Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, 35(2), 261-285. <https://doi.org/10.1006/jmla.1996.0015>

- Paz-Alonso, P. M., Ghetti, S., Donohue, S. E., Goodman, G. S., & Bunge, S. A. (2008). Neurodevelopmental correlates of true and false recognition. *Cerebral Cortex, 18*(9), 2208-2216. <https://doi.org/10.1093/cercor/bhm246>
- Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin, 144*(1), 48–76. <https://doi.org/10.1037/bul0000124>
- Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology, 108*(4), 455–473. <https://doi.org/10.1037/edu0000079>
- Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17*, 398–422. <http://dx.doi.org/10.3758/BF03202613>
- Peters, M. J., Jelicic, M., Gorski, B., Sijstermans, K., Giesbrecht, T., & Merckelbach, H. (2008). The corrective effects of warning on false memories in the DRM paradigm are limited to full attention conditions. *Acta Psychologica, 129*(2), 308-314. <https://doi.org/10.1016/j.actpsy.2008.08.007>
- Peterson, L.R., & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193–198.
- Poirier, M., Dhir, P., Saint-Aubin, J., Tehan, G., & Hampton, J. (2011). The influence of semantic memory on verbal short-term memory. In B. Kokinov, A. Karmiloff-Smith, & N. J. Nersessian (Eds.), *European Perspectives on Cognitive Science*.
- Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology, 48*(2), 384-404. <https://doi.org/10.1080/14640749508401396>
- Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 50*(4), 408. <https://doi.org/10.1037/1196-1961.50.4.408>
- Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and human behavior, 23*(5), 517-537. <https://doi.org/10.1023/A:1022344128649>

- Pratt, D. D. (1991). Conceptions of self within China and the United States: Contrasting foundations for adult education. *International Journal of intercultural relations*, 15(3), 285-310. [https://doi.org/10.1016/0147-1767\(91\)90003-Y](https://doi.org/10.1016/0147-1767(91)90003-Y)
- Principe, G. F., Guiliano, S., & Root, C. (2008). Rumor mongering and remembering: How rumors originating in children's inferences can affect memory. *Journal of Experimental Child Psychology*, 99(2), 135-155. <https://doi.org/10.1016/j.jecp.2007.10.009>
- Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2007). Refreshing: A minimal executive function. *Cortex*, 43(1), 135-145. [https://doi.org/10.1016/S0010-9452\(08\)70451-9](https://doi.org/10.1016/S0010-9452(08)70451-9)
- Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. (Les problems.) [The psychological examination in cases of traumatic encephalopathy. Problems]. *Archives de Psychologie*, 28, 215–285.
- Reyna, V. F., Holliday, R., & Marche, T. (2002). Explaining the development of false memories. *Developmental Review*, 22(3), 436-489. [https://doi.org/10.1016/S0273-2297\(02\)00003-5](https://doi.org/10.1016/S0273-2297(02)00003-5)
- Reyna, V.F., & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, 30, 178-191. <https://doi.org/10.1037/0012-1649.30.2.178>
- Reyna, V.F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences*, 9, 95–123. [https://doi.org/10.1016/S1041-6080\(97\)90002-9](https://doi.org/10.1016/S1041-6080(97)90002-9)
- Roberts, K. P., & Blades, M. (1998). The effects of interacting in repeated events on children's eyewitness memory and source monitoring. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 12(5), 489-503. [https://doi.org/10.1002/\(SICI\)1099-0720\(199810\)12:5<489::AID-ACP535>3.0.CO;2-%23](https://doi.org/10.1002/(SICI)1099-0720(199810)12:5<489::AID-ACP535>3.0.CO;2-%23)
- Robin, F., Ménétrier, E., & Beffara Bret, B. (2022). Effects of visual imagery on false memories in DRM and misinformation paradigms. *Memory*, 30(6), 725-732. <https://doi.org/10.1080/09658211.2021.1895221>
- Roebbers, C. M., Bjorklund, D. F., Schneider, W., & Cassel, W. S. (2002). Differences and similarities in event recall and suggestibility between children and adults in Germany and the United States. *Experimental Psychology*, 49(2), 132. <https://doi.org/10.1027/1618-3169.49.2.132>
- Roebbers, C. M., Howie, P., & Beuscher, E. (2007). Can private reports enhance children's event recall, lower their suggestibility and foster their metacognitive monitoring compared to

- face-to-face interviews? *Computers in human behavior*, 23(1), 749-769.
<https://doi.org/10.1016/j.chb.2004.11.007>
- Roebbers, C. M., & Schneider, W. (2002). Stability and consistency of children's event recall. *Cognitive Development*, 17(1), 1085-1103. [https://doi.org/10.1016/S0885-2014\(02\)00075-8](https://doi.org/10.1016/S0885-2014(02)00075-8)
- Roebbers, C. M., & Schneider, W. (2005). The strategic regulation of children's memory performance and suggestibility. *Journal of Experimental Child Psychology*, 91(1), 24-44.
<https://doi.org/10.1016/j.jecp.2005.01.001>
- Roediger, H. L. III, Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95–115). American Psychological Association. <https://doi.org/10.1037/10394-006>
- 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(4), 803–814.
- Roediger, H. L., & McDermott, K. B. (2000). Tricks of memory. *Current directions in psychological science*,9(4), 123-127.
- Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic bulletin & review*, 8(3), 385-407.
- Romani, C., Mcalpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *Quarterly Journal of Experimental Psychology*, 61(2), 292-323. <https://doi.org/10.1080/17470210601147747>
- Roodenrys, S., Hulme, C., & Brown, G. (1993). The development of short-term memory span: Separable effects of speech rate and long-term memory. *Journal of Experimental Child Psychology*, 56(3), 431-442. <https://doi.org/10.1006/jecp.1993.1043>
- Rosa, N. M., & Gutchess, A. H. (2013). False memory in aging resulting from self-referential processing. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 68(6), 882-892. <https://doi.org/10.1093/geronb/gbt018>
- Rose, N. S., Buchsbaum, B. R., & Craik, F. I. M. (2014). Short-term retention of a single word relies on retrieval from long-term memory when both rehearsal and refreshing are disrupted. *Memory & Cognition*, 42, 689–700. <https://doi.org/10.3758/s13421-014-0398->

- Rose, N. S., Craik, F. I. M., & Buchsbaum, B. R. (2015). Levels of processing in working memory: Differential involvement of frontotemporal networks. *Journal of Cognitive Neuroscience*, 27, 522–532. https://doi.org/10.1162/jocn_a_00738
- Rose, N. S., Myerson, J., Roediger, H. L., & Hale, S (2010). Similarities and differences between working memory and long-term memory: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning Memory and Cognition*, 36(2), 471-483. <https://doi.org/10.1037/a0018405>
- Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological review*, 96(2), 341. <https://doi.org/10.1037/0033-295X.96.2.341>
- Ross, D. F., Marsil, D. F., Benton, T. R., Hoffman, R., Warren, A. R., Lindsay, R. C. L., & Metzger, R. (2006). Children’s susceptibility to misidentifying a familiar bystander from a lineup: When younger is better. *Law and Human Behavior*, 30(3), 249-257.
- Rosselet-Jordan, F. L., Abadie, M., Mariz-Elsig, S., & Camos, V. (2022). Role of attention in the associative relatedness effect in verbal working memory: Behavioral and chronometric perspectives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0001102>
- Rousselle, M., Abadie, M., Blaye, A., & Camos, V. (2022). False memories in children’s working memory. Retrieved from <https://osf.io/9gfmd/>
- Rousselle, M., Abadie, M., Blaye, A., & Camos, V. (2023). Children’s gist-based false memory in working memory. *Developmental Psychology*. Advance online publication. <https://doi.org/10.1037/dev0001476>
- Rousselle, M., Blaye, A. & M., Abadie (in prep.). The development of false memory in working memory tasks: implication of maintenance mechanisms at 5- and 8-year-old.
- Ruci, L., Tomes, J. L., & Zelenski, J. M. (2009). Mood-congruent false memories in the DRM paradigm. *Cognition and Emotion*, 23, 1153– 1165. <http://dx.doi.org/10.1080/02699930802355420>
- Rudy, L., & Goodman, G. S. (1991). Effects of participation on children's reports: Implications for children's testimony. *Developmental psychology*, 27(4), 527. <https://doi.org/10.1037/0012-1649.27.4.527>
- Ruffman, T., Rustin, C., Garnham, W., & Parkin, A. J. (2001). Source monitoring and false memories in children: Relation to certainty and executive functioning. *Journal of Experimental Child Psychology*, 80(2), 95-111. <https://doi.org/10.1006/jecp.2001.2632>
- Russell, J. A. (1991). Culture and the categorization of emotions. *Psycho-logical Bulletin*, 110, 426–450. <http://dx.doi.org/10.1037/0033-2909.110.3.426>

- Ryback, R. S. (1971). The continuum and specificity of the effects of alcohol on memory; a review. *Quarterly journal of studies on alcohol*, 32(4), 995-1016. <https://doi.org/10.15288/qjsa.1971.32.995>
- Rybash, J. M., & Colilla, J. L. (1994). Source memory deficits and frontal lobe functioning in children. *Developmental Neuropsychology*, 10(1), 67-73. <https://doi.org/10.1080/87565649409540567>
- Saint-Aubin, J., & LeBlanc, J. (2005). Word frequency effects in immediate serial recall of pure and mixed lists: tests of the associative link hypothesis. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 59(4), 219. <https://doi.org/10.1037/h0087477>
- Saint-Aubin, J., & Poirier, M. (1999a). The influence of long-term memory factors on immediate serial recall: An item and order analysis. *International Journal of Psychology*, 34(5-6), 347-352. <https://doi.org/10.1080/002075999399675>
- Saint-Aubin, J., & Poirier, M. (1999b). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information?. *The Quarterly Journal of Experimental Psychology: Section A*, 52(2), 367-394. <https://doi.org/10.1080/713755814>
- Sapolsky, R. M. (1992). *Stress, the aging brain, and the mechanisms of neuron death*. the MIT Press.
- Saywitz, K. J., Goodman, G. S., Nicholas, E., & Moan, S. F. (1991). Children's memories of a physical examination involving genital touch: Implications for reports of child sexual abuse. *Journal of consulting and clinical psychology*, 59(5), 682. <https://doi.org/10.1037/0022-006X.59.5.682>
- Schacter, D. L., & Addis, D. R. (2007). The ghosts of past and future. *Nature*, 445(7123), 27-27. <https://doi.org/10.1038/445027a>
- Schacter, D. L., Cendan, D. L., Dodson, C. S., & Clifford, E. R. (2001). Retrieval conditions and false recognition: Testing the distinctiveness heuristic. *Psychonomic Bulletin & Review*, 8(4), 827-833. <https://doi.org/10.3758/BF03196224>
- Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing: behavior, theory, and biological mechanisms. *Cognitive science*, 27(3), 525-559. https://doi.org/10.1207/s15516709cog2703_8
- Schopen, K., Otgaar, H., Howe, M. L., & Muris, P. (2022). Effects of forewarnings on children's and adults' spontaneous false memories. *European Journal of Developmental Psychology*, 19(2), 177-197. <https://doi.org/10.1080/17405629.2021.1904877>

- Seamon, J. G., Luo, C. R., & Gallo, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. *Psychological Science*, 9(1), 20-26. <https://doi.org/10.1111/1467-9280.0000>
- Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., & Schwartz, M. A. (2002). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Memory & Cognition*, 30, 1054-1064. <https://doi.org/10.3758/BF03194323>
- Sharman, S. J., Garry, M., Jacobson, J. A., Loftus, E. F., & Ditto, P. H. (2008). False memories for end-of-life decisions. *Health Psychology*, 27(2), 291. <https://doi.org/10.1037/0278-6133.27.2.291>
- Sheng, L., Byrd, C. T., McGregor, K. K., Zimmerman, H., & Bludau, K. (2015). List memory in young adults with language learning disability. *Journal of Speech, Language, and Hearing Research*, 58(2), 336-344. https://doi.org/10.1044/2015_JSLHR-L-13-0143
- Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research*, 31(4), 713-724. <https://doi.org/10.1086/426605>
- Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: a similarity-based model. *Journal of Experimental Psychology: General*, 133(2), 166. <https://doi.org/10.1037/0096-3445.133.2.166>
- Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review*, 5(4), 710-715. <https://doi.org/10.3758/BF03208850>
- Smith, R. E., & Hunt, R. R. (2020). When do pictures reduce false memory?. *Memory & Cognition*, 48(4), 623-644. <https://doi.org/10.3758/s13421-019-00995-5>
- Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, 283(5408), 1657-1661. <https://doi.org/10.1126/science.283.5408.1657>
- Snodgrass, J. G., & Vanderwart, M., (1980). A standardized set of 260 pictures: Norms for name agreement, Image agreement, Familiarity; and Visual complexity. *Journal of experimental Psychology: Human Learning and Memory*, 6(2), 174-215.
- Spear, N. E. (1973). Retrieval of memory in animals. *Psychological Review*, 80(3), 163. <https://doi.org/10.1037/h0034326>
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1. <https://doi.org/10.1037/h0093759>
- Stahl, C., & Klauer, K.C. (2008). A simplified conjoint recognition paradigm for the measurement of gist and verbatim memory. *Journal of Experimental Psychology:*

- Learning, Memory, and Cognition, 34(3), 570. <https://doi.org/10.1037/0278-7393.34.3.570>
- Stahl, C., & Klauer, K. C. (2009). Measuring phantom recollection in the simplified conjoint recognition paradigm. *Journal of Memory and Language*, 60(1), 180-193. <https://doi.org/10.1016/j.jml.2008.08.001>
- Stern, W. (1910). Abstracts of lectures on the psychology of testimony and on the study of individuality. *The American Journal of Psychology*, 21(2), 270-282.
- Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, 437(7063), 1272-1278. <https://doi.org/10.1038/nature04286>
- Storbeck, J. (2013). Negative affect promotes encoding of and memory for details at the expense of the gist: Affect, encoding, and false memories. *Cognition and Emotion*, 27, 800–819. <http://dx.doi.org/10.1080/02699931.2012.741060>
- Storbeck, J., & Clore, G. L. (2005). With sadness comes accuracy; with happiness, false memory: Mood and the false memory effect. *Psychological Science*, 16, 785–791. <http://dx.doi.org/10.1111/j.1467-9280.2005.01615.x>
- Storbeck, J., & Clore, G. L. (2011). Affect influences false memories at encoding: Evidence from recognition data. *Emotion*, 11, 981–989. <http://dx.doi.org/10.1037/a0022754>
- Strange, D., Sutherland, R., & Garry, M. (2006). Event plausibility does not determine children's false memories. *Memory*, 14(8), 937-951. <https://doi.org/10.1080/09658210600896105>
- Stuart, G., & Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 796. <https://doi.org/10.1037/0278-7393.26.3.796>
- Sugrue, K., & Hayne, H. (2006). False memories produced by children and adults in the DRM paradigm. *Applied Cognitive Psychology*, 20, 625–631. <https://doi.org/10.1002/acp.1214>
- Sugrue, K., Strange, D., & Hayne, H. (2009). False memories in the DRM paradigm: Age-related differences in lure activation and source monitoring. *Experimental Psychology*, 56(5), 354. <https://doi.org/10.1027/1618-3169.56.5.354>
- Sulin, R. A., & Dooling, D. J. (1974). Intrusion of a thematic idea in retention of prose. *Journal of experimental psychology*, 103(2), 255. <https://doi.org/10.1037/h0036846>
- Surprenant, A. E. M., & Neath, I. A. N. (2008). The nine lives of short-term memory. In *Interactions between short-term and long-term memory in the verbal domain* (pp. 28-55). Psychology Press. <https://doi.org/10.4324/9780203848760>

- Swannell, E. R., & Dewhurst, S. A. (2012). Phonological false memories in children and adults: Evidence for a developmental reversal. *Journal of Memory and Language*, 66 (2), 376–383. <https://doi.org/10.1016/j.jml.2011.11.003>.
- Swanson, H. L., & Alloway, T. P. (2012). Working memory, learning, and academic achievement. *APA Educational Psychology Handbook, Vol 1: Theories, Constructs, and Critical Issues.*, 327–366. <https://doi.org/10.1037/13273-012>
- Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: a meta-analysis. *Psychological bulletin*, 121(3), 371.
- Tam, H., Jarrold, C., Baddeley, A.D., & Sabatos-DeVito, M. (2010). The development of memory maintenance: Children's use of phonological rehearsal and attentional refreshment in working memory tasks. *Journal of Experimental Child Psychology*, 107, 306–324. <https://doi.org/10.1016/j.jecp.2010.05.006>
- Tehan, G. (2010). Associative relatedness enhances recall and produces false memories in immediate serial recall. *Canadian Journal of Experimental Psychology*, 64(4), 266-272. <https://doi.org/10.1037/a0021375>
- Tehan, G., & Humphreys, M. S. (1988). Articulatory loop explanations of memory span and pronunciation rate correspondences: A cautionary note. *Bulletin of the Psychonomic Society*, 26, 293-296. <https://doi.org/10.3758/BF03337662>
- Templeton, L. M., & Wilcox, S. A. (2000). A tale of two representations: The misinformation effect and children's developing theory of mind. *Child Development*, 71(2), 402-416. <https://doi.org/10.1111/1467-8624.00153>
- Tipper, S. P., Bourque, T. A., Anderson, S. H., & Brehaut, J. C. (1989). Mechanisms of attention: A developmental study. *Journal of Experimental Child Psychology*, 48(3), 353-378. [https://doi.org/10.1016/0022-0965\(89\)90047-7](https://doi.org/10.1016/0022-0965(89)90047-7)
- Toth, S. L., & Cicchetti, D. (1998). Remembering, forgetting, and the effects of trauma on memory: A developmental psychopathology perspective. *Development and Psychopathology*, 10(4), 589-605.
- Towse, J. N., & Hitch, G. J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity?. *The Quarterly Journal of Experimental Psychology*, 48(1), 108-124. <https://doi.org/10.1080/14640749508401379>
- Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory and Language*, 39(2), 195-217. <https://doi.org/10.1006/jmla.1998.2574>

- Towse, J. N., Hitch, G. J., & Skeates, S. (1999). Developmental sensitivity to temporal grouping effects in short-term memory. *International Journal of Behavioral Development, 23*(2), 391-411. <https://doi.org/10.1080/016502599383883>
- Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of verbal learning and verbal behavior, 5*(4), 381-391. [https://doi.org/10.1016/S0022-5371\(66\)80048-8](https://doi.org/10.1016/S0022-5371(66)80048-8)
- Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology, 77*(4), 593–601. <https://doi.org/10.1037/h0026069>
- Tulving, E., & Thomson, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology, 87*(1), 116–124. <https://doi.org/10.1037/h0030186>
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review, 80* (5), 352-373.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of Memory and Language, 28*(2), 127-154. [https://doi.org/10.1016/0749-596X\(89\)90040-5](https://doi.org/10.1016/0749-596X(89)90040-5)
- Uittenhove, K., Chaabi, L., Camos, V., & Barrouillet, P. (2019). Is working memory storage intrinsically domain-specific? *Journal of Experimental Psychology: General, 148*(11), 2027-2057. <https://doi.org/10.1037/xge0000566>
- Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70*(1), 122. <https://doi.org/10.1037/h0022014>
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104. <https://doi.org/10.1037/0033-295X.114.1.104>
- Valentino, K., Toth, S. L., & Cicchetti, D. (2009). Autobiographical memory functioning among abused, neglected, and nonmaltreated children: The overgeneral memory effect. *Journal of Child Psychology and Psychiatry, 50*(8), 1029-1038. <https://doi.org/10.1111/j.1469-7610.2009.02072.x>
- Van Damme, I. (2013). Mood and the DRM paradigm: An investigation of the effects of valence and arousal on false memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 66*, 1060–1081. <http://dx.doi.org/10.1080/17470218.2012.727837>

- Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The impact of storage on processing: How is information maintained in working memory?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1072.
- VoiceBooking Team (2021). [Website]. Retrieved from <https://www.voicebooking.com/fr>
- Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1256. <https://doi.org/10.1037/0278-7393.25.5.1256>
- Wang, J., Otgaar, H., Howe, M. L., & Cheng, S. (2021). Self-referential false associations: A self-enhanced constructive effect for verbal but not pictorial stimuli. *Quarterly Journal of Experimental Psychology*, *74*(9), 1512-1524. <https://doi.org/10.1177/17470218211009772>
- Wang, J., Otgaar, H., Howe, M. L., Dong, Q., & Zhou, C. (2022). Self-Enhanced False Memory Across the Life Span. *The Journals of Gerontology: Series B*. <https://doi.org/10.1093/geronb/gbac020>
- Wang, J., Otgaar, H., Howe, M. L., & Zhou, C. (2019). A self-reference false memory effect in the DRM paradigm: Evidence from Eastern and Western samples. *Memory & Cognition*, *47*, 76-86. <https://doi.org/10.3758/s13421-018-0851-3>
- Waroquier, L., Abadie, M., & Dienes, Z. (2020). Distinguishing the role of conscious and unconscious knowledge in evaluative conditioning. *Cognition*, *205*, 104460. <https://doi.org/10.1016/j.cognition.2020.104460>
- Watkins, O. C., & Watkins, M. J. (1977). Serial recall and the modality effect: Effects of word frequency. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(6), 712. <https://doi.org/10.1037/0278-7393.3.6.712>
- Watson, J. M., Balota, D. A., & Roediger III, H. L. (2003). Creating false memories with hybrid lists of semantic and phonological associates: Over-additive false memories produced by converging associative networks. *Journal of Memory and Language*, *49*(1), 95-118. [https://doi.org/10.1016/S0749-596X\(03\)00019-6](https://doi.org/10.1016/S0749-596X(03)00019-6)
- Watson, J. M., Balota, D. A., & Sergent-Marshall, S. D. (2001). Semantic, phonological, and hybrid veridical and false memories in healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology*, *15*(2), 254. <https://doi.org/10.1037/0894-4105.15.2.254>
- Watson, J. M., Mcdermott, K. B., & Balota, D. A. (2004). Attempting to avoid false memories in the Deese/Roediger—McDermott paradigm: Assessing the combined influence of

- practice and warnings in young and old adults. *Memory & cognition*, 32(1), 135-141.
<https://doi.org/10.3758/BF03195826>
- Weekes, B. S., Hamilton, S., Oakhill, J. V., & Holliday, R. E. (2008). False recollection in children with reading comprehension difficulties. *Cognition*, 106(1), 222-233.
<https://doi.org/10.1016/j.cognition.2007.01.005>
- Wilburn, C., & Feeney, A. (2008). Do development and learning really decrease memory? On similarity and category-based induction in adults and children. *Cognition*, 106(3), 1451-1464. <https://doi.org/10.1016/j.cognition.2007.04.018>
- Wilkinson, D. A., & Poulos, C. X. (1987). The chronic effects of alcohol on memory: A contrast between a unitary and dual system approach. *Recent Developments in Alcoholism: Memory Deficits Sociology of Treatment Ion Channels Early Problem Drinking*, 5-26.
https://doi.org/10.1007/978-1-4899-1684-6_1
- Wilson, A., & Ross, M. (2003). The identity function of autobiographical memory: Time is on our side. *Memory*, 11(2), 137-149. <https://doi.org/10.1080/741938210>
- Wimmer, M. C., & Howe, M. L. (2009). The development of automatic associative processes and children's false memories. *Journal of Experimental Child Psychology*, 104(4), 447-465.
- Yakolev, P. I., & Lecours, A. R. (1967). The myelogenetic cycles of regional maturation of the brain. In A. MINKOWSKI (Ed.). *Regional development of the brain early in life* (pp. 3-170). Oxford, UK: Blackwell
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46(3), 441-517.
<https://doi.org/10.1006/jmla.2002.2864>
- Zaragoza, M. S. (1991). Preschool children's susceptibility to memory impairment. In J. Doris (Ed.), *The suggestibility of children's recollections: Implications for eyewitness testimony* (pp. 27-46). Washington, DC: American Psychological Association.
- Zaragoza, M. S., & McCloskey, M. (1989). Misleading postevent information and the memory impairment hypothesis: Comment on Belli and reply to Tversky and Tuchin. *Journal of Experimental Psychology: General*, Vol 118(1), Mar 1989, 92-99.
- Zeelenberg, R., Boot, I., & Pecher, D. (2005). Activating the critical lure during study is unnecessary for false recognition. *Consciousness and Cognition*, 14(2), 316-326.
<https://doi.org/10.1016/j.concog.2004.08.004>
- Zhu, B., Chen, C., Shao, X., Liu, W., Ye, Z., Zhuang, L., ... & Xue, G. (2019). Multiple interactive memory representations underlie the induction of false memory. *Proceedings*

of the National Academy of Sciences, 116(9), 3466-3475.
<https://doi.org/10.1073/pnas.1817925116>

Zlotnik, G., & Vansintjan, A. (2019). Memory: An Extended Definition. *Frontiers in Psychology, 10.* <https://doi.org/10.3389/fpsyg.2019.02523>