



**HAL**  
open science

# Towards a unified view of metacognition : Insights from metamemory

Audrey Mazancieux

► **To cite this version:**

Audrey Mazancieux. Towards a unified view of metacognition : Insights from metamemory. Psychology. Université Grenoble Alpes [2020-..], 2020. English. NNT : 2020GRALS004 . tel-03088370

**HAL Id: tel-03088370**

**<https://theses.hal.science/tel-03088370v1>**

Submitted on 26 Dec 2020

**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

Pour obtenir le grade de

### DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : PCN - Sciences cognitives, psychologie et neurocognition

Arrêté ministériel : 25 mai 2016

Présentée par

**Audrey MAZANCIEUX**

Thèse dirigée par **Christopher MOULIN**, Université Grenoble Alpes

et codirigée par **Céline SOUCHAY**, Université Grenoble Alpes

préparée au sein du **Laboratoire Laboratoire de Psychologie et Neuro Cognition**  
dans l'**École Doctorale Ingénierie pour la Santé la Cognition et l'Environnement**

### **Vers une vision unifiée de la métacognition : Apports de la métamémoire**

### **Towards a unified view of metacognition: Insights from metamemory**

Thèse soutenue publiquement le **12 mars 2020**,  
devant le jury composé de :

**Monsieur CHRISTOPHER MOULIN**

PROFESSEUR, UNIVERSITE GRENOBLE ALPES, Directeur de thèse

**Madame CELINE SOUCHAY**

DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES, Co-directrice de thèse

**Monsieur PASCAL MAMASSIAN**

DIRECTEUR DE RECHERCHE, CNRS DELEGATION PARIS-CENTRE, Président

**Madame MARIE IZAUTE**

PROFESSEUR, UNIVERSITE CLERMONT AUVERGNE, Rapportrice

**Madame CHRISTINE BASTIN**

PROFESSEUR ASSISTANT, UNIVERSITE DE LIEGE - BELGIQUE, Examinatrice

**Monsieur ZOLTAN DIENES**

PROFESSEUR, UNIVERSITE DE SUSSEX - GRANDE BRETAGNE, Examineur

**Monsieur NATHAN FAIVRE**

CHARGE DE RECHERCHE, CNRS DELEGATION ALPES, Examineur





## *Abstract*

**M**etacognition and especially metacognitive judgments have been largely studied within separate cognitive fields such as episodic memory (metamemory) or visual perception (metaperception). Despite this historical tradition of evaluating metacognition in a disparate manner, similarities in methodological and theoretical frameworks can be observed and recent work compares metacognitive judgements across a variety of tasks (first-order task), proposing the idea that metacognition could be domain-general. This thesis focuses on the cue-utilization view stemming from the metamemory literature to explore the breadth of metacognition across two correlational and three experimental studies. In particular, we investigated whether people use a common resource in their metacognitive judgements across different types of first-order tasks and whether this resource is also shared across different metacognitive judgements. Moreover, we focused on the metacognitive cue of fluency as a potential domain-general cue in the formation of metacognitive judgements. Overall, our results suggest that whereas prospective judgements are domain-specific, retrospective judgements can be supported by a domain-general resource. The study of the involvement of fluency suggests differing influence of this cue on both the type of first-order task and the type of metacognition judgement. This suggests that fluency effects are less homogeneous than previously thought. In light of these results, we propose a novel view of metacognitive judgment formation in order to have a more unified view of metacognitive research. Finally, we suggest implications for both research on recognition memory and neuropsychological and psychiatric research.

**Keywords:** metacognition, metamemory, confidence judgments, prospective judgements, domain-general processes

## Resumé

La métacognition et en particulier les jugements métacognitifs ont été largement étudiés de façon séparée dans le domaine de la mémoire épisodique (métamémoire) ou de la perception visuelle (méta-perception). Malgré cette tradition historique d'évaluer la métacognition de façon disparate, des similitudes dans les cadres méthodologiques et théoriques peuvent être observées et des travaux récents proposent de comparer les jugements métacognitifs à travers une variété de tâches (nommés tâches de premier ordre) proposant l'idée que la métacognition pourrait être domaine-général. Dans cette thèse, nous nous sommes concentrés sur la « cue-utilization view » issue de la littérature sur la métamémoire afin d'explorer l'étendue de la métacognition via deux études corrélationnelles et trois études expérimentales. En particulier, nous avons cherché à savoir si les individus utilisent une ressource commune dans leurs jugements métacognitifs pour différents types de tâches de premier ordre et si cette ressource est également partagée entre différents jugements métacognitifs. De plus, nous nous sommes concentrés sur l'indice métacognitif de fluence comme un potentiel indice domaine-général dans la formation des jugements métacognitifs. Dans l'ensemble, nos résultats suggèrent que, alors que les jugements prospectifs sont domaine-spécifiques, les jugements rétrospectifs peuvent être sous-tendus par une ressource commune. L'étude de l'implication de l'indice de fluence suggère une influence différente de ce signal sur le type de tâche de premier ordre et le type de jugement de métacognition suggérant que l'effet de fluidité est moins systématique qu'on ne le pensait auparavant. À la lumière de ces résultats, nous proposons une nouvelle approche quant à la formation de jugements métacognitifs dans le but d'avoir une vision plus unifiée de la recherche sur la métacognition. Enfin, nous suggérons des implications à la fois pour la recherche sur la mémoire de reconnaissance et pour la recherche en neuropsychologie et en psychiatrie.

**Mots-clefs :** métacognition, métamémoire, jugements de confiance, jugements prospectifs, processus domaine-généraux

# Publications

Chapter 4 has been accepted as (this thesis version includes Supplementary results as a part of the Results Section):

**Mazancieux, A.**, Fleming, S. M., Souchay, C., Moulin, C. J. A. (2020). Is there a G factor for metacognition? Correlations in retrospective metacognitive sensitivity across tasks. *Journal of Experimental Psychology: General*. Advance online publication. <http://dx.doi.org/10.1037/xge0000746>

Chapter 5 has been accepted as:

**Mazancieux, A.**, Dinze, C., Souchay, C., Moulin, C. J. A. (2020). Metacognitive domain-specificity in feeling-of-knowing but not retrospective confidence. *Neuroscience of Consciousness*, 2020(1), niaa001.

Appendix A has been published as:

**Mazancieux, A.**, Souchay, C., Casez, O., Moulin, C. J. A. (2019). Metacognition and self-awareness in Multiple Sclerosis. *Cortex*, 111, 238-255

Appendix B has been submitted as:

**Mazancieux, A.**, Souchay, C., Casez, O., Moulin, C. J. A. (*in revision*). A multidimensional assessment of metacognition across domains in Multiple Sclerosis

Appendix C has been published as:

Bertrand, J. M.\* , **Mazancieux, A.\***, Moulin, C. J. A., BÉjot, Y., Rouaud, O., Souchay, C. (2019). In the here and now: Short term memory predictions are preserved in Alzheimer's disease. *Cortex*, 119, 158-164

\* *Equal contribution*

## *Acknowledgements*

**F**or some reason, this section appeared as one of the most difficult to write. Probably even more difficult than the thesis itself. I wanted to find the right words. I have to admit that I have never been the best for properly thanking people and expressing how much I care about them. But we only have one PhD thesis in our life – or at least I do not plan to write another one – and I have heard that we only live once, so I will make an effort this time. These last few days before the defence I received lots of absolutely lovely and supportive messages from all my friends and family. I have realised how lucky I am to be surrounded by so many well-wishers.

Luck has been central in this PhD. I am not trying to minimise the work I have done but, truly, luck has been insanely present these last 3 and a half years. Luck starts with Chris, my brilliant PhD supervisor. I guess that choosing to work with someone is always a blind bet. Five years ago, I did not realise I made a great one. Chris, you are a brilliant researcher. You are passionate, relentless, always questioning and looking for the best course of action. But you are also funny, amazingly supportive, and you truly care about people. All this makes you an all-round top notch supervisor. This is my opinion, but I am sure (I can even say I know) I am not the only one who thinks that.

As you said, we drank plenty of coffee. We also had plenty of talks about England and France comparisons, about the research world, society and politics, and about “how can we precisely translate this particular damn word.” And I really liked that (although I am still struggling to know if I prefer these talks or the coffee). I really feel like we worked together with our own organisational style that is based on having no deadline and just seeing “how things progress.” Sometimes I wonder how it worked but – magically perhaps – it did.

I was also very lucky to receive feedback and advice from Céline who co-supervised my PhD. Céline, I admire the woman that you are. You do so many great things for research and things “beyond research” that are so important to making a great working environment. I loved discussing with you about research organisation. I learned a lot. In essence, it was great to have you both as supervisors and I sincerely want to thank you for that.

I want to thank Steve Fleming for hosting me in his lab for two months, thanks to whom I had the amazing opportunity to go to the FIL in London which was a great experience in my PhD student life. I also wanted to thank all the members of my PhD jury.

These years would not had been the same without all my colleagues and friends in the LPNC. I am very grateful for the good spirit and the support of all the girls in the “old” E118 office: Alexia, my homonym Audrey, Emilie, H el ene, and Lisa. Girl power. You are all truly the best. I have always been happy to come to the lab every single day and it was mainly because I knew we would all meet up around a cup of coffee, laugh, grumble, and drink more coffee. I also wanted to thank all people in the “new” E118 office, Camille, M eline, Martin, and R emi. You are amazing officemates. You were so great with me when I bothered you with my PhD writing problems. I wish you all the best. Big thanks to Ladislav, our almost E118-mate. I am really grateful for all your advice, help, and interest in general. Thanks also to Am elie and Brice.

I am also grateful for the nice atmosphere in the “Moulin-Souchay lab”. It is fantastic teamwork. It truly makes sense for research. I particularly want to thank students I had the opportunity to directly co-supervise: Adeline, Anna, Carole, Chlo e, David, Esther, Stef, and Tiffany. Tiffany, you were the first student I co-supervised and we have now published together. I am really happy about it and I hope you enjoyed your job as a neuropsychologist. Stef, I co-supervised you during two years. I made you measure distances between points with me for an indecent amount of times. I am again sorry about that. But now I see you as a friend and I hope we will have the opportunity to work together again. On a more practical note, I want to thank Claire and David for helping me collecting data for Chapter 5 and Experiment 2 in Chapter 6.

More generally, thank you to all the members of the LPNC and especially the PhD students and postdocs. I think we are lucky to have such a great working atmosphere and it is thanks to all of you. Keep up the great work and all the exciting initiatives as always. You are the heart of the lab and you should all be proud of what you do. You have been an amazing support during these years. I also want to mention my international academia friends, Giuseppe, Joan, Katarina, and Kasia.

I want to thank my perfect flatmates, Emma and Loulou who got my head out of the academic world every evening. I cannot believe I will move out and leave you. Many thanks to all my closest friends outside academia, David, Lulu, Marion, Nina and all the others. Thanks for just being as you are. Special thanks to Jonathan. Finally, thanks to my parents for their support.

My last thank you goes to my friend Lisa who has been by my side since my very first day at the university. You were the first person I met there and I hope to keep you as a good friend my whole life. You are a brilliant, funny, and truly kind woman, researcher, climber, musician, cycling and beer drinking friend. I dedicate this thesis to you.





# Contents

<b>Abstract</b>	<b>ii</b>
<b>Resumé</b>	<b>iii</b>
<b>Publications</b>	<b>iv</b>
<b>Acknowledgements</b>	<b>v</b>
<b>List of Abbreviations</b>	<b>xvi</b>
<b>1 The diversity of metacognition</b>	<b>1</b>
1.1 How to define metacognition? . . . . .	2
1.1.1 Representational dimension: Meta-level and object-level . . . . .	2
1.1.2 Behavioral dimension: First-order and second-order . . . . .	3
1.1.3 Consciousness dimension: State of awareness . . . . .	5
1.2 Different forms of metacognition . . . . .	7
1.2.1 Direct tests of metacognition . . . . .	8
High-level metacognition . . . . .	9
Low-level metacognition . . . . .	11
1.2.2 Indirect tests of metacognition . . . . .	12
Evidence in human animals . . . . .	12
Evidence in non-human animals . . . . .	13
1.3 Summary . . . . .	15
<b>2 A comparison of the field of metaperception and metamemory</b>	<b>17</b>
2.1 Methodological comparison . . . . .	17
2.1.1 Measures of metacognition: direct tests . . . . .	18
2.1.2 The quantification of metacognition . . . . .	23
Metacognitive sensitivity . . . . .	23
Metacognitive bias . . . . .	29
2.2 Theoretical comparison . . . . .	32

2.2.1	Relationship between metacognition and task performance . . .	32
	Models of confidence formation . . . . .	33
	Theoretical accounts of metamemory . . . . .	35
2.2.2	Dissociation between metacognition and task performance . . .	37
	Behavioral and neuronal evidence . . . . .	37
	Post-decisional and second-order models of confidence . . . . .	39
	Cue-utilization in metamemory judgements . . . . .	41
2.3	Conclusions . . . . .	44
<b>3</b>	<b>Introduction to empirical chapters</b>	<b>47</b>
3.1	Investigating metacognition across domains . . . . .	47
3.1.1	The domain-generalty and domain-specificity debate . . . . .	47
3.1.2	Domain-generalty in metacognition . . . . .	48
3.2	Methodological considerations . . . . .	50
3.2.1	Objectives . . . . .	50
3.2.2	Material selection . . . . .	50
3.2.3	Data and statistical modelling . . . . .	51
3.3	Aims of the thesis . . . . .	54
<b>4</b>	<b>Is there a G factor for metacognition? Correlations in retrospective metacognitive sensitivity across tasks</b>	<b>56</b>
4.1	Introduction . . . . .	57
4.2	Method . . . . .	60
4.2.1	Participants . . . . .	60
4.2.2	Materials and procedure . . . . .	61
4.2.3	Data and statistical analyses . . . . .	63
4.3	Results . . . . .	66
4.3.1	Type 1 performance . . . . .	66
4.3.2	Confidence level . . . . .	67
4.3.3	Metacognitive efficiency . . . . .	68
4.3.4	Supplementary results . . . . .	70
	Preregistered analyses . . . . .	70
	Comparison between hierarchical and non hierarchical models .	74
4.4	Discussion . . . . .	78

<b>5</b>	<b>Metacognitive domain specificity in FOKs but not RCJs</b>	<b>84</b>
5.1	Introduction . . . . .	85
5.2	Method . . . . .	90
5.2.1	Participants . . . . .	90
5.2.2	Materials and procedure . . . . .	90
5.2.3	Data and statistical analyses . . . . .	91
5.3	Results . . . . .	93
5.3.1	Type 1 performance . . . . .	93
5.3.2	Metacognitive bias . . . . .	93
5.3.3	Metacognitive efficiency . . . . .	95
5.3.4	Exploratory analyses . . . . .	95
	Logistic regressions . . . . .	95
	Correlational analyses . . . . .	98
5.4	Discussion . . . . .	98
<b>6</b>	<b>Investigate ease-of-processing in metacognition for memory and visual perception</b>	<b>103</b>
6.1	Experiment 1: Perceptual fluency induces by prior exposure effects on metacognition for visual perception . . . . .	108
6.1.1	Overview . . . . .	108
6.1.2	Method . . . . .	110
	Participants . . . . .	110
	Material and procedure . . . . .	110
	Data and statistical analyses . . . . .	112
6.1.3	Results . . . . .	113
	First-order performance . . . . .	113
	Metacognitive efficiency . . . . .	113
	Metacognitive bias . . . . .	114
	Relationship with recognition memory . . . . .	115
	Exploratory analyses . . . . .	116
6.1.4	Brief discussion . . . . .	117
6.2	Experiment 2: The effect of conceptual fluency on metacognitive judgments for visual perception . . . . .	119
6.2.1	Overview . . . . .	119
6.2.2	Method . . . . .	120

	Participants . . . . .	120
	Material and procedure . . . . .	120
	Data and statistical analyses . . . . .	121
6.2.3	Results . . . . .	121
	First-order performance . . . . .	122
	Metacognitive efficiency . . . . .	122
	Metacognitive bias . . . . .	124
	Exploratory analyses . . . . .	124
6.2.4	Brief discussion . . . . .	126
6.3	Experiment 3: The word heuristic as conceptual fluency effect on metacog- nition for episodic memory . . . . .	126
6.3.1	Overview . . . . .	126
6.3.2	Method . . . . .	127
	Participants . . . . .	127
	Material and procedure . . . . .	128
	Data and statistical analyses . . . . .	128
6.3.3	Results . . . . .	129
	First-order performance . . . . .	130
	Metacognitive efficiency . . . . .	130
	Metacognitive bias . . . . .	131
	Exploratory analyses . . . . .	132
6.3.4	Brief discussion . . . . .	134
6.4	General discussion . . . . .	135
<b>7</b>	<b>General discussion and perspectives</b>	<b>138</b>
7.1	Summary of results . . . . .	138
7.2	Implications for research on metacognition . . . . .	141
7.2.1	Experience-based and information-based metacognition . . . . .	143
	Information-based metacognition . . . . .	143
	Experience-based metacognition . . . . .	145
	A proposal for metacognitive judgement formation . . . . .	147
7.2.2	Metamemory and metaperception . . . . .	149
	Qualitative differences . . . . .	149
	Using the meta-d framework in memory and perception tasks . . . . .	152
7.2.3	Implications for dimensions of metacognition . . . . .	153

The representational dimension . . . . .	153
The behavioural dimension . . . . .	154
The consciousness dimension . . . . .	155
7.3 Other implications . . . . .	157
7.3.1 Implication for memory research . . . . .	157
7.3.2 Implication for neuropsychology and psychiatry . . . . .	158
7.3.3 Conclusion . . . . .	161
<b>A A review of metacognition and self-awareness in Multiple Sclerosis</b>	<b>163</b>
<b>B A multidimensional assessment of metacognition across domains in Multiple Sclerosis</b>	<b>182</b>
<b>C In the here and now: Short term memory predictions are preserved in Alzheimer's disease</b>	<b>214</b>

# List of Figures

2.1	Example of the JOL paradigm . . . . .	19
2.2	Examples of the FOK paradigms . . . . .	20
2.3	Example of AUROC2 and calibration curves . . . . .	25
2.4	SDT for a discrimination task between stimuli S1 and S2 . . . . .	26
2.5	Calibration curves . . . . .	31
2.6	Diffusion model . . . . .	40
2.7	The cue-utilization approach on JOLs . . . . .	43
4.1	Summary of the four tasks . . . . .	63
4.2	Performance and confidence distributions . . . . .	68
4.3	Posterior distributions over $\mu$ meta- $d'/d'$ ratio . . . . .	69
4.4	Posterior distributions over $\rho$ parameters . . . . .	71
4.5	AUROC2 values per participant and per tasks. . . . .	74
4.6	Posterior distributions over $\rho$ parameters for simulated data with 40 trials	76
4.7	Posterior distributions over $\rho$ parameters for simulated data with 400 trials . . . . .	78
5.1	Procedure of Chapter 5 . . . . .	92
5.2	Type 1 performance of Chapter 5 . . . . .	94
5.3	Metacognitive bias of Chapter 5 . . . . .	94
5.4	Metacognitive efficiency of Chapter 5 . . . . .	96
5.5	Cross-task correlation for metacognitive efficiency . . . . .	97
6.1	Procedure of Chapter 6 - Experiment 1 . . . . .	111
6.2	Main results of Chapter 6 - Experiment 1 . . . . .	114
6.3	Additional results of Chapter 6 - Experiment 1 . . . . .	116
6.4	Procedure of Chapter 6 - Experiment 2 . . . . .	122
6.5	Main results of Chapter 6 - Experiment 2 . . . . .	123
6.6	Additional results of Chapter 6 - Experiment 2 . . . . .	125
6.7	Procedure of Chapter 6 - Experiment 3 . . . . .	129

6.8	Main results of Chapter 6 - Experiment 3 . . . . .	131
6.9	Additional results of Chapter 6 - Experiment 3 . . . . .	133
7.1	Proposed metacognitive judgement formation . . . . .	148



# List of Tables

1.1	Proposal of classification for types of metacognitive tests . . . . .	9
2.1	Type 2 SDT and classification of responses . . . . .	24
4.1	Cross-task performance correlations . . . . .	67
4.2	Cross-task confidence correlations . . . . .	68
4.3	Difference distributions between group-level meta-d'/d' estimates . . . . .	70
4.4	Group-level correlations in meta-d'/d' . . . . .	70
4.5	Pearson correlations for individual meta-d'/d' between pairs of tasks . . . . .	72
4.6	Pearson correlations for paired correlations of metacognitive bias . . . . .	73
4.7	Pearson correlations for paired correlations of AUROC2 . . . . .	74
4.8	Pearson correlations for individual Mratio for simulated data (40 trials) . . . . .	76
4.9	Parameter recovery for simulated data (40 trials) . . . . .	77
4.10	Pearson correlations for individual Mratio for simulated data (400 trials) . . . . .	78
4.11	Parameter recovery for simulated data (400 trials) . . . . .	79
5.1	Review of studies comparing episodic FOK and semantic FOK. . . . .	88
5.2	Difference distribution for each two-by-two comparisons . . . . .	95
6.1	Review of fluency effects on metacognitive judgements. . . . .	105
7.1	Summary of our main results . . . . .	140

# List of Abbreviations

**2AFC** Two alternative force-choice.

**ANOVA** Analysis of the Variance.

**AUROC2** Area Under the Type 2 ROC.

**eFOK** Episodic Feeling-Of-Knowing.

**EOL** Ease-Of-Learning.

**eRCJ** Episodic Retrospective confidence judgement.

**FOK** Feeling-Of-Knowing.

**HDI** Highest Density Interval.

**JAGS** Just Another Gibbs Sampler.

**JOL** Judgement-Of-Leaning.

**MCMC** Markov chain Monte Carlo.

**OSF** Open Science Framework.

**PFC** Prefrontal Cortex.

**RCJ** Retrospective confidence judgement.

**ROC** Receiver Operating Characteristic.

**RT** Reaction time.

**SDT** Signal Detection Theory.

**sFOK** Semantic Feeling-Of-Knowing.

**sRCJ** Semantic Retrospective confidence judgement.

**TOT** Tip-Of-the-Tongue.

## Chapter 1

# The diversity of metacognition

Imagine that you are enjoying holidays with your friends and decide to visit Edinburgh castle. When you enter this historic fortress, you are directly invaded by a powerful feeling of familiarity. It feels like you already have a prior experience with this place and with this exact location. At the exact same time of experiencing such a strong impression, a thought comes to your mind: "I have never been in Scotland before". Thus, you are immersed in a strange state where you ultimately know that this feeling of familiarity is erroneous. This kind of dissociative experience (also known here as a *deja-vu* experience) is a perfect example of the human ability (and perhaps not only humans, as we will see in this chapter) of self-evaluation.

Our cognitive functions such as perception, memory, or language are subject to errors or inaccuracies. One can experience hallucinations, retrieve false contextual details from a prior event, or be unable to remember the name of a famous actress. Moreover, and as in our Edinburgh example, we have the ability to be aware of such cognitive errors. When realizing that these percepts or memories are misleading, such self-assessments of our cognitive abilities refer to metacognition. By the use of the prefix 'meta', metacognition is a theoretical idea 'beyond' cognition and in that sense, [Flavell \(1979\)](#) defined it as 'cognition about cognition'. Metacognition is a broad construct. Definitions differ according to the field of research in which it is studied. Traditionally stemming from the learning and memory literature, metacognition was mainly studied in cases with direct applications. For instance, many studies study metacognition in the context of eyewitness memory or focus on the evaluation of self-knowledge in learners and even populations with memory impairments. More recently, metacognition has been also used to evaluate the state of consciousness associated with a cognitive function, especially in visual perception. Henceforth, in this guise, it has become a full-blown field and models of metacognition have been developed.

The study of metacognition is at a crossroads of different constructs and fields including psychology, cognitive science, neuroscience, philosophy, and economics. Thus, to answer to the question “What is metacognition?” one might consider its multifaceted nature. The current introduction will be based on three dimensions suggested by [Fleming, Dolan, and Frith \(2012\)](#) which are the representational dimension, the behavioural dimension, and the consciousness dimension. Then, we will develop different forms of metacognition in human and non-human animals according to how it is measured in laboratory.

## 1.1 How to define metacognition?

### 1.1.1 Representational dimension: Meta-level and object-level

The first models of metacognition focused on the idea that an internal observer allows the self-perception of performing a cognitive activity. Therefore, they assumed that metacognition has a representation dimension: the observer has a “model” of the current activity supposed to be “beyond” this activity. [Nelson and Narens \(1994\)](#)’s model of metacognition (1990) developed in the context of memory functions is a good example of such a representational view of metacognition. These authors distinguish two levels of processing. The object-level refers to the level of the cognitive activity (e.g., learning a maths lesson) whereas the meta-level is a representation of this activity which can be formulated in a propositional and self-referenced way (e.g., “I am very bad at maths”). To communicate, these two levels are linked by two types of metacognitive processes. Monitoring processes stem from the object-level and inform the meta-level about the actual state of the cognitive activity (e.g., “This particular equation is difficult to understand”). Once the meta-level has been updated, control processes have a direct action on the cognitive level by the use of adaptive strategies to improve the cognitive activity (e.g., “I will spend more time on this problem”). According to such a view of metacognition, there is a behavioural control loop (the monitoring-affects-control hypothesis ([Nelson & Leonesio, 1988](#))) that involves an explicit evaluation of the cognitive activity and a declarative implementation of strategies. An important feature of this model is that the meta-level and the object-level can operate simultaneously, which proposes that metacognition is independent from cognitive processes.

A striking example of this representational and dissociative structure is a phenomenon known as the Tip-Of-the-Tongue (TOT; [Brown & McNeill, 1966](#)). The TOT is a state in

which one cannot quite recall a word but has the knowledge that he or she knows the particular word. This phenomenon is therefore an experience of dissociation between the object- and the meta-level: the self-knowledge (i.e., “I know what the capital of Australia is”) is in contradiction with the cognitive performance (i.e., not recalling the answer). Other dissociations are known in the memory literature such as the *déjà vu* phenomenon mentioned above and where there is a “phenomenological experience of recognizing a current situation and the awareness that this feeling of recognition is inappropriate” (p.2; O’Connor & Moulin, 2010). Another practice example of the object- and the meta-level is given by Nelson (1996). In the sentence “This sentence contains three errors”, the spelling of “this” and “three” refer to two object-level errors and the meaning of the sentence represents a meta-level error. Thus, to be aware of the meta-level error, one must be aware of the object-level error suggesting a hierarchical structure in metacognition. However, as mentioned by Nelson (1996) after realizing the meta-error, one can realize that “there is a total of three errors after all” leading to the idea of a structure containing more than two levels with different monitoring-control loops. Metacognition has therefore a representational function that paradigms tend to measure using behavioural observation. The majority of research has been conducted on monitoring processes probably as they are seen as a starting point of the loop: there is a need for a self-evaluation to perform an adaptive behavioural control.

### 1.1.2 Behavioral dimension: First-order and second-order

Although a classical definition of metacognition involves a representational dimension of a cognitive activity, research has also shown the involvement of information directly stemming from cognitive activity (i.e., the object-level) without necessarily involving another level of representation. Traditionally, inferences about monitoring processes have been measured using subjective reports of introspection known as metacognitive judgments Nelson, 1990. These judgements are extensively used in the laboratory to study metacognition and refer to “decisions about decisions” (or “behaviour about behaviour, see Fleming et al., 2012). Experimentally, metacognitive judgements occur when two decisions are made: one known as first-order response refers to an answer in a cognitive task (e.g., to select with stimulus in the brightest) and the other or second-order response is a metacognitive judgement related to the first-order response (e.g., was the previous response correct?). Several metacognitive judgements<sup>1</sup> exist in the

---

<sup>1</sup>As metacognitive judgements are a main focus of this thesis, they will be presented in more detail in Chapter 2.

literature but the most commonly used are Retrospective Confidence Judgements (RCJs) which are an evaluation of the level of confidence in a previous decision.

In the memory literature, these judgements are seen as a measurement of monitoring processes. However, the “second-order” terminology is now preferably used since metacognition is studied in various cognitive domains beyond the memory domain (e.g., visual perception, decision making, reasoning, motor function, etc.). Moreover, it is assumption-free as second-order judgements can be performed using a range of processes (see Chapter 2). On one hand they can be the result of introspection by the use of metacognitive representations as a model of the current cognitive activity. Conversely and on another hand, they can use information directly stemming from the object-level without necessarily involving another level of representation. A good example of such absence of meta-level representation derives from the notion of uncertainty. Uncertainty can be defined as the variability or the reliability of a representation, a stimulus, or an outcome (Bach & Dolan, 2012). A certain amount of work has shown that uncertainty can be used in order to control behaviours such as modifying decision-making therefore acting as a metacognitive component. In some cases, these adaptive behaviours do not necessarily imply the existence of a meta-level in both human and non-human animals<sup>2</sup> (Kepecs & Mainen, 2012).

From this perspective, the existence of metacognition is suggested by the fact that one is able to produce accurate decisions. Metacognition has a function of evaluation of cognition with the idea that an appropriate evaluation will lead to an adaptive behaviour. In this sense, another important aspect is error detection; the ability to be aware of our own mistakes. This has been especially studied in the context of reaction time tasks when participants have to make decisions under time pressure which leads to an increase in the number of errors. Although initially studied separately, models and methods of confidence formation and error detection shared similarities leading to the idea that they are two outcomes underpinned by common metacognitive processes (Yeung & Summerfield, 2012). Here again, quantifying confidence or detecting an error have been explained as being based on the reliability of the signal during the decision process (akin to uncertainty) which does not necessarily need the involvement of a meta-level representation (see Section 2.1.2 “Low-level metacognition and implicit knowledge” and for a philosophical discussion about the involvement of meta-representations see Proust, 2007, and see section 2 of this Chapter).

---

<sup>2</sup>See section 2.2 of this chapter “Indirect tests of metacognition”.

### 1.1.3 Consciousness dimension: State of awareness

Nelson and Narens (1994) suggest that the main method for investigating monitoring is to ask people to give the output of their introspection. Although metacognition and introspection can overlap, they also differ to some extent. Overgaard and Sandberg (2012) propose a useful distinction between the two: a classification from the subjective point of view or from the functional point of view. Whilst both are second-order processes, metacognition is about cognition (as a function that can be either conscious or unconscious) and introspection is restricted to a conscious state. In this sense, they both have a higher-order function and one can see introspection as “a special case of metacognition” (p.1288, Overgaard & Sandberg, 2012). This suggests that metacognition involves conscious self-reflected mechanisms but also others processes that are not present during introspection.

Whilst it is beyond the scope of this thesis to propose explanations and a complete account of consciousness, it is important to note that the concepts of consciousness and metacognition are relatively close. For the purposes of this thesis, consciousness has been conceptualised as the classical notion of “having subjective experiences” (Frith, 2019). Similarly, metacognition and particularly metacognitive judgements (see next section and Chapter 2) are self-evaluation which also has a subjective value. Thus, these judgements can be seen as reports of the content of subjective experiences and are even used as a measure of consciousness (e.g., Seth, Dienes, Cleeremans, Overgaard, & Pessoa, 2008). Some implications for our understanding of consciousness will be discussed in Chapter 7, however it is not a direct aim of this thesis to use metacognition to examine the nature of consciousness.

Although there is a close relationship between metacognition and consciousness, these are nonetheless dissociable to some extent. Metacognition refers to consciousness in terms of the “access of consciousness” that is the ability to be aware of conscious states as a knowledge but do not account for “the phenomenology of consciousness” that refers to subjective experience as a feeling (Block, 2011). It suggests that one can be aware of a state (e.g., I am in a sad mood, I am seeing a green colour) whilst having no metacognitive process involved in this state of awareness (e.g., Am I sure this is green?). Similarly, if consciousness does not necessary imply metacognition, several works have also shown that metacognitive behaviours can occur at the fringe of consciousness. To investigate this issue, an interesting study case is blindsight. The blindsight phenomenon occurs when patients are able to detect and identify visual stimuli

during forced-choice experiments in the absence of conscious awareness of these stimuli when they are presented in the blind visual field (e.g., Poppel, Held, & Frost, 1973). In their commentary, Kentridge and Heywood (2000) report the case of a patient that had to detect the presence or the absence of a stimulus. During their experimental procedure they added visual cues which indicate the likely location of the stimuli and which result in a decrease in reaction time to detect stimuli. When the patient was told that in 2/3 of cases the cue indicated the correct target location, his reaction time to detect the target was shorter when the cue was indeed correct, even in his blind field. Interestingly, when instructions changed (when the cue indicated the correct target location in only 1/3 of cases), he was able to have shorter reaction times for incongruent cue/target trials after several series of trials. This means that even if the explicit information was not helpful for the patient at first, he was still able to adapt his behaviour to the new situation through implicit learning. Moreover, the patient was not aware of this change as he reported being unaware of all targets and cues.

As also stressed by Reder (1987), this example supports the idea that some strategy selection can occur without any awareness of having used such strategies. Similarly, patients with a lack of consciousness as in a vegetative state are still able to have adaptive behaviours such as learning (Bosco et al., 2009). The same metacognitive function in the absence of conscious awareness has also been highlighted with error detection (Charles, Van Opstal, Marti, & Dehaene, 2013). This range of evidence supports the idea that metacognition is not necessarily a conscious process (Spehn & Reder, 2000). As for some second-order behaviours, this is incongruent with the idea of a meta-level that would consciously monitor and control an object-level. It rather suggests the existence of a loop of cognitive regulation without the involvement of a meta-representation and the conscious awareness of monitoring and control.

Overall, although metacognition and conscious awareness overlap to some extent, some manifestations of conscious awareness occur without a metacognitive value (e.g., the phenomenology of consciousness) and metacognitive behaviour does not necessarily involve consciousness (e.g., unconscious cognitive control). Overall, it seems that metacognition can occur at different levels: either with the involvement of introspection and self-awareness at the extreme case or as a result of changes in behaviours without conscious awareness. In this thesis, we appealed to these three conceptions of metacognition. Although primary work on metacognition has been developed in the context of memory with Nelson and Narens (1994) framework, here we point out several limits to this view of metacognition. We will detail more different forms of metacognition than



we have drawn out so far, and will classify them according to methods that are used to measure them.

## 1.2 Different forms of metacognition

Because of its multidimensional aspect, the expression of metacognition can occur in various forms. Whilst the classical metacognitive framework distinguishes between monitoring and control, we will present here metacognition as knowledge that can be used either for monitoring or for control that is “what people know about cognition in general” (p. 290; [Koriat, 2007](#)). As with many concepts in psychological science, these forms of metacognition can map onto at least two classical distinctions (for an example in the memory literature see [Richardson-Klavehn & Bjork, 1988](#)). The first focuses on the type of processes that underlie metacognitive behaviours, either explicit knowledge or implicit knowledge. Knowledge is said to be explicit when “there is an internal state whose function is to indicate the content of the knowledge” (p. 737, [Dienes & Perner, 1999](#)). Here we propose that some processes use explicit metacognitive knowledge when a metacognitive state (e.g., being confident in an answer) indicate the content of the knowledge that can be communicated by explicit statement (“I am confident *because* [...]”). Other used knowledge would however be implicit (e.g., the fluency heuristic, see Chapter 2).

The second distinction relies on the ways metacognition is measured, either direct or indirect testing. In direct tasks, participants have instructions to have a self-reflection about a mental process. Conversely, in other paradigms participants are not directly asked to introspect and evaluate themselves. In these indirect tests, other behaviours are used to infer metacognitive abilities. This direct/indirect distinction maps onto the monitoring and control processes described by [Nelson and Narens \(1994\)](#). As monitoring is an online self-assessment during a cognitive task, it necessarily implies a direct test of metacognition. On the contrary, regulation of cognitive activity and change in strategies (referring to control processes) are mainly indirect measures of metacognition as they rely on changes in behaviours. Although these two distinctions have overlaps (words have been often used interchangeably<sup>3</sup>), we suggest that some forms of metacognition based on explicit knowledge can occur during indirect testing (although the above indirect tests are mainly underpinned by implicit knowledge see Table 1.1).

---

<sup>3</sup>Note that French translations of these notions are even more confusing as both monitoring and control have been translated in “contrôle”. The word “surveillance” also occurred sometimes but it is also a synonym of control.

Conversely, as we will see direct tests of metacognition benefit from the influence of both explicit and implicit knowledge. Direct tests of metacognition will be detailed in Chapter 2, as they are the core measures used in this thesis.

The implicit/explicit distinction in metacognition has been similarly suggested as a dual view of metacognition. The classical ‘dual process’ or ‘dual system’ framework has been extensively developed in cognition (Evans, 2008; Stanovich, 1999). It distinguishes between a “system 1” or “type 1 process” referring to processes that are supposed to be automatic, fast, non-conscious, associative, and effortless and a “system 2” or “type 2 process” that is supposed to include process that are more controlled, slow, conscious, declarative, and require more effort. According to Shea et al. (2014), one can distinguish between the “intra-personal cognitive control” as a “system 1” metacognition and the “supra-personal cognitive control” as a “system 2” metacognition.

Finally, we make a third distinction in terms of access of conscious states that are involved in these metacognitive behaviours. To do so, we use the distinction made by Schooler (2002) who distinguishes between meta-conscious, conscious, and non-conscious processes. Nonetheless, because consciousness is often proposed to refer to subjective experiences which has the subject as referential (e.g., Frith, 2019), we will not use the term “processes” but rather the term of “state” to refer to this third category. As such, we distinguish between meta-conscious, conscious, and non-conscious states. This third classification overlaps with the explicit/implicit distinction and the direct/indirect distinction without matching perfectly with one or the other. Table 1.1 summaries our proposal.

### 1.2.1 Direct tests of metacognition

Direct tests of metacognition are situations where participants are directly asked to have a self-reflection about a cognitive activity and refer mainly to metacognitive judgements or decisions. As these judgements are the main focus of this thesis, the different existing paradigms will be detailed in Chapter 2 and we will here focus on processes that are supposed to be involved in such judgements and decisions. As suggested by Arango-Muñoz (2011) one can distinguish between two types of metacognitive processes that are at play in direct measures: high-level and low-level metacognition (see also in Chapter 2 the distinction between information-based and experience-based metacognition, Koriat, 1997. These two levels of metacognition have also been linked to the “system 1”/“system 2” distinction described above (Arango-Muñoz, 2011; Dokic, 2014).

Table 1.1: Proposal of classification for types of metacognitive tests according to the type of test (direct and indirect), type of knowledge (explicit and implicit), and state of consciousness (meta-conscious, conscious, non-conscious). The different metacognitive judgements will be detailed in Chapter 2.

DIRECT TESTS		INDIRECT TESTS	
<b>Metacognitive judgements</b>		<b>Other tasks</b>	
Confidence judgements		Opt-out paradigms	
Judgements-of-learning		Re-study choice	
Feeling-of-knowing		Study time allocation	
Ease-of-learning		Post-decision wandering	
Confidence forced-choice		<b>Other manifestations</b>	
		Facial expressions	
EXPLICIT KNOWLEDGE		IMPLICIT KNOWLEDGE	
<b>High-level metacognition</b>		<b>Low-level metacognition</b>	
Information-based		Experience-based	
Beliefs, theories, concepts		Epistemic feelings	
"System 2"		"System 1"	
slow and analytic		fast and based on heuristics	
META-CONSCIOUS	CONSCIOUS		NON-CONSCIOUS

### High-level metacognition

High-level metacognition refers to the highest degree of self-reflection about our cognitive abilities. Therefore, it involves a meta-level, a re-representation of the cognition, and theories or beliefs about the nature of the cognition. In that sense, it is associated with meta-consciousness that is the "explicit awareness of the content of consciousness" (p.339; [Schooler, 2002](#)). This type of metacognition can be easily pictured in the following example: a student has to learn a list of words in a foreign language. At some point during the learning process, she notices that a word is more difficult to learn. Such evaluation is a metacognitive judgement as it refers to a representation of the cognitive activity and can be based on belief related to memory (e.g., words with more than three syllables are difficult to learn). Therefore the thought "I believe that this word is more difficult to learn" has a meta-representational and conceptual structure.

High-level metacognition is underpinned by explicit knowledge that is supposed (as

with all explicit knowledge) to be verbally accessible and reportable (Dienes & Perner, 1999). Such knowledge has a function of representation of the content and can be intentionally used in metacognitive behaviours such as in the above example. According to Flavell (1979), this knowledge or theory about cognition (“metacognitive knowledge”) is related to the task (e.g., “recognizing a word is easier than recalling it”), the strategy (e.g., “I will remember better if I write the words I have to learn”), and the person (e.g., “I live in Europe so I know more countries in Europe than states in the United States of America”). This last class of belief is interesting because Flavell (1979) suggests that it can refer to both the subject and other people. Therefore, there is a co-existence of self-oriented and other-oriented theories. According to this view, high-level metacognition can be seen as a self-oriented mindreading ability Carruthers (2009). Mindreading is this capacity to “represent the mental states of the people around us” (p.121; Carruthers, 2009) and allows the ability to have a theory of mind. Studies in infancy have shown that there is a developmental co-occurrence for mindreading and high-level metacognition (see Gopnik, 1993). The idea that high level metacognition has similarities with mindreading is also supported by neuroimaging findings. Vaccaro and Fleming (2018) performed a meta-analysis of 47 MRI studies using metacognitive judgements and found common engagement of regions in the prefrontal cortex when they compared their work to meta-analyses of the neural correlates of mindreading.

As these beliefs and theories are also about the subject herself, metacognitive judgements also involve self-awareness and self-efficacy (Bandura, 1977; Borkowski, Carr, Rellinger, Pressley, et al., 1990). There is a self-perspectival aspect to this type of metacognition that implies self-consciousness (Metcalf & Son, 2012). According to Rosenthal (2000) and the “high-order thought” (HOT) theory, high level metacognition allows consciousness. As it states: “one is conscious of one’s conscious states because every such state is accompanied by a high-order thought to the effect that one is in that state” (p.270-271). From our point of view, high-level metacognition pertains to meta-conscious states as it involves a meta-representational structure. On the contrary, from a HOT point of view, this type of metacognition actually allows conscious states. Again, it is not a direct aim of this thesis to use metacognition to examine the nature of consciousness, however we will discuss further implications in terms of consciousness in the General Discussion (Chapter 7).

Finally, and because of its relationship with self-awareness, metacognitive tasks have been largely used with neurological and psychiatric populations (for a review of metamemory in neurological population see Pannu & Kaszniak, 2005). It is a useful framework

in order to explore the degree of disease and symptom awareness that these patients<sup>4</sup> have. Metacognitive deficits have been found in patients with psychiatric symptoms (see [Hoven et al., 2019](#) for a review) and multiple metacognitive processes emerge according to psychiatric symptoms dimensions in healthy adults ([Rouault, Seow, Gillan, & Fleming, 2018](#)).

### Low-level metacognition

In contrast with high-level metacognition, low level metacognition does not depend on conceptual representations and self-reflection but rather on subjective feelings. These feelings can be feelings of doubt, of knowing ([Hart, 1965](#)), of certainty ([Ha, Haury, & Nehm, 2012](#)), of rightness ([Thompson & Morsanyi, 2012](#)), and of familiarity ([Whittlesea, 2001](#)). In terms of access to consciousness, low-level metacognition has a particular status where the process involved in subjective feelings are largely unconscious, however the output of such processes result in a conscious state in a way that it is accessible to the person but does not involve reflexive cognition ([Koriat, 1993](#)). Low-level metacognition is therefore based on processes based on implicit knowledge (such as heuristics) but the feeling is conscious and these experiences “allow some degree of personal control over processes that would otherwise influence behaviour directly and automatically, outside the person’s consciousness” (p.315; [Koriat, 2007](#)). Therefore, low-level metacognition involves conscious states. These subjective feelings are named “epistemic feelings” ([de Sousa, 2008](#)) or “noetic feelings” ([Dokic, 2014](#)) as they have an informative value about the state of the system. As suggested by [Arango-Muñoz \(2011\)](#), “the feeling itself is metacognitive in a sense of being directed towards a mental disposition [...], but the content of the epistemic feeling that determines decision-making is non-conceptual and thus not meta-representational” (p.77). Conversely, these feelings are based on heuristics that are most of the time non accessible to participants (e.g., the fluency heuristic, which is manipulated experimentally in Chapter 6).

Nonetheless, as with beliefs and theories, these feelings have been shown to influence subjective report of metacognition (e.g., experience-based metacognition, [Koriat, 2007](#); see Chapter 2) and feelings seem to be a basis for judgements in a broader sense ([Greifeneder, Bless, & Pham, 2011](#)). Therefore, direct tests of metacognition such as

---

<sup>4</sup>Appendices focus on this question. Appendix A provides a brief overview of the use of the metacognitive framework in neurological populations and reviews work conducted during the thesis related to metacognition and self-awareness in Multiple Sclerosis. Appendix B proposes an experimental protocol that investigates metacognitive judgement in Multiple Sclerosis across memory and executive functions. Appendix C focuses on metacognitive judgements for short-term memory in Alzheimer’s disease.

metacognitive judgements are based on both high-level and low-level metacognition. For instance, a judgement of confidence can be the result a feeling of rightness and a conceptual self-evaluation of the situation based on beliefs and theories. Because the two types of metacognition occur in metacognitive judgements, some patients can have a pattern of performance where some judgements are impaired despite other are spared as they can be performed using mainly low level metacognition (e.g., fractionation of metacognition in Alzheimer's disease, [Souchay, 2007](#)).

## 1.2.2 Indirect tests of metacognition

In indirect tests, participants are not directly asked for a self-evaluation of their own function but measure other behaviours that are used to infer metacognition. These tests have been largely used in situations when verbalisation is not (or is less) possible, such as with children, populations with cognitive impairments, and in non-human animals studies. We propose that these tests are mainly driven by implicit knowledge. It is a form of metacognition that [Proust \(2003\)](#) calls the “know-how to decide” and is proposed to generate conscious feelings, but the latter remain epiphenomenal ([Dokic, 2014](#)).

### Evidence in human animals

Indirect tests of metacognition in humans have been often presented as measure of metacognitive control because they refer to the way participants modify their behaviour in the task. However, as we have suggested in the beginning of this section, in order to have efficient control, one must have an accurate monitoring (even if it is implicitly represented and sometimes non-conscious). In these tasks, we can infer that participants have some knowledge about a particular aspect of the task in their behaviour. Mainly used in visual perception tasks, authors sometimes focus on opt-out paradigms ([Kiani & Shadlen, 2009](#)). In these tests, participants have the opportunity to not responding in a trial if they are not confident enough in their choice. Therefore, metacognition is embedded in the first-order decision and not performed sequentially as an explicit judgement. In memory, when there are two tests using the same material, it has been shown that participants allocate more time to re-study an item that they did for previously unrecalled item leading to the idea that they have some knowledge about previous failure ([Son & Metcalfe, 2000](#)).

Most of the time, the access to consciousness in these tasks is limited. In humans, successful performance on these tasks can be associated with both conscious states

and unconscious states. In this later case, [Schooler \(2002\)](#) proposes that it is a type of tacit-monitoring which is a monitoring of a conscious state (e.g., a perceptual decision) that occurs non-consciously, continuously, and is involved in checking for failures. However, some of these tests can also involve conscious states such as feelings. As with low level metacognition, these feelings are conscious although processes underpinning these feelings are based on implicit knowledge (and are not available for conscious report).

Other more spontaneous behaviours can be used to infer the presence of metacognition in humans. A useful framework to investigate this question is to focus on patients that exhibit self-awareness disorders. With these patients there is often a classical pattern where they exhibit impairments in direct tests and preserved in more indirect paradigms. Thus, patients with Alzheimer's disease have shown some deficits<sup>5</sup> in metacognitive judgements (see [Souchay, 2007](#), for a review) despite other manifestations of implicit metacognitive knowledge. Whilst overestimating their performance on a memory task, Alzheimer's disease patients exhibit congruent failure/success facial expressions as it is the case for controls ([Mograbi, Brown, Salas, & Morris, 2012](#)). Similarly, if patients with hemiplegia that deny their motor deficit are presented heavy objects that usually need two hands to be carried, it is a means of assessing indirectly if their are aware of their difficulties. Some patients with such difficulties are able to switch their hand to the midpoint of the object which is necessary to correctly carry the object using one hand ([Moro, Pernigo, Zapparoli, Cordioli, & Aglioti, 2011](#)). This adaptive behaviour suggests that patients have some implicit knowledge about their deficit despite being able to verbally and explicitly reporting them. To account for this phenomenon, [Mograbi and Morris \(2013\)](#) labelled this "implicit awareness" that they define as an "indirect demonstration of some level of knowledge about a deficit" (p. 181). Overall, indirect tests of metacognition in humans allow the assessment of metacognitive abilities with non-verbal populations (e.g., preverbal infants, [Goupil & Kouider, 2016](#)) or patients for whom direct tests are less appropriate.

### **Evidence in non-human animals**

A large body of work focuses on the assessment of metacognition in non-human animals. As these studies exclusively focus on behavioural observations, they necessarily imply the use of indirect tests. We suggest that these studies mainly reflect the use of

---

<sup>5</sup>These deficits relate to notion of accurate (or rather inaccuracy in the present case) of metacognitive judgements that will be developed in Chapter 2.

implicit metacognition however we let as an open question the possibility that some manifestation of explicit components occur during these indirect tasks.

One of the first studies focusing on this question was carried out with dolphins ([Smith et al., 1995](#)) in a tone discrimination task. Dolphins had to categorize a sound as either a 2,100 Hz tone or another tone, however they could also use a third option when they were not sure of the answer. As this third option was chosen for the most difficult trial (tone that where close to 2,100 Hz), the authors suggested that dolphins were able to evaluate their level of confidence in a response. This paradigm has been extensively used in animal metacognition studies, and is a form of opt-out paradigm. In these protocols, participants are typically asked to make simple decisions. When they are uncertain in their answer, they can choose to skip the trial. This paradigm is especially interesting as it has been used in both humans ([Gherman & Philiastides, 2015](#)) and non-human animals. For instance, macaque monkeys were presented dot motion stimuli and had to categorise the direction of the motion ([Kiani & Shadlen, 2009](#)). As for dolphins, they selected the opt-out option for stimuli with a higher visual noise level (stimuli with the most uncertainty). Critically, monkeys were able to adapt their behaviour according to the situation. The opt-out option was presented in only half of the trials and performance was lower when the opt-out was not available suggesting that monkeys did not just categorise stimuli with high noise as a third category.

An interesting case is whether non-human animals are able to judge the quality of their memory. Because such evaluations are about an internal state it is therefore less susceptible to suffer from the “third option” categorisation as is the case in visual perception. In a perceptual task it is possible that the animal uses surface characteristics of an “objective difficulty” of stimuli (i.e., visual noise level) to categorise it in a dimension which is about the percept rather than the difficulty itself. As such, they would categorise stimuli with higher noise as a third type of stimuli. In memory, this is presumably less of an issue since the object of the evaluation is not a stimulus in the environment to be classified, but the outputs of a cognitive process. [Kornell, Son, and Terrace \(2007\)](#) used an animal version of post-decision wagering paradigm for a memory recognition task. This study is particularly interesting as the first-order task is memory and that the paradigm used is not likely to suffer from the third-option categorisation. Rhesus monkeys were presented samples of pictures to learn. After this presentation, nine pictures appeared on the screen containing one that had been presented before and monkeys had to identify by touching the screen. Then, monkeys had to choose between two options. The first one was bet with high risk where monkeys received three tokens



that they can exchange for food is the answer was correct. In the case of incorrect responses, they lost three tokens. In the second option, the bet was a low risk bet where monkeys received a small (one tokens) but sure reward. Results revealed that monkeys chose the high risk bet more often when their response was correct in comparison to the low risk bet suggested an abilities to accurately evaluate their memory.

Although the literature in animal metacognition is growing, it is often debated in the same manner as more non-conscious form of metacognition in humans (tacit-monitoring). For instance, it has been suggested that associative learning can also explain behaviours described above and does not necessary therefore imply metacognition (Smith, Beran, Couchman, & Coutinho, 2008). However, as noted by Kentridge and Heywood (2000), “implicit learning of a novel schema may not involve metacognitive regulation per se. Substitution of one automatic process by another as a result of the inadequacy of the former as circumstances change does, however, clearly involve metacognition” (p. 308). The idea that metacognitive processes might be underlined by a form of implicit learning raises the question of the kind of behaviours we shall consider as metacognitive behaviours. Thus, these behaviours do not need an awareness level and are consider as anoetic metacognition that is stimulus-driven (Metcalf & Son, 2012). Even though they might imply such awareness, the parsimony principle leads us to consider them as mainly non-conscious.

### 1.3 Summary

Research to date has shown the diversity of metacognition. Fundamentally representational and associated with access to consciousness, behavioural evidence has also shown different forms of metacognition. As such, some behaviour that do not necessarily imply high-order representations are also metacognitive in essence. In that sense, metacognition pertains to “moderate representationalism” in a sense that metacognition can “represent a first-order state, but without representing the fact that this state has a certain representational function” (p. 6, Beran, Brandl, Perner, & Proust, 2012). The cognitive state is therefore implicitly represented though the use of heuristics (e.g., fluency, see Chapter 5) that are largely non-conceptual. Moreover, second-order judgments are about internal states but do not necessarily imply the use of a self-reflexion involving self-consciousness as these can be driven by implicit knowledge and heuristics (low level metacognition). By its functional definition of monitoring and control of cognitive activity, metacognition results in a large range of judgments and behaviours.

These can be based on either explicit knowledge with the involvement of introspection and self-awareness in extreme cases or can be based on more implicit knowledge as a result of changes in behaviours without conscious awareness.

Other useful classifications of metacognition have been made. For instance, [Metcalf and Schwartz \(2016\)](#) highlight the discrepancy between manifestations of metacognition in the laboratory (the way it is studied) and manifestations in real life (which they call spontaneous metacognition). They argue that whilst the study of metacognition mainly focuses on explicit reports about the monitoring of responses that can be either correct or incorrect, metacognition that occurs spontaneously “indicate[s] that the knowledge base is wrong, that one has misunderstood, that one does not know what one thinks one knows, or one is unable to smoothly and fluently retrieve the solution one needs”. This distinction echoes an existing distinction in memory between voluntary explicit memory and involuntary explicit memory ([Richardson-Klavehn & Bjork, 1988](#)). It directly raises the question of the function of metacognition in daily life that can be seen as a signal supposed to inform the cognitive system of a specific failure.

This thesis focuses on direct tests of metacognition. As we will see in chapter 2, these tests and especially metacognitive judgements have been particularly studied in the metacognition for memory tasks and visual perceptions tasks namely metamemory and metaperception. We will investigate metacognitive judgements across different cognitive domains (e.g., visual perception, memory, etc.) in order to observe if common resources are involved in different domains for the same measure (Chapter 4) and for several measures (Chapter 5). Then, we will turn to processes that are involved in these direct tests and will investigate the role of heuristics and especially the fluency heuristic (low level metacognition) as a potential domain-general process (Chapter 6; for an introduction to the domain-generality and domain-specificity debate see Chapter 3 Section 1).

## Chapter 2

# A comparison of the field of metaperception and metamemory

Metacognition is a broad construct and can therefore be applied to different cognitive domains<sup>1</sup> (e.g., memory, visual perception, language, etc.). In the memory domain, metacognition (metamemory) developed from direct applications. For instance, many studies evaluate confidence in the context of eyewitness memory (e.g., [Perfect & Hollins, 1996](#)) or focus on the evaluation of self-knowledge in learning with students (e.g., [Veenman, Van Hout-Wolters, & Afflerbach, 2006](#)) or the degree of disease awareness in patients (e.g., Alzheimer’s disease; [Souchay, 2007](#)). More recently, second-order behaviours have also been used to evaluate the state of consciousness associated with a cognitive function especially in perception (metaperception; e.g., [Cheesman & Merikle, 1986](#)). From this point of view, measuring metacognition allows a better understanding of the level of consciousness people have about their cognition and how it can be used to regulate cognitive activity. As such, two fields of metacognition that we aim to compare in this chapter emerge: metamemory and metaperception. Because this thesis focus on potential common metacognitive resources involved in different domains, we propose both a methodological and a theoretical comparison of these two fields of metacognition. In particular, we aim to produce a synthesis of metamemory and metaperception given that these two literatures have developed somewhat separately.

### 2.1 Methodological comparison

We distinguish two methodological aspects. Firstly, we label “measures of metacognition” what is recorded in the participant’s behaviour directly in relation with the task. We only report here direct tests of metacognition. Secondly, we refer to “quantification

---

<sup>1</sup>See the notion of cognitive domain in more details in Chapter 3 section 1.

of metacognition” for metacognitive indices that can be calculated from these measures namely metacognitive sensitivity and metacognitive bias.

### 2.1.1 Measures of metacognition: direct tests

These direct tests of metacognition are mainly metacognitive judgements and metacognitive decisions. Metacognitive judgements are termed second-order behaviours as they are ‘about’ a first-order task. Thus, they are performed during a cognitive task (e.g., a memory task, a visual perception task, etc.) and for each trial of the task. The metamemory literature distinguishes two types of metacognitive judgements according to the moment when they are made (e.g., [Dunlosky, Mueller, & Thiede, 2018](#)): prospective judgements are a prediction of a future performance whereas retrospective judgements focus on a previous performance.

An extensively-studied type of prospective judgement in memory is Judgements-Of-Learning (JOL; [Arbuckle & Cuddy, 1969](#)). JOLs are performed in episodic memory tasks during the encoding phase (Figure 2.1). In this paradigm, participants have typically to learn pairs of words; a cue that will be presented to the participant during the recall phase and a target that they will have to remember. For each pair of words, participants have to make a JOL referring to their ability to remember the target in the future recall test. Their response can be either dichotomic (“yes, I will recall the target” or “no, I will not recall the target”) or using a several-point certainty scale from ranging from 0% (“I am certain that I will not recall the target”) to 100% (I am certain that I will recall the target). Moreover, these judgements can be performed either when participants encode the to-be-remembered stimuli (immediate-JOLs) or after a first encoding (delayed-JOLs).

Another type of prospective metacognitive judgement that has been largely studied in the metamemory literature is the Feeling-Of-Knowing (FOK). In experiments, FOKs refer to the predictions of future stimulus recognition when this stimulus has not been recalled. They are a special kind of judgement in that they are cued by a retrieval attempt, but pertain to future performance, being therefore defined as prospective judgements. In the episodic FOK (eFOK) paradigm ([Schacter, 1983](#); [Souchay, Isingrini, & Espagnet, 2000](#)), participants typically learn cue-target paired-words. In a recall phase, a cue word is presented and the participants have to recall the target word. In the case where participants are not able to remember the stimulus, they predict if they think they will be able to recognize this particular word amongst a set of words. This prediction is the FOK. Finally, participants perform the recognition task. In the semantic FOK (sFOK)

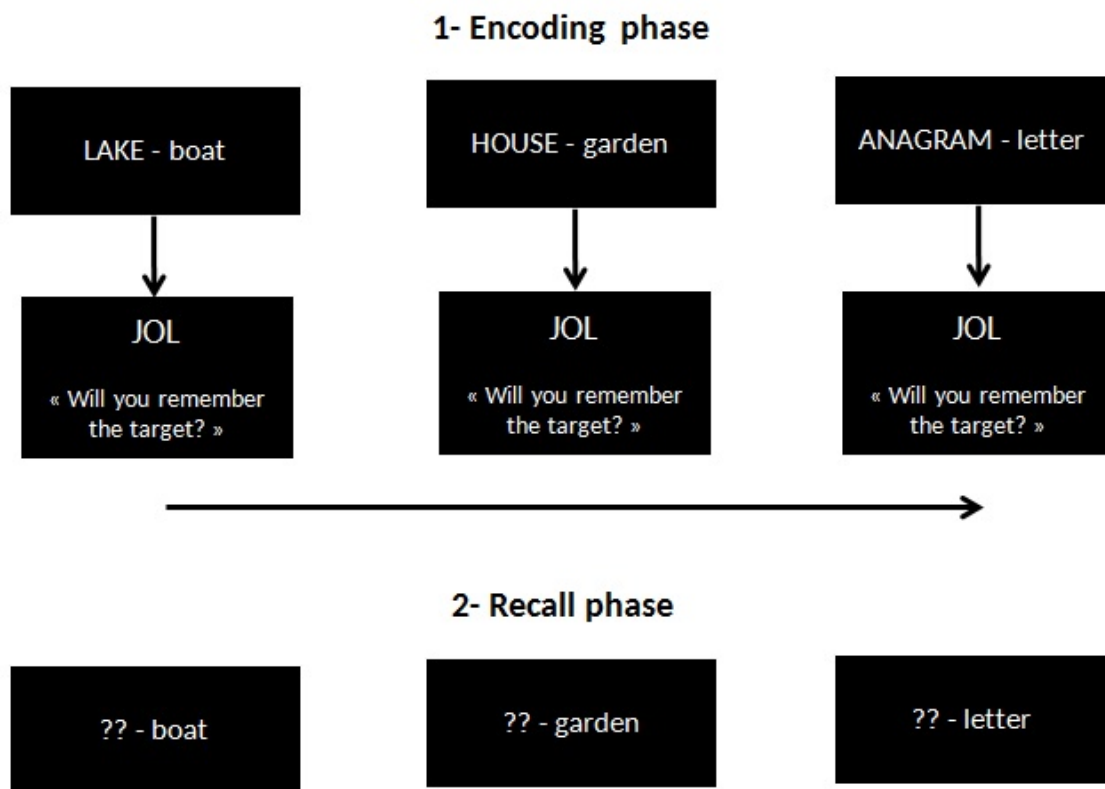


Figure 2.1: Example of the JOL paradigm. Participants learn pairs (e.g. LAKE – boat, and are later cued with the word ‘boat’ and are asked to predict their later performance.

paradigm (Hart, 1965; Nelson, 1990) instead of learning paired-words, word definitions or general knowledge questions are presented to the participants and they have to recall the word referring to this definition. As in the eFOK task, they make FOKs if they are not able to find the correct word, and perform a recognition task (see Figure 2.2 for an example of both paradigms). Although often classified as a prospective judgement, they are made after a retrieval attempt therefore having a retrospective dimension.

A last class of prospective judgements in metamemory is the Ease-Of-Learning judgement (EOL; Underwood, 1966). In this less-utilised paradigm, participants typically have to say how easily it will be to learn a to-be-remembered item. Therefore, these judgements are made before or during the encoding phase.

Contrary to prospective judgements, there is only one type of retrospective judgement which is the RCJ. This is used in both the metamemory and the metaperception fields. RCJs refer to the level of confidence that a participant has in a given answer using a multiple-point scale. As they can be performed for any kind of answer they have been

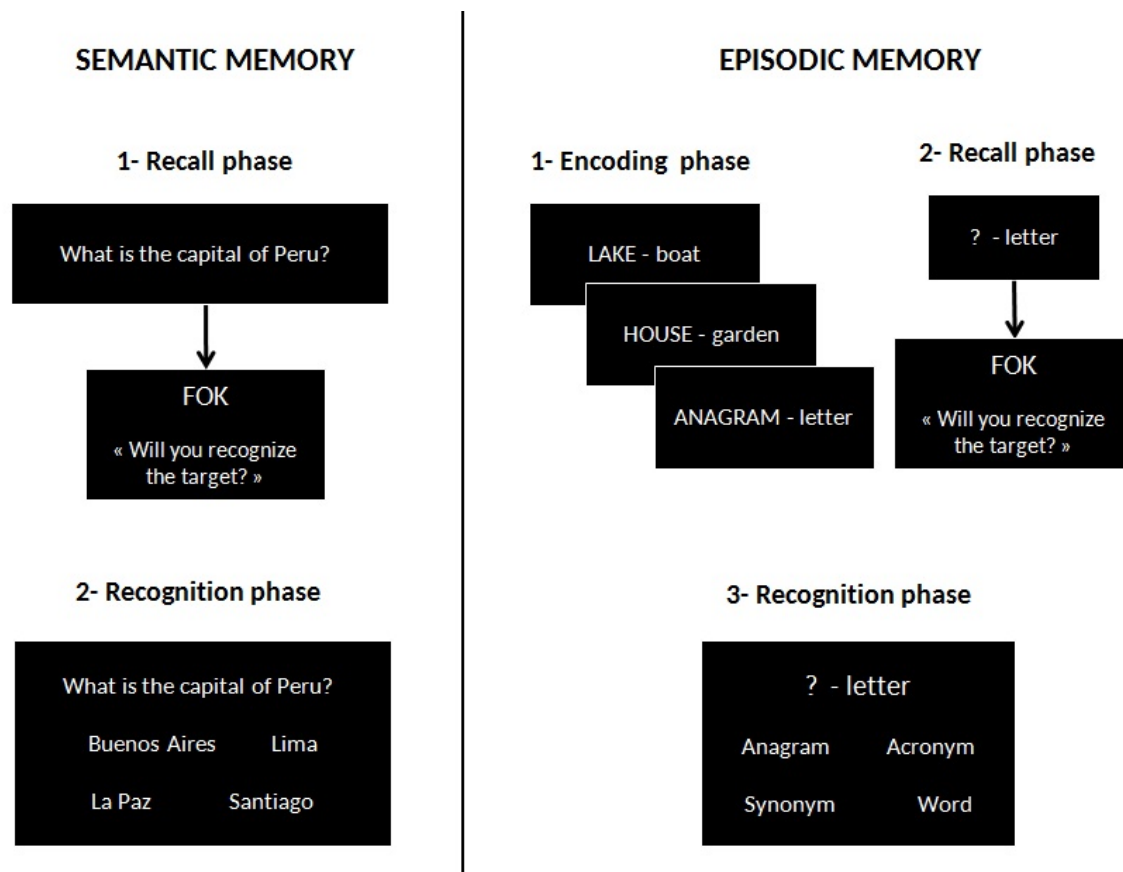


Figure 2.2: Examples of eFOK and sFOK paradigms.

the main judgment used in cross-domain comparative studies (see Chapter 4).

RCJ are the most used metacognitive judgement in the metaperception field and only a few studies in this domain have focused on prospective judgements. In the few works on prospective judgements in perception, a prospective judgement is not operationalised in the same way as for memory (where the study-test phases provide the possibility of predicting performance on an upcoming test or decision). In metaperception, the prospective judgement is not strictly a prediction before having encountered the trial, but a judgement made just before the motor response of the first-order decision, and with all the on-screen information necessary to perform the task (Siedlecka, Paulewicz, & Wierchoń, 2016; see Barthelme & Mamassian, 2010 for a similar example using other metacognitive decision). That is, participants first give their level of confidence, and then commit to a decision.

In metamemory, retrospective and prospective judgements have been shown to differ in terms of their accuracy: participants are less able to do not discriminate between

correct and incorrect responses in prospective judgements (i.e., metacognitive sensitivity, see next section for details) compared to retrospective judgements (e.g., [Kelemen, Frost, & Weaver, 2000](#)) as more information is available after task completion. Nonetheless, prospective judgements remain overall accurate. To the best of our knowledge, only [Fleming, Massoni, Gajdos, and Vergnaud \(2016\)](#) used prospective judgements in perception that are close to the metamemory definition. In their experiment, participants performed a typical dot-density discrimination task followed by RCJs for each trial. Every fifth trial, they had to give their confidence on their ability to discriminate correctly the next trial. The comparison between the two types of judgements showed that prospective judgements in visual perception decision making do not discriminate between correct and incorrect responses. This result points to a first difference between the metamemory field and the metaperception field.

All these judgements can be performed on different scales and it is also interesting to compare the scales that are used across fields. In the metamemory literature, a common approach is to ask people to judge on a percentage scale ranging from 0% to 100% with multiple steps (e.g., [Finn & Metcalfe, 2007](#)). Other works also use dichotomic judgements (e.g., Will you retrieve the correct answer? yes/no"; [Souchay & Isingrini, 2012](#)). In the metaperception field, RCJs are usually on a 4-point or 6-point scale and rarely use percentage (e.g., [Fleming et al., 2015](#); [Song et al., 2011](#), although see [Fleming et al., 2016](#)).

A few studies have considered whether these different types of scale influence the results. [Tunney and Shanks \(2003\)](#) compared two types of metacognitive scale in artificial grammar learning. After an encoding phase where participants had to learn strings that obey to one of two proposed rules, they had to classify the strings according to the rule they should obey. After each classification, participants had to judge their level of confidence using either a dichotomic scale (high or low confidence) or a continuous scale from 50% (referring to random responses for the first-order task as it was a 2-alternative force choice task; 2AFC) to 100% (the highest level of confidence). The authors found that the point scale was a better predictor of the task performance. [Dienes \(2007\)](#) reproduced this experiment and found the opposite result. However, task difficulty was not the same in the two experiments and it is possible that scale use was different according to this variable. He also proposed an experiment which compared six different scales and no difference between them was highlighted. Thus, although few studies focus on the scale question, it seems that differences are minor.

In the field of metaperception other direct tests of metacognition exist that can be defined as metacognitive decision. For instance, in the confidence forced-choice paradigm (Barthelme & Mamassian, 2010; de Gardelle & Mamassian, 2014; de Gardelle, Le Corre, & Mamassian, 2016), participants have to select the stimulus they feel is more likely to be correct for each trial. Metacognition refers here to the comparison of the first-order performance for stimuli that were chosen (i.e., associated with confidence) and those that were not chosen (i.e., associated with no confidence). These decisions are often used to assess uncertainty (that is seen as the inverse of confidence, Meyniel, Sigman, & Mainen, 2015) in perception. As suggested in Chapter 1, such evaluation of the reliability of the signal during the decision process can be used in order to control behaviours (e.g., modifying decision-making; Bach & Dolan, 2012). It is therefore seen as metacognitive component. Although such a method pertains to direct test of metacognition, we do not consider them as metacognitive judgements as these judgements imply a by-trial evaluation and not a cross-trial comparison.

Finally, a last class of direct test refers to error-detection paradigms. Also mainly used in visual perception tasks, participants are often aware of their own mistakes (e.g., Rabbitt, 1966). These protocols are very similar to the RCJs one as a first-order decision is performed followed by a second-order questions such as “did you make an error?”. On the contrary to RCJs paradigm where reaction time (RT) for the first-order decision is not necessarily controlled, error-detection protocols are most of the time pressure tasks. Although works dedicated to RCJs and error-detection have been developed separately, they now tend to converge and common frameworks have been proposed (e.g. Yeung & Summerfield, 2012).

As reported here, there is a wide range of direct tests of metacognition. The focus of this thesis is metacognitive judgements (both prospective and retrospective) as they have been the most studied in the metaperception and the metamemory field. We highlighted that prospective judgements have been mainly develop in context of memory tasks (especially episodic memory) whereas these judgements are less present in perception. Moreover, that retrospective judgements are always more accurate than prospective judgements regardless of the field, these latter judgements being completely inaccurate in visual perception tasks. One of our aims here will be to compare these two kinds of judgements in visual perception tasks and episodic memory tasks (see Chapter 5 and Chapter 6).



### 2.1.2 The quantification of metacognition

Several methods exist to quantify metacognition in relation with the experimental design used to measure it. However, here we exclusively focus on the methods that can be applied to metacognitive judgements. As they have an evaluative value, the core notion is the accuracy of these judgements that is if they are related to the task performance. As mentioned above, two aspects of metacognition can be separated: bias (or absolute accuracy) and sensitivity (or relative accuracy). When reviewing the metacognitive literature, we can notice that these two notions have slightly differences according to the field of metaperception or metamemory.

#### Metacognitive sensitivity

Regarding metacognitive sensitivity, definitions across the fields converge toward the idea that it refers to the ability to discriminate between correct and incorrect responses. The calculation of such a simple notion has been widely discussed and reviewed (Barrett, Dienes, & Seth, 2013; Benjamin & Diaz, 2008; Dunlosky et al., 2018; Fleming & Lau, 2014; Masson & Rotello, 2009; Sherman, Barrett, & Kanai, 2015). We here briefly review the most used methods according to the respective field of studying metacognition. The memory literature has mainly relied on Goodman-Kruskall gamma correlations (Kruskal & Goodman, 1954; Nelson, 1984), non-parametric correlations comparing the number of concordant judgements (a high confidence for a correct response or a low confidence for an incorrect response) and discordant judgements (a high confidence for an incorrect response or a low confidence for a correct response). There has been plenty of discussions of the suitability of this measure and even Nelson himself (Nelson, 1988) found split-half reliability of only  $\rho = -0.02$  for 110 general knowledge FOK items. Moreover, it has been shown that gamma correlations are sensitive to metacognitive bias leading to interpretation problems when comparing two groups with different biases (Masson & Rotello, 2009). To avoid such a confounding factor, other methods have turned to signal detection theory which is known to assess independently bias and sensitivity (Green, Swets, et al., 1966).

We can separate between the Type 1 SDT which refers to discrimination in the first-order task and the Type 2 SDT referring to the discrimination between correct and incorrect responses about the first-order performance. In Type 1 SDT,  $d'$  refers to the ability to discriminate between different states of the world (i.e., signal and noise). This parameter can be calculated as  $d' = z(\text{hits}) - z(\text{false alarms})$ , where  $z$  is the inverse of the

cumulative normal distribution function, hits are the proportion of ‘signal’ responses when signal is present, and false alarms are the proportion of ‘signal’ responses when noise is present (assuming Gaussian distribution for signal and noise with equal variance). The same logic can be applied for a second-order task. In Type 2 SDT, the sensitivity parameter can be calculated as  $Type\ 2\ d' = z(H2) - z(FA2)$ , where  $H2$  are the proportion of ‘high confidence’ responses when first-order decision is correct, and false alarms are the proportion of ‘high confidence’ responses when the first-order decision is incorrect (see Table 2.1). However, this method also assumes that distributions of correct and incorrect responses are Gaussian with equal variance, which is rarely the case (Galvin, Podd, Drga, & Whitmore, 2003). Therefore, a suitable solution is to use non-parametric methods for metacognitive sensitivity. With multipoint rating scales, it is possible to construct a Type 2 receiver operating characteristic (ROC) curve (Figure 2.3A) in which each point of the confidence scale allows a separation between ‘low confidence’ and ‘high confidence’ according to the correctness of the Type 1 response ( $H2$  rate and  $FA2$  rate). Then, it is possible to calculate the area under this Type 2 ROC curve ( $AUROC2$ ) as an index of metacognitive sensitivity.  $AUROC2$  can be interpreted akin to a proportion of discrimination between correct and incorrect responses regardless of metacognitive bias: an area equal to 0.5 suggests that second-order performance is at chance and 1 refers to perfect metacognitive sensitivity (Benjamin & Diaz, 2008; Galvin et al., 2003).

Table 2.1: Type 2 SDT and classification of responses.

Type I decision	High confidence	Low confidence
Correct	Type 2 hit ( $H2$ )	Type 2 miss ( $M2$ )
Incorrect	Type 2 false alarm ( $FA2$ )	Type 2 correct rejection ( $CR2$ )

Although this method has been largely used to quantify metacognitive sensitivity in metaperception in the visual modality (e.g., Ais, Zylberberg, Barttfeld, & Sigman, 2016; Fleming, Weil, Nagy, Dolan, & Rees, 2010; Song et al., 2011; Weil et al., 2013) or tactile modality (e.g., Whitmarsh, Oostenveld, Almeida, & Lundqvist, 2017) and also in metamemory mainly for RCJs (e.g., Higham, 2007; Higham, Perfect, & Bruno, 2009) but also JOLs (e.g., Benjamin & Diaz, 2008; Masson & Rotello, 2009), some methodological concerns arise. Galvin et al. (2003) showed that  $AUROC2$  are sensitive to both  $d'$  and

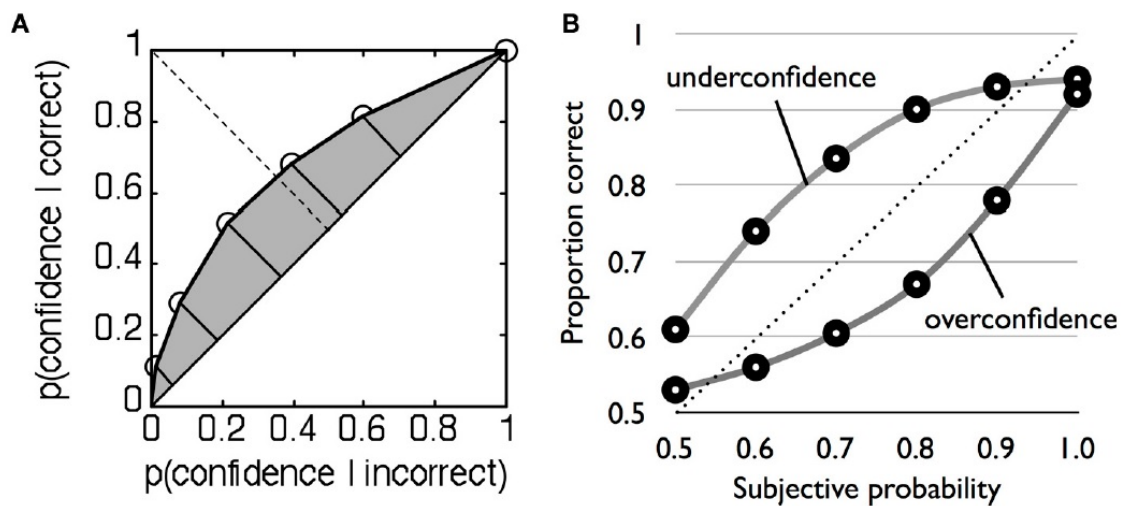


Figure 2.3: (A). Example of a AUROC2 where each point refers to the H2 rate (the probability to give high confidence for correct responses) and FA2 rate (the probability to give high confidence for incorrect Responses). Thus, the number of point is N-1 points of the metacognitive rating scale. (B) Calibration curves. Estimated probabilities are each judgement scale points. The upper curve refers to underestimation and the lower curve refers to overestimation of performance. Reprint from [Fleming and Lau \(2014\)](#).

Type 1 criterion. As such, metacognitive sensitivity proved to be greater for higher first-order performance. This can be intuitively appreciated in the extreme case of a participant performing a recognition task at chance. In this example, correct responses rely on random effects and the participant should have an inability to introspect on these correct responses (although see Blind insight effect using Type 2  $d'$ ; [Scott, Dienes, Barrett, Bor, & Seth, 2014](#)) and therefore have a low metacognitive sensitivity ([Fleming & Lau, 2014](#)). Figures 2.4B and 2.4C propose examples of how  $d'$  value can affect the proportion of H2 and FA2 (and therefore AUROC2). Let us suppose an unbiased observer performing a 2AFC task with confidence judgements on a 3-point scale. Her Type 1 criterion,  $c = 0$ , meaning that she has no tendency to give one response more than the other. For each response "S1" or "S2", there are two Type 2 criteria delimiting the level of confidence associated with the response (Figure 2.4A). For instance, if the internal response falls over the second  $c_2$ , "S2" the associated confidence rating is 3. As suggested by Figure 2.4B and 2.4C, the proportion of H2 and FA2 changes according to  $d'$  for the same values of  $c$  and  $c_2$ . When  $d'$  is large (Figure 2.4B) the proportion of FA2 is different from 0 only when confidence equals 1 (an "S2" response for a S1 presentation corresponding to the hatched area). However, when  $d'$  is lower (Figure 2.4C, the distance

between mean S2 distribution and  $c$  has been shortened), a greater proportion of FA2 appears for a confidence level of 2 (corresponding to the red hatched area). AUROC2 is therefore smaller in this condition. Thus, when task performance is likely to differ across participants (e.g., memory tasks which have high between-subject variability) or group comparisons (e.g., patients vs. controls) spurious differences in metacognitive sensitivity may emerge driven by variation in first-order task performance, rather than metacognitive capacity itself.

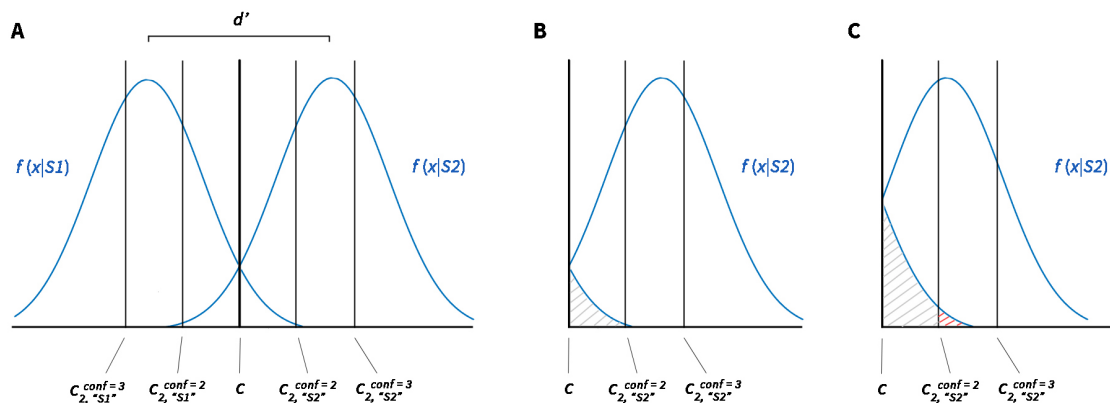


Figure 2.4: (A). SDT plot for a discrimination task between stimuli S1 and S2. The x axis is the internal response and participants answer “S1” when the internal response falls under the Type 1 criterion ( $c$ ) and “S2” when it falls over  $c$ . The  $d'$  value refers to the ability to discriminate between S1 and S2 stimuli (the distance between the two distribution) and  $c_2$  are the responses-specific Type 2 criteria (confidence ratings being a 3-point scale in this example). (B). SDT plot for “R2” responses for a large  $d'$ . The proportion of FA2 is different from 0 only when confidence = 1 (grey hatched area). (C). SDT plot for “R2” responses for a small  $d'$ . The proportion of FA2 is different from 0 when confidence = 1 (grey hatched area) and confidence = 2 ( $d'$ ) (red hatched area).

One recent measure that achieves this control is metacognitive efficiency,  $\text{meta-}d'/d'$ . The  $\text{meta-}d'$  framework models the relationship between performance and metacognition where  $\text{meta-}d'$  is defined as the Type 1  $d'$  that would lead to the observed Type 2 ROC curve in the absence of noise or imprecision in confidence estimates (Maniscalco & Lau, 2012). Because of the relationship between Type 1 and Type 2 performance, it is possible to determine the shape of the Type 2 ROC curve given a particular  $d'$  Type 1 criterion (Galvin et al., 2003). This AUROC2 refers to the ideal metacognitive sensitivity for given tasks information. The idea of  $\text{meta-}d'$  is to reverse this relationship by computing the observed Type 2 ROC curve for a participant (i.e., assuming that this ROC curve is his or her ideal curve) and expressed it in terms of Type 1 parameter ( $\text{meta-}d'$ ). Therefore,

meta- $d'$  quantifies the sensitivity of confidence ratings to performance in units of  $d'$ , which is the signal available for a subject to perform the type 2 task (Maniscalco & Lau, 2012). Because  $d'$  and meta- $d'$  are in the same units, they can be compared which allows derivation of a measure of metacognitive efficiency, controlling for task performance. If this measure ( $M_{ratio}$ ; meta- $d'/d'$ ) is close to 1, then metacognitive efficiency is optimal under the SDT. Because this index is a ratio, the value can be extremely high in the case of very low first-order performance. A solution to deal with this issue is to either use a logarithmic transformation of  $M_{ratio}$  or calculate the difference between meta- $d'$  and  $d'$  (Rounis, Maniscalco, Rothwell, Passingham, & Lau, 2010). Meta- $d'$  can be fitted with minimization of squared-sum-error but the most used fitting method is maximum likelihood estimation (Maniscalco & Lau, 2012). Bayesian estimations have also now been developed (Fleming, 2017; see Chapter 3 section 2.2 for more detail).

More specifically, the meta- $d'$  model consists of the maximization of the likelihood of the response-specific confidence data for the model parameters  $\theta$  (the Type 1 parameters that are generated by the model which are meta- $d'$ , meta- $c$ , and response-specific meta- $c_2$ ). When  $\theta$  values are specified, we can derive the probability with which the model generates confidence ratings for each stimuli and responses (Type 2 model). Model assumptions are that meta- $c = c'$  (where  $c'$  is the Type 1 relative criterion calculated as  $c' = c / d'$ ; Macmillan & Creelman, 2004) and that Type 1 and Type 2 criteria should have an ordinal relationship. Hence, the model looks for the best concordance between the Type 2 ROC curve generated by the model (the expected Type 2 ROC curve) and the actual type 2 data (the observed Type 2 ROC curve).

The Type 2 SDT model allows the estimation of ideal Type 2 ROC curves given Type 1 parameters ( $d'$ ,  $c$ , and  $c_2$ ). It is possible to construct either directly the overall Type 2 ROC curve (Galvin et al., 2003; assumption of a Type 2 decision axis so symmetry in the placement of Type 2 criteria for S1 and S2) or the response-specific Type 2 ROC curve and average them to have the overall Type 2 ROC curve (Maniscalco & Lau, 2014). However the meta- $d'$  framework uses the second type of fitting these curves as an assumption regarding the type 2 decision variable is not required (the overall curve implies a symmetry in the placement of Type 2 criteria for S1 and S2 responses; Maniscalco & Lau, 2014). The construction of response-specific Type 2 ROC curve includes estimations of the probabilities of each confidence rating given stimuli (S1 or S2) and responses (S1 or S2). This leads to 4 probabilities per confidence rating. In Figure 2.4C, the different regions delimited by the multiple  $c_2$ , " $s_2$ " represents these probabilities for the S2 responses. We can then obtain these probabilities for the times where S2 stimuli

were indeed presented (i.e., correct responses) and for the times where S1 stimuli were presented (i.e., incorrect responses). As an example, the probability for confidence = 2, stimuli = S1, and response = "S2" is given by the area under the S2 curve between the confidence boundaries  $c_{2, "S2"} = 2$  and  $c_{2, "S2"} = 3$  (the red area on the Figure) divided by the area under the S2 curve that is above the Type 1 criterion (all the "S2" responses). This measure has been extensively used in the metaperception field (e.g., Fleming et al., 2015; Maniscalco & Lau, 2015; Rouault, Seow, et al., 2018) but rarely in the metamemory literature. Nonetheless, recent studies comparing metacognition for visual perception and for memory (e.g., Baird, Smallwood, Gorgolewski, & Margulies, 2013; Fitzgerald, Arvaneh, & Dockree, 2017; McCurdy et al., 2013; Morales, Lau, & Fleming, 2018; Palmer, David, & Fleming, 2014; Sadeghi, Ekhtiari, Bahrami, & Ahmadabadi, 2017; Samaha & Postle, 2017) have now used meta- $d'$  in the context of metamemory for RCJs.

Overall, the method to quantify metacognitive sensitivity may also change according to the experimental procedure. When task performance is controlled, there is no need to compute metacognitive efficiency and AUROC2 or meta- $d'$  are sufficient. However, when task performance is likely to differ across participants (e.g., memory tasks) or group comparisons (e.g., patients vs. controls), meta- $d'/d'$  allows control of such variability. Although the field of metaperception focuses on SDT-based quantification of metacognitive sensitivity, the memory field mainly relies on non-parametric correlations between judgements and first-order performance such as gamma correlations. The difference is therefore a limit in the comparison of studies across fields especially since gamma correlations have been shown to be biased by both task performance and metacognitive bias (Masson & Rotello, 2009).

Another way of computing metacognitive sensitivity is the use of logistic regression (Sandberg, Timmermans, Overgaard, & Cleeremans, 2010). Logistic regressions quantify the relationship between a binary response and a categorical or continuous predictor. These logistic regressions are particularly useful when they are used with random effects that take into account the hierarchical nature of metacognitive judgements (multiple trials per participants). However, it has been shown that this quantification of metacognition is not independent from both Type 1 and Type 2 criteria (Rausch & Zehetleitner, 2017).

Finally, psychometric curves are also used for specific paradigms that involve systematic stimuli variations according to stimuli characteristics (e.g., contrast or luminance). One curve referring to the association between varying stimulus characteristics

and task performance can be plotted per metacognitive judgement. For instance, in the confidence forced-choice paradigm (Barthelme & Mamassian, 2010; de Gardelle et al., 2016), one psychometric curve refers to stimuli that have been chosen (i.e., stimuli with high confidence) and a second one refers to stimuli that have not been chosen (i.e., stimuli with low confidence). When the stimuli characteristic that varies is diagnostic of task performance (e.g., stimuli motion direction and subject categorisation), the steeper the curve, the higher the metacognitive sensitivity.

### **Metacognitive bias**

While metacognitive sensitivity is the ability to discriminate between correct and incorrect responses, metacognitive bias is defined in the metaperception field as “the tendency to give high confidence ratings, all else being equal” (p.5, Fleming & Lau, 2014). It is mainly quantified by the average confidence ratings over trials and therefore refers more to the magnitude of judgements. This definition slightly differs from the metamemory one where “absolute accuracy pertains to the degree to which the magnitude of a person’s judgement matches the magnitude of performance” (p. 30, Dunlosky et al., 2018). From this perspective, there is a need to compare with the actual performance in order to detect overestimation (judgements being higher than the mean performance) or underestimation (judgements being lower than the mean performance). The logic is the following: if a participant gives higher confidence in Condition A compared to Condition B, this difference should not be considered as a bias if the actual performance is higher in Condition A compared to Condition B. We suggest that this difference in terms of quantification of bias occurs because task performance varies in memory (also voluntarily due to experimental manipulation) which is less the case in perception, because task performance is often controlled (for example, using a staircase procedure). This allows a very low between subject and between experimental condition variability. Therefore there is no need to compare the magnitude of judgements to the actual task performance. However, this is still routinely labeled as metacognitive bias, to distinguish it from sensitivity (as laid out in the taxonomy of Fleming & Lau, 2014). Nonetheless, there is a consensus to define bias as the tendency to give either high confidence or low confidence irrespective of task performance that is controlling for performance. Control of performance is either directly included in the experimental procedure or in the quantification of metacognition. As for metacognitive sensitivity, the method to quantify metacognitive bias may change according to the experimental procedure.

Another means of quantifying metacognitive bias that is common across fields refers to calibration curves (Harvey, 1997; Keren, 1991). Calibration refers to average performance for each point of the judgement scale. As this suggests a comparison between percentages of performance and prediction, it needs a scale which represents the percentages of both performance and prediction. If the curve is under the perfect calibration curve (Figure 2.3B), there is an underestimation of performance. The reverse refers to an overestimation.

As is the case for metacognitive sensitivity, another measure of metacognitive bias is based on SDT. Sherman, Seth, and Barrett (2018) developed the m-distance measure in order to quantify what they called a metacognitive threshold. Because using directly the  $c_2$  from the Type 1 SDT as a measure of metacognitive bias would make the strong assumption that participants use the same decision axis to perform the first-order and the second-order task (see section 2.2. of this Chapter), m-distance is based on the meta-d' model. Thus, this measure uses meta- $c_2$  that are  $c_2$  estimated by the meta-d' model (criteria that an ideal-observer would use given the actual data of the participant). Furthermore, it is a distance in a sense that it allows the estimation of the additional bias from the meta-c (actually the same as Type 1 c in the meta-d' model) which results in a metacognitive bias that is not biased by the placement of the first-order criterion (see Figure 2.5). Finally, as it is the case for the relative criterion  $c'$  in Type 1 SDT, m-distance is divided by meta-d' in order to control for metacognitive sensitivity and to compare this threshold across subjects. This distance between the Type 1 criterion and the Type 2 criterion can be calculated for each response as follows (in the case of a 2-point confidence scale):

$$mdist_{R1} = \frac{metac - metac_{2,R1}}{metad} \quad mdist_{R2} = \frac{metac_{2,R2} - metac}{metad} \quad (2.1)$$

When thresholds for confidence are supposed to be placed symmetrically around the Type 1 criterion (as it is the case in Figure 2.5) then  $mdist_{R1}$  equals the opposite of  $mdist_{R2}$ . Therefore, the lower the distance, the bigger the tendency to report high confidence, as the criterion can be reached with less internal evidence (Figure 2.5). This measure overall allows a good estimate of confidence criterion within the SDT framework since it allows a control for any shift in Type 1 criterion and difference metacognitive sensitivity measured by meta-d'.

To conclude this section, the quantification of metacognitive judgements is a primary focus in the field. It has been an important distinct topic of metacognitive research since



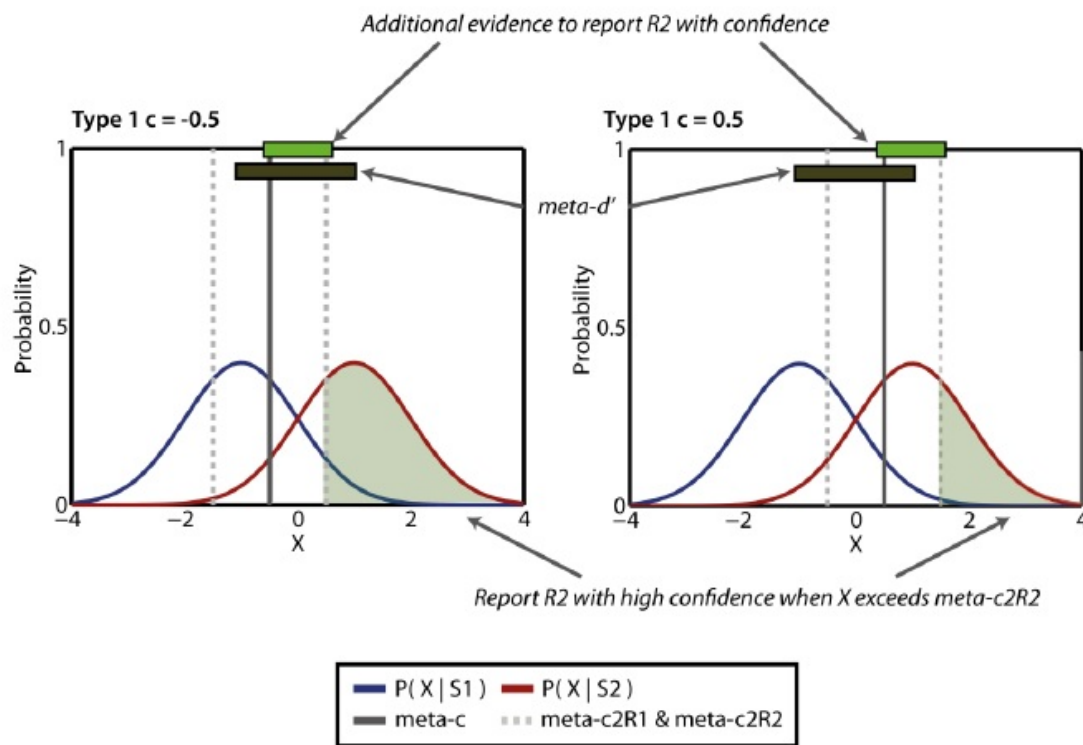


Figure 2.5: Representation of the meta- $d'$  model for a metacognitively ideal observer. The two dashed lines are meta- $c_2$ , “R1” and meta- $c_2$ , “R2” and are symmetrically disposed around meta- $c$ . The additional evidence to report “R2” with high confidence ( $m_{\text{dist}}^{\text{R2}}$ ) is the same regardless of Type 1 criterion. Reprint from [Sherman et al. \(2018\)](#)

early work (e.g., [Nelson, 1984](#)), however indexes are thus far, very field-dependant. The field of decision making especially in visual perception has a strong tendency to use the meta- $d'$  framework or other SDT measures although some works also extensively use psychometric curves or logistic regressions for measuring metacognitive sensitivity. On the other hand, the field of metamemory has mainly been focused on gamma for metacognitive sensitivity and either calibration curves or discrepancy score for bias.

In this thesis, we will focus on the meta- $d'$  framework (see Chapter 3 for more details on methodological choices for both metacognitive sensitivity and metacognitive bias) as this measure is the most independent from both first-order performance and metacognitive bias as it is the case in recent literature (e.g., [Baird et al., 2013](#); [Fitzgerald et al., 2017](#); [McCurdy et al., 2013](#); [Morales et al., 2018](#); [Palmer et al., 2014](#); [Sadeghi et al., 2017](#); [Samaha & Postle, 2017](#)). The core idea is to propose to use the most powerful approach that has been developed for metaperception to examine metamemory processes. Here we also propose to apply for the first time such quantification of metacognition for

prospective metamemory judgements (Chapter 5 and Chapter 6). Although SDT-based approaches have provided the most reliable estimations, it is nonetheless important to note that they imply the use of a constrained protocol (e.g., 2AFC or yes/no designs), which is less applicable to applied issues in eyewitness testimony or memory impairment.

## 2.2 Theoretical comparison

This section proposes a comparison of different theoretical frameworks and models that have been developed in the field of metaperception and metamemory. Although it is possible to notice common points in these models, it is important to highlight that in some cases they have been proposed to explain different aspects of metacognitive judgements. The field of metaperception exclusively focuses on retrospective confidence whereas metamemory has proposed larger theories of metacognitive judgements including both prospective and retrospective judgements. Moreover, models of confidence formation have mainly focused on confidence in decision making when an observer has to decide between two answers. On the contrary, the metamemory field has proposed models of metacognitive judgement in the context of both recall and recognition (as a potential memory decisions). This difference in terms of first order task is important to take into account when comparing these models. The main aim of this section is to give a brief overview of theoretical models in both fields. Interestingly, although metaperception and metamemory have been developed separately, similarities across model emerge. We organise this section according to models that have been proposed to account for (1) the observed relationship between metacognitive judgements and first-order task performance, and (2) dissociations between metacognition and task performance. Finally, we focus on a particular framework (stemming from the metamemory literature) that can be used in metacognition across a wide range of domains (i.e., first-order tasks; see Chapter 3 for more insights on the notion of “domain”) and judgements.

### 2.2.1 Relationship between metacognition and task performance

In both metaperception and metamemory, the first models of metacognitive judgements claimed that there is a strong relationship between first-order and second-order judgements. The main claim of these models is that there is a dependency between the quality of a cognitive function (e.g., visual discrimination, visual detection, strength of the

memory trace, the amount of general knowledge) and metacognitive accuracy. Models of metaperception and metamemory merge when they both refer to confidence formation decision making in the case of visual perception decision making and recognition memory. These models suppose that confidence is driven by a subjective internal decision variable which depends on the sensory evidence. Two main frameworks have been developed: the SDT framework and the evidence accumulation framework (see [Mamassian, 2016](#) for a review). The main difference between these two frameworks relates to the information that is modelled. Whilst SDT exclusively focuses on response proportions, evidence accumulation models take into account reaction times. On the other hand, theoretical models developed in the metamemory field do not exclusively focus on confidence. Here we sketch out a few of the critical models of confidence formation in either or both domains.

### **Models of confidence formation**

The SDT framework has been used as a model for predicting behaviour but also as a measurement tool to quantify metacognition. As mentioned in the methodological section, this framework supposes that first-order decision and confidence are computed from the same evidence axis. However, SDT first-order models propose that the decision variable support first-order decision and confidence and that both are computed at the same time.

On the contrary, the accumulation of evidence framework proposes that evidence is accumulated until a decision is made either when evidence has reached a predefined boundary or when a certain amount of time is exceeded. Many accumulation of evidence models have been developed for two-choice situations with different stopping rules and different types of accumulation ([Ratcliff & McKoon, 2008](#)). These models have also been extended to explain confidence formation in decision making and are named “decision locus” models ([Yeung & Summerfield, 2012](#)) as confidence is computed while the decision is being made (akin to first-order SDT models). For instance, confidence can be computed from a comparison between the evidence for the decision that has been made and the evidence for the other alternative (i.e., balance of evidence; [Vickers, 1979](#)). Thus, the bigger the difference, the higher the level of confidence. In drift diffusion models, there is only one variable which accumulates evidence (see Figure 2.6A) and one suggested possibility for computing confidence is to consider the distance between one of the bounds and the accumulated evidence when the decision is made ([Kiani & Shadlen, 2009](#)). Another model proposed is that one accumulator exists per confidence

rating (Ratcliff & Starns, 2013). The variable accumulating evidence that first reached the decision boundary for a given confidence determines this level of confidence as the winner. The accumulation of evidence framework allows a good explanation of the relationship between decision time and confidence – higher confidence are observed for decision with a low RT – that have been found across lots of cognitive domains (Baranski & Petrusic, 1994; Vickers & Packer, 1982). These models assume that the quality and the quantity of evidence are used in confidence formation and therefore first-order and second-order decisions are computed from the same internal state. Note that the majority of these models have been developed in the context of tasks as described in the first section; that is where participants have to perform two tasks sequentially: first a two-choice decision and second a confidence judgement. Therefore, it is possible that both decisions are computed separately (or somewhat separately).

From a neuronal point of view, studies have also highlighted that the neural signature of confidence emerges from the decision process in humans (Gherman & Philiastrides, 2015) and that same neurons can represent both confidence and decision in non-human primate (Kiani & Shadlen, 2009). Similarly, fMRI has revealed that activity of the ventromedial PFC reflects both value-based choice and confidence formation in this choice (De Martino, Fleming, Garrett, & Dolan, 2013). Thus, it suggests that brain regions that account for first-order decisions are also involved in the computation of confidence.

Overall, both SDT models and accumulation of evidence models propose that confidence reflect the quantity or quality of evidence which suggest that “observers have a direct access on this [information]” (p.1312, Yeung & Summerfield, 2012). However, alternative models based on these frameworks also account for a dissociation between first-order performance and metacognition (see next section).

Finally, it is important to note that a large literature on recognition memory proposes to use RCJs in the context of the recognition memory decision making SDT framework. Although these models are not model of confidence per se, we would like to briefly mention them here as they integrate confidence in an opposite way as the field of metaperception. Whilst this latter field (metaperception) suggests that confidence stems from evidence in the first-order task, models of recognition memory propose the use of confidence to infer memory processes (akin to a measure of trace strength).

In recognition memory, two main SDT models have been proposed: the dual-process signal-detection model (DPSD; Yonelinas, 1994; Yonelinas, 2002) and the unequal-variance

signal-detection model (UVSD). The DPSD supposes that recollection (a process associated with the retrieval of contextual details in the encoding situation) is a threshold process working as all-or-none whereas familiarity (a process allowing the judgement of prior exposure to an item without recalling contextual details) is a SDT continuous process. Model assumptions are that when making a memory decision, participants tend to recollect the item and if recollection fails, they use the familiarity process. An alternative model, the UVSD proposes that memory is based on a unidimensional trace. However and unlike classical SDT, the variance for targets exceeds the variance for lures which allows explanations of a range of empirical findings not accounted by the equal-variance model (for a review see [Wixted, 2007](#)). Research to date has extensively focused confidence judgements and ROC curves in order to discriminate between models. Confidence is used to infer trace strength or the involvement of different memory processes and thus it is seen as a first-order decisional process instead of a second order process. If models of confidence in perception and models of memory that use confidence both suggests that there is an intrinsic relationship between confidence and the first-order processes (memory, visual perception), inferring memory processes using confidence suggests an important caveat in view of findings that dissociate metacognition and task performance (see Section 2.2). Note that other models that have been mainly developed in the context of memory and that do not rely on either SDT or evidence accumulation have also integrated confidence. As such, models of recognition memory have suggested that confidence is directly computed from the strength of the perceived familiarity between a probe and a memory trace ([Hintzman, 1988](#); [Murdock, 1983](#)).

### **Theoretical accounts of metamemory**

This idea of “direct-access” is also present in the field of metamemory. In this literature, models similar to first-order models have also been developed to explain both prospective and retrospective metacognitive judgements. According to the direct-access hypothesis ([King, Zechmeister, & Shaughnessy, 1980](#)), people estimate their memory performance by having a direct access to the strength of their memory trace. As the objective performance to memory tests and the subjective evaluation of performance are based on the same memory trace, it explains why people are overall accurate in their metacognitive judgements. This model has been especially used to account for the accuracy of JOLs ([Dunlosky & Nelson, 1994](#); [Mazzoni & Nelson, 1995](#)) and FOKs ([Koriat, 1993](#); [Schwartz, 1994](#)). The main idea is that people can monitor online their memory trace and that objective performance and subjective evaluation are based on

a direct access to the strength of evidence (see [Koriat, Lichtenstein, & Fischhoff, 1980](#) for the same idea in RCJs). Even if gamma correlations between predicted and actual performance are often not very strong, this does not contradict the direct-access view. Because JOLs and FOKs<sup>2</sup> are made item-by-item, but in block that is not at the same time as the first-order task (i.e., a later recall or recognition phase) it is possible that task performance depends on other factors than the strength of the memory trace that existed when the metamemory judgements are made (which is not the case for RCJs that are made at the same time as task performance). Thus, it has been shown that delayed-JOLs are more accurate than immediate-JOLs ([Nelson & Dunlosky, 1991](#)). In this paradigm, participants had to perform JOLs either directly after having encoded the word-pair or after a short delay. Whilst these two conditions yield to the same recall rate, metacognitive sensitivity was higher for delayed-JOLs. It is proposed that delayed-JOLs have a more similar context to recall than immediate-JOLs as the latter would be more biased by short-term memory that is not used in episodic recall.

More recently, the non-criterial recollection hypothesis suggests that different metamemory judgments are enhanced by information such as contextual details from the encoding phase of an item (see [Isingrini et al., 2016](#) for FOKs, and [McCabe & Soderstrom, 2011](#) for JOLs). Therefore, the better the ability to retrieve (“recollect”) the encoding context, the better the metacognitive sensitivity. Akin to AUROC2 being sensitivity to Type 1 d’, it suggests that good memory performance relates to good metacognition. The idea that an effective memory is needed to be metacognitively accurate has also been highlighted by studies in neuropsychological populations (see [Ernst, Moulin, Souchay, Mograbi, & Morris, 2016](#); [Souchay, 2007](#) in Alzheimer’s disease). For instance, differences in FOKs in different patients and older adults are concomitant with a deficit in either episodic recall or recognition (except in autism, see [Wojcik, Moulin, & Souchay, 2013](#); see Chapter 5 for more details). However, this literature largely relies on gamma correlations to quantify metacognitive sensitivity which is known to be influenced by task performance (see previous section). Therefore, the observed impaired metacognition in these populations could be a spurious deficit due to the chosen index.

Overall, the relationship between metacognition and first-order performance has been explained by both models of confidence and models of metamemory. Models of confidence have been mainly developed in the context of visual perception despite the fact that confidence is also used in recognition memory but rather to infer first-order

---

<sup>2</sup>With this respect, FOKs exhibit a special status as they occur after a recall failure (so at the same time as the first-order task) but the question relates to recognition that is made in another future block.

processes. As such, confidence is not seen as a metacognitive process per se which is questionable especially regarding empirical dissociation between metacognition and task performance.

## 2.2.2 Dissociation between metacognition and task performance

As mentioned in the previous section, it is possible that both decisions are computed separately or partially separately because during the task RCJs are performed as a second decision. This idea is particularly enhanced by evidence suggesting that dissociation between Type 1 performance and metacognition can occur. Moreover, this is enhanced by the ability to detect errors in the Type 1 decision. This section will firstly review behavioural and neuroimaging evidence in favour of separable processes between first-order and second-order decisions. Then, we will turn to theoretical models that support the idea of the involvement of additional factors in the formation of metacognitive judgements.

### Behavioral and neuronal evidence

In metacognition for both memory and perception, a certain amount of work has shown that dissociations can occur between first-order performance and metacognition. These dissociations have been identified at different levels: in a young adult population, across the lifespan, and in the context of pathologies, at both the behavioural and the neurological level.

In visual perception decision-making, opposite patterns between type 1 performance and the magnitude of confidence have been observed. For instance, Transcranial Magnetic Stimulation (TMS) applied on bilateral dorsolateral PFC created a decrease in metacognitive sensitivity without affecting perceptual discrimination (Rounis et al., 2010). Conversely, single-pulse TMS on the occipital lobe allows a decrease in a tilt orientation discrimination task performance whilst increasing the magnitude of confidence judgements (Rahnev, Maniscalco, Luber, Lau, & Lisanby, 2011). Similarly, without directly using confidence judgements but rather subjective awareness decisions (dichotomic decision between “seen” and “guess”), Lau and Passingham (2006) have the amount of time between the target and a particular mask have no effect on task performance whereas increasing the proportion of “seen” responses. This magnitude of response was also specifically correlated with the activity of Broadman’s area 46. Controlling

first-order performance using a staircase procedure has also revealed an extreme variability in metacognitive sensitivity across people, and this variability has been linked to the grey matter volume in the anterior PFC (Fleming et al., 2010). Neurological damage and experimental situations suggest double dissociations between first-order performance and metacognitive sensitivity. Patients with lesions to the anterior PFC have shown a specific metacognitive sensitivity deficit in visual perception decision making (Fleming, Ryu, Golfinos, & Blackmon, 2014). Moreover, blindsight patients exhibit above chance task performance whilst being unaware of such outcome (e.g., Poppel et al., 1973; Weiskrantz, Warrington, Sanders, & Marshall, 1974).

Implicit learning, such as artificial grammar learning offers a similar situation where participants can perform above chance in classification tasks in the absence of confidence (e.g., Dienes, Altmann, Kwan, & Goode, 1995). Using the same paradigm, the opposite pattern has also been shown: participants that were not able to learn in the implicit task (first-order performance was at chance) were nonetheless able to discriminate between correct and incorrect responses (i.e., “blind insight” phenomenon; Scott et al., 2014). This range of findings implies that in some situations first-order evidence is non-available for the second-order task and that in other situations second-order evidence cannot be accessed during the first-order task.

Interestingly, such double dissociations have also been shown in the metamemory literature. Amnesic patients have been shown good metacognitive predictions measured by eFOK despite their severe memory deficit (except for Korsakoff patients; Shimamura & Squire, 1986). On the contrary, patients with autism exhibit no episodic memory deficit despite having a lower eFOK sensitivity compared to neurotypical participants (Wojcik et al., 2013). Moreover, participants can be unaware of some factors that can contribute to memory performance. For instance, whereas spacing at encoding increases learning comparing to massed information, the magnitude of JOL for these two types of encoding is the same (Logan, Castel, Haber, & Viehman, 2012; Zechmeister & Shaughnessy, 1980). Conversely, participants can identify variables as likely to influence their memory and therefore can increase the magnitude of their judgements whilst first order performance remains the same. These effects are known as metacognitive illusions (see the cue-utilization approach in the next session for more details).

Finally, empirical works have repetitively shown “hyper-metacognitive sensitivity” (see Fleming & Daw, 2017) suggested by a meta- $d' > d'$ . However, according to the underlying model, such situations are not supposed to occur as Type 1 parameters define the maximal AUROC2 that it is theoretically possible to obtain (i.e., given the



available signal; Galvin et al., 2003) which suggests that the first-order model is not completely appropriate to model metacognitive behaviours.

### Post-decisional and second-order models of confidence

More recent models of confidence formation support the idea of the involvement of both a representation of sensory evidence during the decision task and additional factors. Contrary to first-order models which suppose that both first-order decision and confidence are computed in a single stage, post-decisional models require that information for both decisions are collected in dual stages. Such models allow an explanation for the dissociations described above as they suggest that independent processes (or noise) influence first-order or second-order decisions independently. Thus, parameters that interfere with the accumulation of evidence in this second stage would affect only the confidence task (Navajas, Bahrami, & Latham, 2016).

In post-decisional accumulation models, there is new information integration because evidence continues to be accumulated for a short while after the first-order decision has been made. In visual perception decision making, the two-stage dynamic signal detection theory is a good example of such models (Pleskac & Busemeyer, 2010). As an extension of the drift diffusion model, it assumes that accumulation of evidence does not stop until one of the two boundaries is achieved and that accumulation of evidence keeps going in order to perform the confidence task after the decision has been made (Figure 2.6B). Post-decisional models are particularly interesting as they can account for the error-detection phenomenon (e.g., Rabbitt & Vyas, 1981). Models of confidence and error-detection tend now to be modelled in a common framework (Fleming & Daw, 2017; Yeung & Summerfield, 2012).

Within the SDT framework, Type 2 criteria suggest that additional information and additional sampling is needed for confidence formation. As in post-decisional accumulation models, this additional information also depends on the first-order state as confidence is supposed to be computed from the same evidence axis and multiple Type 2 criteria are placed along the axis according to the Type 1 criterion (see Figure 2.4A). Thus, metacognitive sensitivity is better when first-order discrimination is high (AUROC2 depends on the Type 1 parameters  $d'$  and  $c$ ). However noise can influence confidence formation (resulting in a hypo-metacognitive sensitivity;  $M_{ratio} < 1$ ), or people can use other relevant information such as prior beliefs (resulting in a hyper-metacognitive sensitivity;  $M_{ratio} > 1$ ).

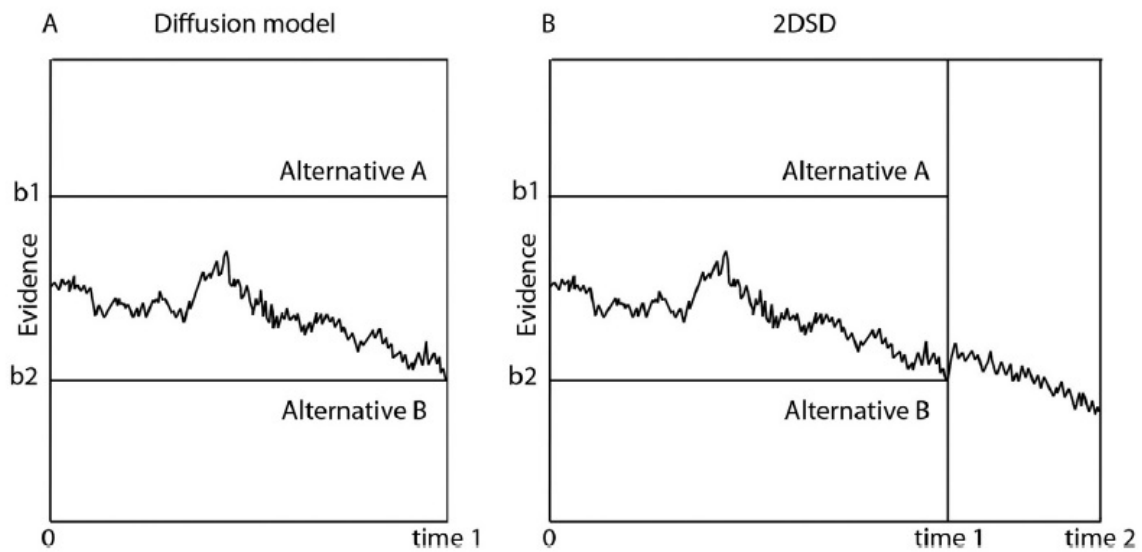


Figure 2.6: (A). Diffusion model. Evidence is accumulated until one of the bounds is reached. (B). The two-stage dynamic signal detection theory (2DSD). After one of the two bounds has been reached, evidence continues to be accumulated for a while which allows the formation of confidence.

Reprint from [Grimaldi et al. \(2015\)](#)

Recently, [Fleming and Daw \(2017\)](#) proposed another type of model that they called the “second-order” model. This model proposes that confidence is computed by second order inference that is independent from the first-order decision evidence. Crucially, they can be correlated, meaning that when perfectly correlated ( $\rho = 1$ ), the model refers to the first-order model. The authors compared a first-order model, a post-decisional model, and the second-order model. Although all of them were able to account the relationship between first-order performance and confidence, only the post-decisional model and the second-order model were able to predict error-detection. Moreover, only the second-order model supports the effect of motor action of the first-order task in the formation of confidence. The emerging literature on this topic suggests that the motor response involved in the first-order decision influences the level of confidence in that decision. This idea mainly stems from the comparison between prospective and retrospective confidence judgements: metacognitive sensitivity is higher when confidence is given after the first-order decision compared to before this decision without impacting first-order performance ([Siedlecka et al., 2016](#); [Siedlecka, Skóra, et al., 2019](#)). This is because, the second-order model supposes, “additional diagnostic information [is] provided by the action” (p. 98; [Fleming & Daw, 2017](#)). This is enhanced by recent

work showing that participants have a higher metacognitive sensitivity when they actively perform the first-order task compared to situations when they are observers and therefore have no action of perform (Pereira et al., 2018). Similarly, artificial delays in first-order response resulting in a sensory-motor conflict may reduce metacognitive sensitivity as well (Faivre et al., 2020).

Finally, the main difference between post-decisional accounts and the second-order model is that the former supposes that first-order decision and confidence is computed sequentially and from the same evidence whereas the latter supposes separated and parallel computations while allowing correlations between the two. Again, these models of confidence have been mainly developed in the context of decision making especially in visual perception. However, models of metamemory have also highlighted the idea that additional factors of task evidence (memory trace strength) are also involved in metacognitive judgements.

### **Cue-utilization in metamemory judgements**

Akin to the idea of second-order inference in confidence formation, current models of metamemory judgments support the idea that metacognitive judgements are inferential and involve cues. On the contrary to the direct-access view, the cue-utilization approach suggests that people have no access to the strength of their memory trace but can infer it using different cues. Therefore, Koriat (1997) explains metacognitive judgements as follows: “[...] their accuracy is not guaranteed but depends on the empirical correlation between the cues used and the criterion memory test” (p.350). Such an idea suggests that in some contexts these inferences might be misleading, therefore explaining that metacognitive judgements are not always accurate (or are illusory). When the cues used to perform judgements are also pertinent to the memory retrieval itself, they are diagnostic of memory performance and the judgements are accurate. However, in some situations the cues used might have no influence of the memory test per se. Thus, these cues are non-diagnostic. This directly leads to the notion of metacognitive illusion: participants might have an erroneous perception of competence by a misuse of these cues (see Chapter 6 for a more detailed description).

Koriat (1997) differentiates three types of cues: intrinsic cues, extrinsic cues, and mnemonic cues. Intrinsic cues are directly related to the characteristics of the stimuli during the task like perceptual characteristics (e.g., size, brightness, clarity, etc.), the emotional aspect of the stimulus, the relationship between targets and cues during pair-word learning, and so forth. On the contrary, extrinsic cues refer to the external factors

that not the related to the items but more to the task and processes involved in the task in general. In the context of memory this depends on the learning or retrieval conditions like the time to encode stimuli, the strategy the participant uses at encoding (e.g., visual imagery) or even knowledge of the type of retrieval test (e.g., recall or recognition). The final class of cues are mnemonic and include the experience of internal indicators or signals that can be used by participants to evaluate their level of memory performance. The metamemory literature has identified several mnemonic cues influencing a variety of metacognitive judgements such as the familiarity of the cue (e.g., [Metcalf, Schwartz, & Joaquim, 1993](#)) or processing fluency (e.g., [Alter & Oppenheimer, 2009](#)).

According to this approach, these three types of cues are used to perform metacognitive judgements by means of two processes: information-based or theory-based metacognition and experience-based metacognition ([Koriat, 1997](#); [Koriat & Bjork, 2006](#); [Koriat, Bjork, Sheffer, & Bar, n.d.](#); [Koriat & Levy-Sadot, 1999](#)). Information-based metacognition (high level metacognition as discussed in Chapter 1) involves inferential processes from explicit theories or beliefs. It refers to a direct influence of intrinsic and extrinsic cues on metacognitive judgements (see Figure 2.7 for an example on JOLs). For instance, participants can use the extrinsic diagnostic cue “type of test” and apply the beliefs that recognition is easier than recall, therefore evaluating their performance according to the task that is proposed to them. As such, information-based metacognition is inferential by nature and results in the use of analytic and controlled processes. On the contrary, experience-based metacognition (low level metacognition see Chapter 1) refers to an indirect influence of influence of intrinsic and extrinsic cues on metacognitive (see also Figure 2.7). The influence of these cues is named indirect as it is mediated by mnemonic cues. For instance, an intrinsic cue such as the size of the stimuli can influence ease of processing and therefore affects JOLs (see the font-size effect detailed in Chapter 6). Experience-based metacognition is also inferential, however it involves the application of heuristics that are automatic and non-analytic processes. As with the cues, the main identified heuristics are the accessibility heuristic, the cue-familiarity heuristic, and the fluency heuristic. The use of these heuristics gives rise to subjective feelings that are then used as a basis of metacognitive judgements. Although several heuristics have been identified, we suggest here that all actually pertain to the fluency heuristic. As such, the cue-familiarity can be seen as conceptual fluency effect as feeling of familiarity has been shown to be influenced by fluency through an attribution process (e.g., [Kelley & Rhodes, 2002](#); [Mandler, 1980](#); for more insights on fluency see Chapter 6). Similarly, the accessibility heuristics suppose that the accessibility of partial information in used

in metacognitive judgements (especially in the case of FOKs, [Koriat, 1995](#)) which can also refer to a conceptual fluency heuristic.

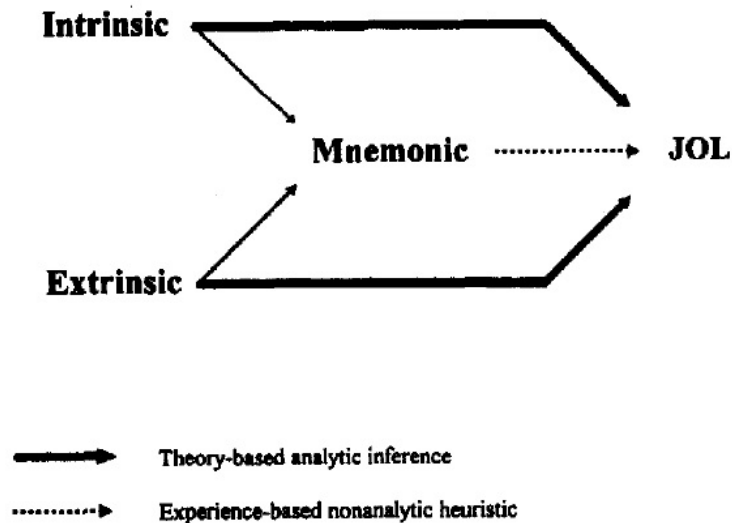


Figure 2.7: Schematic representation of the cue-utilization approach on JOLs. Intrinsic and extrinsic cues have direct and indirect effects on JOLs. The indirect effect is mediated to mnemonic cues. Reprint from [Koriat \(1997\)](#)

Overall, and in comparison with field of confidence formation more generally and in metaperception, there is a lack of formal models in the traditional field of metamemory. A few exceptions nonetheless exist. [Sikström and Jönsson \(2005\)](#) proposed a model that accounts for the difference between immediate-JOLs and delayed-JOLs. By modelling large and weak drifts of memory strength it also explained differences in JOL predictability of recall performance. A second more formal model that has been used in the context of JOL is the stochastic detection and retrieval model (SDRM; [Jang, Wallsten, & Huber, 2012](#)). Whereas classical SDT models assume the same decision axis for both first-order and second-order decisions, the SDRM supposes two decision axes possibly related (X-Y bivariate memory-strength distribution). From this perspective, SDRM is akin to the second-order model of confidence ([Fleming & Daw, 2017](#)). As [Sikström and Jönsson \(2005\)](#)'s proposal, this model allows explanations of the delayed-JOL effect but also of the testing-JOL effect which suggest that JOL sensitivity increases with practicing the same memory material. It suggests that people partly base their JOLs on the outcome of the previous recall (an effect also know as “memory for past test”, [Finn & Metcalfe, 2007](#)) which is explained by a greater correlation between memory sampling during JOL and during recall for the second practice time. However, the authors suggest

that it is “not a model for predicting behaviour but rather a measurement tool” (p.187; [Jang et al., 2012](#)). Note, again, that models of recognition memory have extensively used confidence but rarely for the purpose of modelling metacognition (except for some accumulation of evidence models focused on decision making in general, e.g., [Ratcliff & Starns, 2013](#)). Therefore, it seems that metamemory models have rather focused on prospective judgements rather than retrospective judgements. Note that whereas prospective judgements take many forms according to the process measured (learning, encoding, failed retrieval), retrospective judgements are all of the form: RCJ.

## 2.3 Conclusions

The comparison of the metaperception and metamemory fields has shown that the main difference occurs in the type of judgements that it proposes to explain rather than the nature of the first-order decision. It seems that although models of confidence have been largely developed in the context of visual perception (but also in other domains; see for instance [De Martino et al., 2013](#) in value-based decision making, and see [Snizek, 1992](#) in group decision making) they pertain to metacognition in decision making (i.e., metadecision). Similarly, models of confidence in memory recognition, that is seen as a decision making process, have been developed (e.g., [Ratcliff & Starns, 2009](#); [Ratcliff & Starns, 2013](#)). Whilst RCJs have been studied in a variety of decision making tasks such as visual perception (e.g., [Rounis et al., 2010](#)), value-based decision making (e.g., [De Martino et al., 2013](#)), memory (e.g., [Koriat et al., 1980](#)), or in group decision making (e.g., [Snizek, 1992](#)), confidence has been poorly compared across these first-order tasks (for an overview of cross-task comparison studies see Chapter 3 and Chapter 4). Models of decision making that include confidence propose the involvement of processes that could be applied to confidence formation regardless of the nature first-order task (e.g., post-decision evidence accumulation). Some models have indeed been applied for memory and visual perception decisions ([Ratcliff & Starns, 2013](#)). Thus, these considerations suggest that confidence can be domain-general (i.e., independent from the first-order task, see Chapter 3 for more details). In this thesis, we will investigate the domain-generalty of metacognition.

A very large literature has also used confidence in recognition memory in order to infer memory processes. From this perspective, confidence is not seen as a metacognitive process per se but as a result of trace strength or processes involved in memory (e.g., [Wixted, 2007](#); [Yonelinas, 1994](#); [Yonelinas, 2002](#)). On the contrary, other metacognitive

judgements that are mainly prospective and memory-dependant (e.g., EOLs and JOLs are directly related to the learning process) had a particular focus. They differ from retrospective confidence in both the quantification of the accuracy of these judgments and theoretical frameworks that explain them. Thus, we can see a picture emerging: metadecision models focus on retrospective judgements whereas metamemory models rather focus on prospective judgements. It is possible that the distinction made in terms of first-order task (memory or perception) rather pertains to a distinction in terms of judgement types. Furthermore and as observed in the literature, prospective judgements are less accurate than retrospective judgements (e.g., [Fleming et al., 2016](#); [Kelemen et al., 2000](#)) probably because they rely less on the first-order evidence. This idea is also suggested by theoretical frameworks of confidence that are actually extension of decision-making models (e.g., post-decisional models). In this thesis, we will handle this question by comparing these two types of judgements in different first-order tasks (see Chapter 5 and Chapter 6).

Nonetheless, similarities across models of retrospective confidence and prospective judgements can be made. In the second-order model of confidence, the idea of proxy of confidence (i.e., information that is used to infer confidence) emerges and echoes the notion of metacognitive cues. The computation of confidence would therefore be based on inferences from these cues that are available during the task (as for other metacognitive judgements). Moreover, it is possible to consider that these proxies of metacognitive judgement are driven by the experience of the task (low level metacognition see Chapter 1) such as fluency or the action of the first-order decision and driven by higher beliefs (high level metacognition see Chapter 1). Recent work has focused on this question and heuristics have also been found in visual perception decision making. For instance, [Maniscalco, Peters, & Lau, 2016](#) suggested that confidence is based on the heuristic of response-congruent evidence in favour of the selected response. If this is non-optimal in a laboratory context, using this heuristic in the real world is much more efficient as choices are rarely (or even never) limited to 2AFC but are rather between multiple alternatives. When cues are also important for the processing goal of the task (e.g., blur in a visual perception task, accessibility of partial information in a memory task) they are valid or diagnostic cues and judgements track task performance. In these situations, participants have a high metacognitive sensitivity. In other situations, cues can be non-diagnostic of task performance and a dissociation between first-order and second-order behaviour can occur. This notion of cue-diagnosticity may also echoes processes in models of confidence. From a second-order model of confidence point of

view, it can refer to the correlation between the variable decision (first-order evidence) and the confidence decision (second-order evidence): a higher correlation increases metacognitive sensitivity because diagnosticity is higher.

As suggested here, the cue-utilisation view of metacognition can be also applied to other models of metacognition such as the second-order model of confidence. Moreover and to the best of our knowledge, only this framework proposes an integration of both prospective and retrospective judgements (Koriat, 1997; Koriat & Bjork, 2006; Koriat et al., n.d.; Koriat & Levy-Sadot, 1999). Therefore, we suggest that it is a good candidate to investigate metacognition across domains. The domain-generality of metacognition will be evaluated according to experience-based and information-based metacognition. As proposed above, we suggest that experience-based metacognition mainly relies on the fluency heuristic. As subjective ease is likely to occur regardless of the task, experience-based metacognition is likely to be domain-general. In this thesis, we will also investigate the domain-generality of fluency in the formation of metacognitive judgements (see Chapter 6).



## Chapter 3

# Introduction to empirical chapters

**M**etacognition is a broad construct that can be measured using direct and indirect tests (Chapter 1). Direct tests of metacognition, particularly metacognitive judgements, have been extensively studied in the field of metaperception and metamemory (Chapter 2). Despite a long tradition of using metacognitive judgements separately within these fields, similarities in methodological and theoretical framework can be observed. Our aim here is to propose a more unified view of metacognition which moves towards the idea that metacognition can be domain-general. In particular, we have identified the cue-utilisation framework as a good theoretical candidate for studying potential processes involved the generality of metacognition. Moreover, we propose to use the meta-d' framework to quantify metacognition as it allows control of numerous biases (as developed in Chapter 2). This chapter will narrow our aims by defining the notion of domain-generality theoretically and by proposing different ways of handling the idea that metacognition can be more unified than it is the case in the literature so far.

### 3.1 Investigating metacognition across domains

#### 3.1.1 The domain-generality and domain-specificity debate

One major question in the study of human behaviour and the brain is whether cognition is domain-general. Such an idea suggests that different domains share resources and should involve similar brain regions and networks. It supposes that similar rules and processes underlie different behaviours. On the contrary, domain-specificity stands for a more independent view of different cognitive phenomena. The notion of specificity has been extensively linked to the notion of modularity. The modular perspective (Fodor, 1983) supports the idea that cognition is subdivided into systems which are

encapsulated (i.e., no access to other signal during information processing). According to Fodor<sup>1</sup>, modules are domain-specific in terms of “the range of questions for which a device provides answers (the range of inputs for which it computes analyses)” (p. 103, Fodor, 1983). A classical example is the evolution of the concept of intelligence. As Miller (2000) suggests “the existence of the g factor (the “general intelligence” factor) in psychometrics appears to contradict the strong modularity view of the mind” (p.42). However, recent works focus on the theory of multiple intelligences (Davis, Christodoulou, Seider, & Gardner, 2011).

The domain-general and domain-specific debate in cognition necessarily points to the definition of a cognitive domain. As highlighted by Frensch and Buchner (1999), “any attempt at defining the notion of domain in some absolute objective sense can never hope to succeed” (p. 142). It appears that the definition of a domain is relative to constraints leading to domain separations. For instance, what is acceptable as a domain can differ according to whether one focuses on the behavioural or neuronal reference. It can also differ according to the choice of measurement. Here we take Frensch and Buchner’s point of view of the domain-generality debate: the width of applicability of a constraint. In other words, determining the domain-generality or domain-specificity debate answers the question: “how widely applicable a particular theoretical statement of empirical finding is” (p.140, Frensch & Buchner, 1999). Thus, it suggests that we should be able to define rules in a particular context in order to assess their possible extension to another context. For instance, the question of whether a particular learning program that is efficient for verbal learning is also efficient for non-verbal learning pertains to the domain-general and domain-specific debate. Asking whether metacognition is being domain-general obliges us to define theoretical statements that we can generalise to other situations.

### 3.1.2 Domain-generality in metacognition

In the metacognitive literature, we propose two versions<sup>2</sup> of the domain-generality and domain-specificity debate. First, a domain-general view proposes that people use a common resource in their metacognitive judgements across different types of tasks. Thus, people who are accurate in their evaluations of one task should also be accurate for another task. We suggest that this is particularly interesting for tasks where first

---

<sup>1</sup>Whilst Fodor proposed that “low level” processes are modular, he also suggested that this modularity is less present for higher order processes and it is sometimes unclear how encapsulated a module is.

<sup>2</sup>Note that several other domain-general debates also exist (e.g., are there common processes involved in direct and indirect tests of metacognition?) but we decided to focus on metacognitive judgements.

order performance is not correlated, that we can, in this sense, consider as “domains”. Second, one can wonder if a general resource is involved in the formation of different types of metacognitive judgements (e.g., prospective and retrospective). Therefore, we propose to investigate how wide are metacognitive processes in terms of first-order tasks and types of judgements. A domain-specific view proposes that there are different metacognitive components at play in different tasks and judgements and therefore predicts that the accuracy of metacognition will differ across tasks and judgements. It is important to stress that these two visions of metacognition are not necessarily exclusive as it is possible to suppose both domain-specific and domain-general resources co-exist in the formation of metacognition (see Chapter 7 for a discussion).

In behavioural studies, there are different ways to assess the specificity or the generality of a cognitive process that we gather here in three main types. A first class of method refers to correlational studies. The involvement of general metacognition suggests that reports of performance should be correlated across different domains. As mentioned before, it suggests that metacognitive accuracy in one domain or judgement type should be correlated with the accuracy in another domain or judgement. This method is the most commonly used in the field of domain-general metacognition (see Chapter 4 for an overview). A similar method pertains to factorial analyses. This method searches independent latent variable(s) called “factors” that can explain variability of several observed variables. For instance, this has been extensively used in intelligence research. Second, it is possible to assess the domain-specificity of a process using functional independence. From this perspective, functions or processes are supposed to be independent if one variable has an effect on one process and no effect or the opposite effect on another process. In that sense, it is possible to observe the influence of cues on metacognitive judgements across domains. Finally, a last class of method is neuropsychological dissociations. The study of behaviour in patients with neurological or psychiatric disorders is informative. Therefore, a selective deficit in metacognition in domains despite a preservation in another domain suggests a dissociation between metacognition in these two domains (and therefore a domain-specificity). The same reasoning occurs with metacognitive judgements. In the context of metacognition, these methods can be used for both metacognitive bias and metacognitive sensitivity. As these two methods of quantification of metacognition are supposed to be independent, it is possible that one is more domain-general than the other.

## **3.2 Methodological considerations**

### **3.2.1 Objectives**

The core question of this thesis is the breadth of metacognition. As mentioned in the last section, we focus on two possible domain-general and domain-specific debates: one related to the first-order tasks and another related to the type of metacognitive judgements. Chapter 4 and Chapter 5 are correlational chapters that investigate separately these two questions. Chapter 4 assesses the domain-general of metacognition across four different first-order tasks, all using RCJs. Chapter 5 focuses solely on metamemory (both episodic and semantic) and compare cross-task correlations in metacognition for both FOKs and RCJs, revisiting a typical dissociation between episodic and semantic memory in the accuracy of FOK judgements. Chapter 6 takes an experimental approach in order to examine domain-general related to the first-order tasks and related to the type of metacognitive judgements using functional independence. Based on the cue-utilisation view of metamemory, we investigate a potential general cue for metacognition (fluency) in a visual perception task (Experiment 1 and 2) and in a memory task (Experiment 3) in both prospective and retrospective judgements. As outlined in Chapter 2, the notion of the quantification of metacognition is highly important in this field. Therefore, the next section focuses on the methodological choices that we made in order to the answer to our theoretical questions, especially since the empirical chapters are all presented in the format of scientific articles, where there is less scope for such a discussion and justification of methods.

Finally, in the appendices some work is presented on investigations of neuropsychological dissociations in metacognition (3 articles), although the questions posed in these articles were not –eventually– related to the main aim of this thesis. Appendix A provides a brief overview of the use of the metacognitive framework in neurological populations and reviews work related to metacognition and self-awareness in Multiple Sclerosis. Appendix B investigates metacognitive judgements in Multiple Sclerosis across memory and executive functions. Appendix C is focused on the comparison of different metacognitive judgements for short-term memory in Alzheimer’s disease.

### **3.2.2 Material selection**

The five experiments that we present in this thesis have similarities in terms of methods and materials selection. First, all first-order tasks were 2AFC tasks because of our

methodological choice to quantify metacognition (see next section). Second, as is the case in the metamemory field, we decided to use a percentage scale for the metacognitive scale. This allows a direct comparison with task performance that can be calculated in percentage as well. In order to have a wide range of possible responses for participants we decided to have an 11-point scale. We began by using a scale ranging from 0% to 100% (Chapter 4) however we changed the scale for experiments in Chapter from a scale ranging to 50% to 100% as we found that there was some ambiguity in the absolute meaning of the scale label 0% confident, given that chance level in 2AFC tasks is 50%. Finally, the metacognitive scale used in Chapter 5 ranges also from 50% to 100% but had only 6 points. Our reasoning for this change was related to the FOK measure that is the main focus in this chapter. FOKs can be seen as a mix between a judgements and a feeling similar to TOT that would be an all-or-none state (people are supposed to either have a TOT or not). As such, FOKs have been mainly measured with yes/no scales (e.g., [Hart, 1965](#); [Souchay & Isingrini, 2012](#)). We thus decided to lower our scale to a 6-point in the experiment in Chapter 5.

In the metaperception field most studies have a high number of trials but a lower number of participants (especially for cross-task correlations, see Chapter 4). Here, however, we are constrained by the limits of human memory, where in episodic memory tasks in particular, there are a limited number of trials where we can reliably expect participants to encode and retrieve information. Thus, here we chose to have the opposite reasoning with fewer trials but a higher number of participants (minimum 100 per study).

### 3.2.3 Data and statistical modelling

As presented in Chapter 2, several means to quantify both metacognitive bias and metacognitive sensitivity exist. We now turn to the selection of a method for the calculation of both that is pertinent to the questions we address here. Our methodological choices differ according to our aims and therefore across correlational studies (Chapter 4 and Chapter 5) and experimental studies (Chapter 6).

Regarding metacognitive sensitivity, we decided to use the meta- $d'$  framework as this measure is the most independent from both first-order performance and metacognitive bias. We chose to estimate meta- $d'$  in a Bayesian framework instead of fitting it with minimisation of squared-sum-error or with maximum likelihood estimation. As it is important for the understanding of our statistical analyses, we (voluntarily briefly

as it is beyond the scope of this thesis) explain the logic of Bayesian statistical modelling (for more details see for instance [Wagenmakers et al., 2018](#)). Broadly speaking, the Bayesian approach uses probabilities to model uncertainty. Prior information is encoded into a prior distribution  $p(\theta)$  representing what is known about model parameters before seeing any data. The likelihood,  $p(D|\theta)$  can be seen as a hypothesis made about the data generating process. After seeing some data, the prior distribution is updated into a posterior distribution,  $p(\theta|D)$  following Bayes rule which states:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \quad (3.1)$$

The probability of the data  $p(D)$  (the marginal likelihood) is a normalising constant which allows that the area under the posterior distribution is equal to 1. When  $p(D)$  is not calculable, one can use algorithms to approximate the posterior distribution. In the meta-d' model, posterior approximation is made using Gibbs sampling (i.e., a type of MCMC algorithm) implemented in JAGS ([Plummer, 2003](#)).

We see several advantages of the use of a Bayesian estimation of meta-d' in this thesis. First, it naturally handles zero cell counts and avoids the use of edge correction which may bias other estimates such as maximum likelihood ([Fleming, 2017](#)). These corrections can occur quite often (e.g., highest confidence level for an incorrect response) which may especially bias estimates when the number of trials is low (as in all our experiments in this thesis) and as the scale has many points, zero cell counts are more likely to occur (11-point scale are used here except in Chapter 5). Second, Frequentist estimates of hit and false alarm rates fail to take into account uncertainty. Thus, point estimates of meta-d' do not provide an estimate of the precision of each single-subject estimate. This estimation is especially useful when group-level analyses are of interest (e.g., cross-task correlations) as a Bayesian approach can take into account the uncertainty about single-subject parameter estimates at the group level, and thus naturally handles both within- and between-participant uncertainty. For instance, it is possible that two participants have the same meta-d' estimate (the mean of the distribution;  $\mu = 1$ ) whilst having different degree of uncertainty (the variability of the distribution;  $\sigma = 0.5$  for Nina and  $\sigma = 0.1$  for Lisa). As the meta-d' estimation of Lisa is more accurate, it is assumed that her estimate is more reliable and should therefore have a bigger weight in the group-level estimate.

Crucially, the Bayesian framework proposed by [Fleming \(2017\)](#) to compute meta-d' also includes a hierarchical estimation of metacognitive sensitivity. More specifically,

the model HMeta-d allows a group-level estimation of Mratio<sup>3</sup> for which a prior is also given (known as hyperprior; see Fleming, 2017 details on priors and hyperpriors in the model). In such hierarchical models, subject-level estimations inform the group-level parameters that also inform subject-level estimations. This allows a more accurate estimation of subject-level parameters by allowing the group-level estimates to constrain subject-level fits, and more stable group-level estimates by limiting the impact of single-subject estimates with high uncertainty on the group.

The model can be easily extended to estimate jointly several group-level Mratios from a bivariate Gaussian distribution. When one Mratio is estimated per first-order task, the covariance between parameters can also be estimated in the hierarchical framework. Then the correlation parameters can be extracted from the covariance matrix:

$$\begin{bmatrix} \log(M1_s) & \log(M2_s) \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu_{M1} \\ \mu_{M2} \end{bmatrix}, \begin{bmatrix} \sigma_{M1}^2 & \rho_{M1M2}\sigma_{M1}\sigma_{M2} \\ \rho_{M1M2}\sigma_{M1}\sigma_{M2} & \sigma_{M2}^2 \end{bmatrix} \right) \quad (3.2)$$

This framework is particularly useful to measure cross-task correlations for metacognitive efficiency and we therefore used it in Chapter 4 and Chapter 5. In Chapter 6, we opted for a non-hierarchical estimation of meta-d'. Our main hypothesis in Chapter 6 is about metacognitive bias. Thus, we decided to first focus on the bias which then contained our measurement choice for metacognitive sensitivity. We chose to mainly measure bias using the recently developed m-distance index as it falls onto the SDT framework as meta-d'. Because m-distance needs the use of single-subject meta-c2 to be calculated, we could not use a hierarchical estimation of meta-d' for metacognitive sensitivity as single-subject parameters are also influenced by group-level parameters in multilevel models. Thus, we performed single-subject Bayesian estimations of parameters in Chapter 6.

We adapted the HMeta-d matlab code to allow an estimation of parameters in R (adapted code is freely available on GitHub at [https://github.com/amazancieux/HMeta-d/tree/R\\_functions/R](https://github.com/amazancieux/HMeta-d/tree/R_functions/R) using the "rjags" package. As in the HMeta-d toolbox, we discarded early samples of the posterior distributions and ran three chains in order to diagnose convergence problems. Convergence diagnostics were computed with the "coda" package using the "potential scale reduction factor"  $\hat{R}$  (Gelman & Rubin, 1992). Three functions have been adapted: one that allows a single-subject meta-d' parameter (used

<sup>3</sup>This group-level parameter is actually a logarithmic Mratio to avoid extremely high values in the case of very low first-order performance (see Chapter 2).

in all experiments of Chapter 6), another that allows both single-subject and group parameters of  $M_{ratio}$ , and a last that allows single-subject estimations and group estimations of  $M_{ratio}$  for two, three, or four tasks with estimations of cross-tasks correlational parameters (used in Chapter 4 and Chapter 5).

### 3.3 Aims of the thesis

The main idea of this thesis is to propose a more unified view of metacognitive research, which has thus far been shown to be disparate (Chapter 1) and mainly studied within fields (Chapter 2). Our rationale is to use a methodological and theoretical framework developed in a particular context, one that we thought likely to be generalised as our core question relates to the domain-generalty of metacognition. We define domain-generalty as theoretical statements that we can generalise to other situations. The comparison of the field of metamemory and metaperception in Chapter 2 suggests that the metacognition literature has grown up in somewhat disparate pockets of explanation which does not help in defining theoretical concepts which are common to all domains. Nonetheless, domain-generalty can refer to either the type of first-order task (e.g., memory or perception) that is used or the type of judgements (prospective and retrospective) that are made. With this respect, the cue-utilisation framework seems pertinent to handle the domain-generalty question. Particularly, we have argued in Chapter 2 that experience-based metacognition can be typified by the fluency heuristic that seems to be common in prospective judgements (see Chapter 6 for an overview in metamemory) as well as in retrospective judgements (see similarities made with models of confidence formation) regardless of the nature of the first-order task. In this thesis, we will therefore investigate the following questions:

- Is there a common resource in metacognition in terms of first-order tasks?
- Is this common resource involved in both prospective and retrospective judgements?
- Is experience-based metacognition a common process used across episodic memory and visual perception for both prospective and retrospective judgements? With this in mind, are visual perception and memory biased by fluency?

These questions will be handled for both metacognitive sensitivity and metacognitive bias using the meta- $d'$  developed in the context of metaperception that we will also apply here to metamemory and for the first time to prospective metamemory.



Finally, this thesis has an Open Science approach that we think to be crucial for increasing reproducibility and transparency of research. Thus, we systematically pre-registered<sup>4</sup> our experiments on the OSF website. Preregistration delineates between confirmatory and exploratory analyses “by requiring researchers to state how they will analyze the data before they observe it allowing them to confront a prediction with the possibility of being wrong” (p. 2605; [Nosek, Ebersole, DeHaven, & Mellor, 2018](#)). Moreover, all materials, data, and analyses scripts are also freely available on the OSF.

---

<sup>4</sup>For transparency reasons, we want to stress that initial analyses in Chapter 4 were planned using the SDT framework but with AUROC2 and that the switch to the Bayesian meta-d' framework was made after Fleming's (2017) paper.

## Chapter 4

# Is there a G factor for metacognition? Correlations in retrospective metacognitive sensitivity across tasks

Is metacognition a general resource shared across domains? Previous research has documented consistent biases in judgments across tasks. In contrast, there is debate regarding the domain-generality or the domain-specificity of the ability to discriminate between correct and incorrect answers (metacognitive sensitivity) because most previous work has documented non-significant correlations across domains. However, such null findings may be due to low statistical power and differences in task structure or performance, thereby masking a latent domain-generality in metacognition. We examined across-domain correlations in confidence level and sensitivity in a large sample (N=181). Participants performed four two-alternative-forced-choice tasks (episodic memory, semantic memory, executive function, and visual perception) with trial-by-trial confidence judgments. We found significant correlations in average confidence level across tasks. By applying a hierarchical Bayesian model to estimate cross-task covariance, we found five out six cross-task correlations in metacognitive efficiency (meta- $d'/d'$ ) were significant, even for pairs of tasks in which first-order performance was not correlated. This suggests that at least some components of metacognitive efficiency in retrospective confidence are domain-general.

## 4.1 Introduction

Metacognition refers to the ability to monitor and control cognitive processes (Flavell, 1979). It is often studied with reference to memory (e.g., Nelson, 1990) but has also recently been quantified for other domains such as visual perception (e.g., Song et al., 2011) decision making (e.g., Yeung & Summerfield, 2012) and motor tasks (e.g., Simon & Bjork, 2001). A critical research question therefore concerns the cross-domain organisation of such metacognitive evaluations of cognition. The core question of this article is whether metacognition is a specific process particular to each cognitive domain (e.g., language, memory, perception) or whether it is a higher-order process with some overlap across multiple cognitive domains. A domain-general view of metacognition proposes that people use a common resource when they evaluate their performance across different types of tasks. In contrast, a domain-specific account proposes that there are different metacognitive components at play in different tasks. By leveraging individual differences it is possible to adjudicate between these two proposals. According to the domain-general view, people who have accurate judgements for one task should also make accurate judgements for another. In contrast, if metacognition relies on domain-specific components, we would expect such abilities to be uncorrelated. The focus of this paper is to investigate this issue using RCJs. RCJs are self-evaluations of certainty in a given response and are appropriate for addressing the question of domain-generality, as they can be applied to decisions made across a variety of tasks.

In the current study, we focus on assessing the domain-generality of both metacognitive bias and sensitivity, two measures which map onto two different aspects of metacognition. Metacognitive bias refers to the overall magnitude of a judgment, such as whether an observer has a tendency to report high or low confidence, irrespective of their performance. Metacognitive sensitivity refers to the ability of a person to discriminate between different levels of performance, such as correct or incorrect trials (Fleming & Lau, 2014).

Previous research using RCJs has provided equivocal findings for metacognitive sensitivity. Whereas a few studies have found positive correlations between metacognitive sensitivity for memory and visual perception tasks (McCurdy et al., 2013; Lee, Ruby, Giles, & Lau, 2018), a majority concluded in favour of domain-specificity due to non-significant correlations (Baird, Cieslak, Smallwood, Grafton, & Schooler, 2015; Baird et al., 2013; Fitzgerald et al., 2017; Morales et al., 2018). Regarding structural MRI data, distinct cerebral areas correlating with individual variation within two tasks has

been observed, also supporting the possibility of neurofunctional independence between domains (Baird et al., 2015; Baird et al., 2013; McCurdy et al., 2013). Specifically, metacognitive sensitivity in a visual perception task has been related to the volume and function of lateral anterior PFC, whereas metacognitive sensitivity in a memory task is associated with the structure and function of precuneus and medial anterior PFC. Accordingly, lesions to anterior PFC have been shown to selectively affect visual perceptual sensitivity while sparing sensitivity on the memory task (Fleming et al., 2014).

However, a recent meta-analysis of cross-domain correlations in metacognitive sensitivity pointed to a heterogeneous pattern of domain-generality (Rouault, McWilliams, Allen, & Fleming, 2018). Although there was an overall cross-domain correlation between different perceptual tasks (e.g., visual, auditory, tactile; see for instance Ais et al., 2016, and Faivre, Filevich, Solovey, Kühn, & Blanke, 2018) there was equivocal evidence for domain-generality across visual perception and memory tasks. Moreover, it was noted that drawing conclusions about domain-specificity relies on accepting the null hypothesis of no correlation, which is problematic if individual experiments are underpowered to detect a correlation. In addition, it was recognised that cross-domain correlations may also be biased by inconsistencies in the sensitivity index calculated in these studies and variability in task structure between domains.

A first important consideration is the method used to assess metacognitive sensitivity. Different techniques are often used to compute sensitivity which makes it difficult to compare results across studies. Moreover, several of these indexes (such as gamma correlation or AUROC2) do not control for the effect of task performance (Fleming & Lau, 2014), and spurious correlations in metacognitive sensitivity may emerge between domains that are driven by variation in task performance (i.e., first-order performance), rather than metacognitive capacity itself (i.e., second-order performance; Rouault, McWilliams, et al., 2018). One recent measure that achieves this control is metacognitive efficiency,  $\text{meta-}d'/d'$ . The  $\text{meta-}d'$  framework models the relationship between performance and metacognition using SDT.  $\text{Meta-}d'$  is defined as the Type 1  $d'$  that would lead to the observed AUROC2 in the absence of noise or imprecision in confidence estimates (Maniscalco & Lau, 2012). Metacognitive efficiency is then defined as the level of metacognitive sensitivity ( $\text{meta-}d'$ ) of a subject relative to the subject's actual Type 1 performance. By estimating  $\text{meta-}d'$  in a Bayesian hierarchical framework (Fleming, 2017) it is possible to directly estimate covariance in metacognitive efficiencies across domains.

A second possible explanation for inconsistencies between results of previous studies is that different task designs have been used in different domains. For instance, several studies have compared metacognitive sensitivity between 2AFC perceptual tasks and yes/no recognition memory tasks. As recently (Lee et al., 2018), these differences in task structure may obscure across-domain correlations in metacognitive ability, particularly given potential asymmetries in metacognitive ability for yes and no responses (Kanai, Walsh, & Tseng, 2010; Meuwese, van Loon, Lamme, & Fahrenfort, 2014). Here we focus on comparing between different 2AFC tasks which are appropriate for fitting an equal-variance meta-d' model. Unlike the debate surrounding metacognitive sensitivity, there is greater agreement in previous literature that metacognitive bias is relatively stable across tasks. People tend to be overconfident in their judgments of general knowledge (Lichtenstein & Fischhoff, 1977) and visual perception (Baranski & Petrusic, 1994; Song et al., 2011), and this degree of confidence is correlated across tasks (Ais et al., 2016). Moreover, the hard-easy effect – overestimation in difficult tasks and underestimation in easy tasks – has also been found in both types of task (e.g., Baranski & Petrusic, 1995). In sum, while previous studies support a domain-generalty in metacognitive bias, both neuroimaging and behavioural findings, albeit in small samples, remain equivocal about the domain-generalty of metacognitive sensitivity.

On a theoretical level, models of metacognition have been developed in two distinct fields: metamemory (metacognition about memory) and metaperception (metacognition about perceptual decision making). Although these frameworks have developed independently, common points can be highlighted. Models of confidence formation in perceptual decision making suggest that confidence is based on a computation of a probability that a decision is correct. A dominant view supports the idea that confidence relies on both evidence from the first-order decision and additional computations beyond this such as post-decisional processes (Navajas et al., 2016) or 'second-order' inference (Fleming & Daw, 2017).

Similarly, in metamemory, the amount and quality of evidence is proposed to be critical in supporting a confidence estimate (e.g., Koriat et al., 1980). One component of such evidence are cues that are intrinsically related to memory processes (e.g. extrinsic information such as number of stimuli to encode, relatedness between targets and distractors, Koriat, 1997), equivalent to the notion of sensory evidence in perceptual decision-making. However, as in metaperception, metamemory confidence (and other metacognitive judgements) is thought to also be inferred from additional information that may not be used to guide first-order memory responses. In the metaperception

field, confidence has been modeled using extensions of SDT and evidence accumulation frameworks, whereas the computational distinction between first- and second-order processes in memory has received less attention. For instance, according to the stochastic detection and retrieval model (Jang et al., 2012), a first sample of evidence informs a recall or recognition response and a second sample of evidence supports the formation of confidence. This model, as in related models of perceptual confidence (Fleming & Daw, 2017), suggests that additional computations (that can more or less correlated with a first-order decision computation) are used to inform confidence judgements. It is therefore possible that both domain-specific (i.e., internal perceptual or mnemonic states supporting first-order decisions in each task) and domain-general resources (i.e., post-decisional computations that could be common across tasks) contribute to confidence judgements in the two domains.

Motivated by these theoretical issues, the aim of the present study was to compare metacognitive judgments across four different 2AFC cognitive tasks and to ask whether correlations in bias (measured by confidence level) and/or sensitivity (measured by meta- $d'$ ) are indicative of a common underlying process of metacognition. The idea was to quantify potential domain-general contributions to metacognition while keeping the task structure similar across first-order decisions. As noted above, it remains possible that an absence of correlations regarding metacognitive sensitivity is explained by a lack of statistical power, as the sample sizes of previously mentioned studies ranged from 23 to 52 participants. It is however important to note that these studies are mainly neuroimaging studies that did not directly aim to test cross-task correlations in behavioural measures of metacognition. To test a correlation hypothesis, it has been suggested that "there are few occasions in which it may be justifiable to go below  $n = 150$ " to obtain stable and reliable correlations (Schönbrodt & Perugini, 2013, p.10). Here we employ a large sample ( $N = 181$ ) based on a priori power calculations and compute the covariance of meta- $d'/d'$  estimates in a hierarchical Bayesian framework, thereby maximizing the sensitivity of our analysis approach to detect shared variance across domains.

## 4.2 Method

### 4.2.1 Participants

The current experiment was conducted in the Laboratoire de Psychologie et Neurocognition in Grenoble, France, and included 181 young adults ( $M = 20.01$ ,  $SD = 3.13$ ; 84%

of women) recruited through an advertisement at the Grenoble-Alpes University. We estimated the required sample size according to Schönbrodt and Perugini (2013) using an expected correlation of 0.4 between metacognitive sensitivity on a memory and a perceptual task (McCurdy et al., 2013). The authors explained that "the true correlation strength uncontaminated by outlier influence, although significant, is likely to be lower than the  $r$  value of 0.471" (p.4), hence our more conservative estimate of 0.4. According to Schönbrodt and Perugini (2013), for a correlation of 0.4 and 80% of power, correlations begin to be stable for 181 participants. All participants were native French speakers and reported having a normal or corrected-to-normal vision. The study was preregistered on the OSF (<https://osf.io/b5ype/>) and preregistered analyses are presented in Supplementary material. We report here non-preregistered analyses (see data and statistical analyses section).

## 4.2.2 Materials and procedure

The entire procedure included four cognitive tasks: an episodic memory task, a semantic memory task, an executive functioning task, and a visual perception task. Task order was randomly assigned for each participant. See Figure 1 for examples and a schematic representation. The episodic memory task was separated into two parts: an encoding phase and a retrieval phase. During the encoding phase, participants were presented with 40 unrelated pairs of words for 2500ms duration in a randomized order. Words were extracted from the French Lexique database (New, Pallier, Brysbaert, & Ferrand, 2004) according to the following criteria: nouns or adjectives with six letters, two syllables, and between 20 to 100 occurrences per million. During the retrieval phase, immediately after the end of the encoding phase, participants were presented with a cue word seen during the encoding phase and had to select which one of the two other presented words was paired with this cue word. Participants had no time limit to give their answer. Distractors were other words extracted from Lexique according to the same criteria as targets and cues. These 2AFC decisions in this task, and in the following, are referred to as the 'first order' task

In the semantic memory task, participants performed a series of 2AFC decisions for general knowledge questions specifically designed for the French participants in this study. These questions included various topics such as cinema, sport, art, history, and geography (e.g., What is the largest department in France? Which painter is the main representative of Cubism?). We pretested the difficulty of 60 questions in 20 participants by calculating the percentage correct for each question. From these 60 questions,

20 were excluded because they were either too easy (above 95% correct answers) or too difficult (below 5% correct answers). Participants had no time limit to give their answer.

The visual perception task was akin to the one used by Fleming et al. (2014) and consisted of two circles (diameter of 11.5°) each containing dots presented for 700ms. After stimuli presentations, participants responded as to which one of the two circles contained more dots with no time limit. Before each new stimulus presentation, participants had to press the space bar. One of the two circles always contained 50 dots and the other either had fewer than or more than 50 dots, randomly defined on each trial. Stimuli were created using a plot function in R software. For each stimulus the number of dots was randomly defined – between 25 and 49 for stimuli with fewer dots and between 51 and 75 for stimuli with more dots.

The fourth task consisted of an attention, flexibility and working memory (executive function) task. Participants were presented a letter-number sequence of five symbols for 1000ms. Half of these sequences had three letters and two numbers and the other half had two letters and three numbers (e.g., 7A5N2). Participants chose which one of the two presented responses corresponded to the sum of all numbers and the relevant letters (in the example above the correct answer would be 14AN). They had no time limit to give their answer and had to press the space bar before each new stimulus presentation. All stimuli were made prior to the task by associating random letters (from A to Z) with numbers (from 0 to 9). Distractors were made by changing either one letter or the sum of all numbers (e.g., if the correct response is 14AN, distractors can be either 16AN or 14BN) from the correct answer. All stimuli had the same structure with numbers embedded in strings of letters.

All four tasks comprised 40 trials each and had similar response requirements. The position of the correct answer was randomly assigned and the order of the four tasks was randomised for all participants. To begin each trial participants pressed the “space” bar. For the first-order decision, participants had to press the “s” letter to select the lefthand answer and the “l” letter to select the righthand answer and they had no time limit for make their decision. Figure 4.1 provides a summary of the four tasks.

After each response on each of the four tasks, participants were asked to evaluate how confident they were in their answer. The scale ranged from 0% of confidence (minimum confidence) to 100% (maximum confidence). Participants could report 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% by using the number keys 0 to 9. Participants used “c” to report 100% confidence. It was explained to the participants that



0% confidence signified a guess response. There was no time limit for either first-order decisions or confidence judgments and participants were not asked to respond as quickly as possible; however we measured decision time in an exploratory analysis.

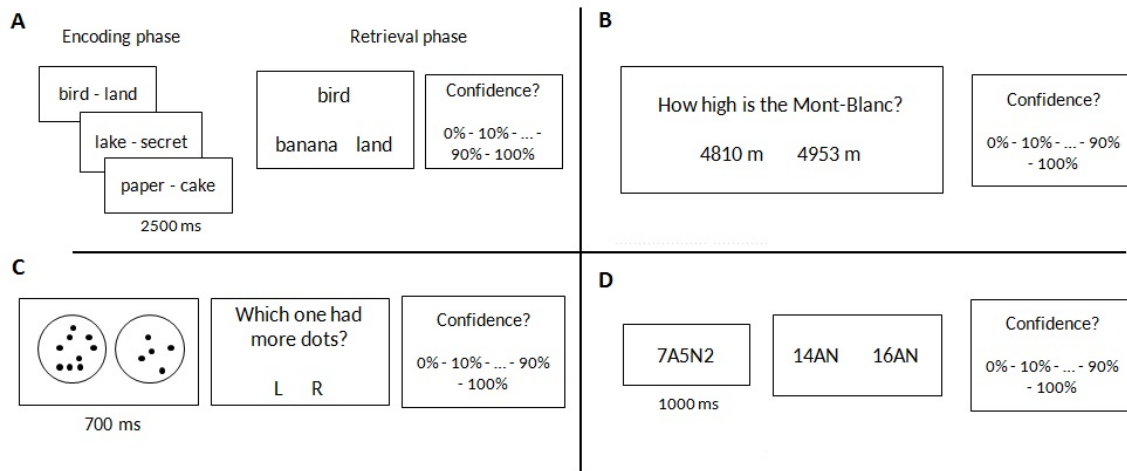


Figure 4.1: Summary of the four tasks. (A) Episodic memory task. (B) Semantic memory task. (C) Visual perception task – real stimuli included between 25 and 75 dots. (D) Working memory/attention task (executive functioning).

### 4.2.3 Data and statistical analyses

As described above, we focused on both metacognitive bias and metacognitive sensitivity. In our initial preregistration, we aimed to measure metacognitive bias by subtracting mean task performance from mean confidence, because we anticipated that first-order performance would differ across the four tasks. Metacognitive sensitivity was proposed to be measured by the AUROC2. We decided to deviate from both of these planned analyses for several reasons (see Supplementary results section for preregistered analyses).

Regarding metacognitive bias, we reasoned that there was some ambiguity in the absolute meaning of the scale label 0% confident, given that chance level in 2AFC tasks is 50%. We therefore decided to measure the average confidence level across trials without subtracting mean task performance, which would rely on subjects having interpreted a scale value of 0% confidence as 50% performance (chance).

We chose to estimate metacognitive efficiency (meta- $d'/d'$ ) – i.e. metacognitive sensitivity corrected for differences in performance – when comparing cross-task correlations in metacognitive capacity. This is because measures of metacognitive sensitivity (such as gamma correlation and AUROC2) are sensitive to differences in first-order performance (e.g., Fleming & Lau, 2014), rendering such scores inappropriate for the current study in which task performance varied across both domains and participants. Using AUROC2, for instance, it is possible that cross-task correlations at the metacognitive level could be partly or fully driven by correlations in first-order performance. The meta- $d'$  framework allows us to control for such variability. In Type 1 SDT,  $d'$  refers to the ability to discriminate between different states of the world (i.e., signal and noise). This parameter can be calculated as  $d' = z(\text{hits}) - z(\text{false alarms})$ , where  $z$  is the inverse of the cumulative normal distribution function, hits are the proportion of "signal" responses when signal is present, and false alarms are the proportion of "signal" responses when noise is present (here, signal was defined arbitrarily as one of the two response options, as two stimulus options were presented on each trial of the 2AFC tasks). In Type 2 SDT, the sensitivity parameter of interest is the ability to discriminate between correct and incorrect responses, rather than signal and noise. Meta- $d'$  refers to the Type 1  $d'$  that would give rise to the observed confidence distributions in the absence of noise or imprecision in the ratings. By modelling the relationship between Type 1 and Type 2 performance (the more information available for the Type 1 task, the more sensitive Type 2 confidence ratings should be), meta- $d'$  quantifies the sensitivity of confidence ratings to performance in units of  $d'$  (Maniscalco & Lau, 2012). Because  $d'$  and meta- $d'$  are in the same units, they can be compared which allows derivation of a measure of metacognitive efficiency, controlling for task performance. If this measure (Mratio; meta- $d' / d'$ ) is close to 1, then metacognitive efficiency is optimal under the SDT model.

Here we used a recent hierarchical Bayesian framework (Fleming, 2017) to estimate meta- $d'/d'$  at the group level (HMeta- $d$ ). This allows a more accurate estimation of subject-level parameters by allowing the group-level estimates to constrain subject-level fits, and more stable group-level estimates by limiting the impact of single-subject estimates with high uncertainty on the group. (Fleming, 2017) showed in simulation that HMeta- $d$  was able to recover stable group-level parameter estimates with as few as 50 trials per subject, which was not the case when averaging single-subject maximum likelihood fits. This framework is also particularly useful to test the question of the domain-generalty of metacognition since it can also be used to estimate covariance

between estimates in a hierarchical framework.

As we have a low number of trial per task ( $N = 40$ ), a Bayesian estimation of meta-d' is more appropriate as it naturally handles zero cell counts and avoids the use of edge correction which may bias maximum likelihood estimates. Moreover, maximum likelihood estimates of parameters based on hit and false alarm rates fail to take into account uncertainty about these rates that is a consequence of finite data. A Bayesian approach takes into account the uncertainty about single-subject parameter estimates at the group level, and thus naturally handles both within- and between-participant uncertainty. This is particularly crucial in the current study given that uncertainty in the model's estimate of meta-d' needs to be incorporated into an assessment of any correlation between the two domains (see Supplementary results).

To extend the existing model, each subject's log metacognitive efficiency ( $\log(\text{meta-d}'/d')$ ) in the four tasks (M1, M2, M3, M4) was specified as a draw from a multivariate Gaussian:

$$\begin{aligned} & \left[ \log(M1_s) \quad \log(M2_s) \quad \log(M3_s) \quad \log(M4_s) \right] \sim \quad (4.1) \\ & \mathcal{N} \left( \begin{bmatrix} \mu_{M1} \\ \mu_{M2} \\ \mu_{M3} \\ \mu_{M4} \end{bmatrix}, \begin{bmatrix} \sigma_{M1}^2 & \rho_{M1M2}\sigma_{M1}\sigma_{M2} & \rho_{M1M3}\sigma_{M1}\sigma_{M3} & \rho_{M1M4}\sigma_{M1}\sigma_{M4} \\ \rho_{M1M2}\sigma_{M1}\sigma_{M2} & \sigma_{M2}^2 & \rho_{M2M3}\sigma_{M2}\sigma_{M3} & \rho_{M2M4}\sigma_{M2}\sigma_{M4} \\ \rho_{M1M3}\sigma_{M1}\sigma_{M3} & \rho_{M2M3}\sigma_{M2}\sigma_{M3} & \sigma_{M3}^2 & \rho_{M3M4}\sigma_{M3}\sigma_{M4} \\ \rho_{M1M4}\sigma_{M1}\sigma_{M4} & \rho_{M2M4}\sigma_{M2}\sigma_{M4} & \rho_{M3M4}\sigma_{M3}\sigma_{M4} & \sigma_{M4}^2 \end{bmatrix} \right) \end{aligned}$$

Priors were specified as follows:

$$\begin{aligned} \mu_{M1}, \mu_{M2}, \mu_{M3}, \mu_{M4} & \sim \mathcal{N}(0, 1) \\ \sigma_{M1}, \sigma_{M2}, \sigma_{M3}, \sigma_{M4} & \sim \text{InvSqrtGamma}(0.001, 0.001) \\ \rho_{M1M2}, \rho_{M1M3}, \rho_{M1M4}, \rho_{M2M3}, \rho_{M2M4}, \rho_{M3M4} & \sim \text{Uniform}(-1, 1) \end{aligned} \quad (4.2)$$

$\mathcal{N}$  is a normal distribution with mean and standard deviation as parameters.  $\mu_M$  and  $\sigma_M$  refer to the mean and the standard deviation of  $\log(\text{meta-d}'/d')$ .  $\rho_{MiMj}$  is the correlation coefficient for  $\log(\text{meta-d}'/d')$  between tasks  $i$  and  $j$ .

The HMeta-d toolbox (<https://github.com/metacoglab/HMeta-d>) uses MCMC sampling to estimate posterior distribution over model parameters using JAGS (Plummer, 2003). We modified the HMeta-d code to allow estimation of parameters in R using rjags. As in the HMeta-d toolbox, we discarded early samples of the posterior distributions and ran three chains in order to diagnose convergence problems. Convergence

diagnostics were computed with the "coda" package using the "potential scale reduction factor"  $\hat{R}$  (Gelman & Rubin, 1992). Material, raw data, model and analysis scripts are available in OSF (<https://osf.io/b5ype/>). Significance of group-level parameters was estimated by calculating whether the 95% HDIs on the posterior distributions of the correlation coefficients  $\rho_{M_i M_j}$  overlapped with zero, which is a Bayesian analogue of a frequentist confidence interval since it is the smallest interval containing 95% of the MCMC samples (Kruschke, 2014).

We complemented the HMeta-d analyses for metacognitive efficiency with non-hierarchical Pearson's  $r$  correlations and paired t-tests for magnitude of judgements and task performance. For paired t-tests, outliers were detected using 3 tests: Leverage, RSS and Cook's distance. When necessary, Bonferroni corrections were applied.

## 4.3 Results

### 4.3.1 Type 1 performance

We assessed task performance using Type 1  $d'$ . This index was calculated for each participant and each task (see Figure 4.2A for mean and confidence intervals). For these analyses a Bonferroni correction was used providing a significance threshold of  $\alpha = 0.05/6 = 0.008$ . Paired t-tests showed that performance on the executive function task ( $M = 2.58$ ;  $SD = 0.74$ ) was better than the episodic memory task ( $M = 1.84$ ;  $SD = 0.88$ ),  $t(180) = 9.42$ ,  $p < .001$ ,  $d_z = 0.70$ , semantic memory task ( $M = 1.19$ ;  $SD = 0.60$ ),  $t(180) = 22.71$ ,  $p < .001$ ,  $d_z = 1.69$ , and visual perception task ( $M = 0.92$ ;  $SD = 0.39$ ),  $t(180) = 30.26$ ,  $p < .001$ ,  $d_z = 2.25$ . The episodic memory task was also better performed than the semantic memory task,  $t(180) = 9.32$ ,  $p < .001$ ,  $d_z = 0.69$ , and the visual perception task,  $t(180) = 13.09$ ,  $p < .001$ ,  $d_z = 0.97$ . Finally, the semantic memory task was better performed than the visual perception task,  $t(180) = 4.98$ ,  $p < .001$ ,  $d_z = 0.37$ .

We next examined intersubject correlations in first-order performance across tasks. Table 4.1 summarises Pearson correlation coefficients between  $d'$  values. These analyses revealed a positive correlation between episodic and semantic memory performance,  $r = 0.23$ ,  $p = .002$ . Executive function performance was also positively correlated with semantic memory performance,  $r = 0.27$ ,  $p < .001$ , and visual perception performance,  $r = 0.21$ ,  $p < .001$ . However, correlations between other task performance

pairings (visual perception and episodic memory; executive function and episodic memory; semantic memory and visual perception) were not significant after correcting for multiple comparisons.

Table 4.1: Pearson correlation coefficients, confidence intervals, and  $p$  values for correlations in task performance between pairs of tasks. Alpha threshold is .008.

	Performance ( $d'$ ) correlations			
	Episodic memory	Visual perception	Semantic memory	Executive function
Episodic memory		$r = 0.04 [-0.11, 0.18]$ $p = .638$	<b><math>r = 0.23 [0.09, 0.37]</math></b> <b><math>p = .002</math></b>	$r = 0.16 [0.02, 0.30]$ $p = .030$
Visual perception			$r = -0.08 [-0.23, 0.06]$ $p = .258$	<b><math>r = 0.25 [0.11, 0.39]</math></b> <b><math>p &lt; .001</math></b>
Semantic memory				<b><math>r = 0.25 [0.11, 0.38]</math></b> <b><math>p &lt; .001</math></b>
Executive function				

### 4.3.2 Confidence level

Mean confidence judgments were calculated for each participant and each task (Figure 4.2B). The pattern of results for confidence judgements was similar to that for task performance. Paired t-tests (corrected for multiple comparisons) showed people were more confident overall on the executive function task than the episodic memory task,  $t(180) = 10.04, p < .001, d_z = 0.75$ , the semantic memory task,  $t(180) = 18.73, p < .001, d_z = 1.39$ , and the visual perception task,  $t(180) = 18.10, p < .001, d_z = 1.35$ . The episodic memory task was also judged with higher confidence than the semantic memory task,  $t(180) = 4.71, p < .001, d_z = 0.35$ , and the visual perception task,  $t(180) = 6.30, p < .001, d_z = 0.47$ . Finally, the semantic memory task was judged with higher confidence than the visual perception task,  $t(180) = 3.37, p < .001, d_z = 0.25$ .

In order to estimate domain-general influences on confidence level, we computed correlations between average confidence levels across tasks (Table 4.2). We observed a significant correlation between confidence levels across all tasks after correction for multiple comparisons (all  $p < .008$ , with  $r$  ranging from 0.21 to 0.39; the exception was a trend-level correlation between visual perception and episodic memory) suggesting that the more participants report high confidence in one task, the more they report high confidence in another task.

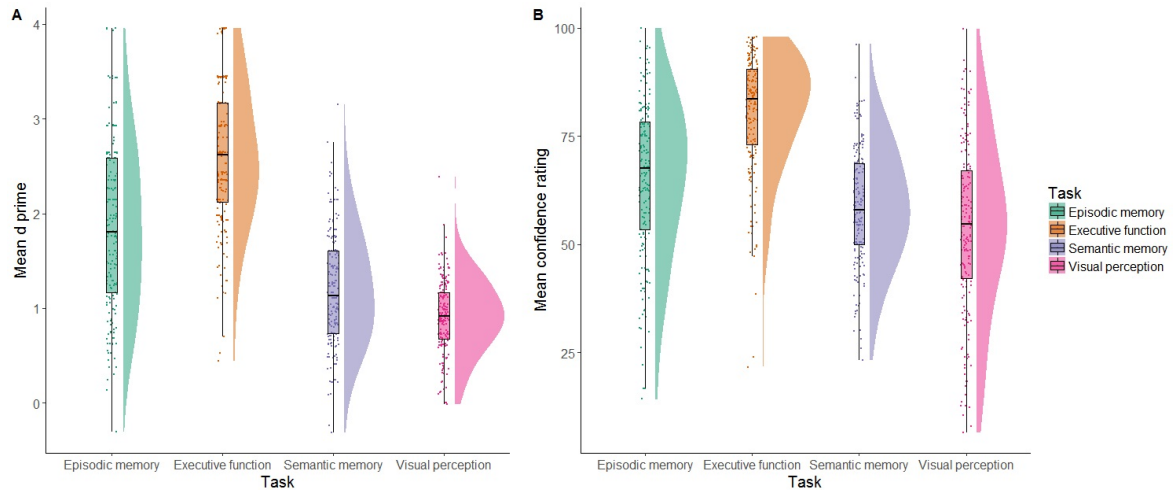


Figure 4.2: (A) Raincloud plots (Allen et al., 2019) for  $d'$  for the four tasks. (B) Raincloud plots for confidence level for the four tasks.

Table 4.2: Pearson correlation coefficients, confidence intervals, and  $p$  values for correlations in confidence between pairs of tasks. Alpha threshold is .008.

	Confidence level correlations			
	Episodic memory	Visual perception	Semantic memory	Executive function
Episodic memory		$r = 0.19$ [0.05, 0.33] $p = .009$	$r = 0.34$ [0.21, 0.46] $p < .001$	$r = 0.21$ [0.06, 0.34] $p = .005$
Visual perception			$r = 0.39$ [0.27, 0.52] $p < .001$	$r = 0.36$ [0.23, 0.48] $p < .001$
Semantic memory				$r = 0.37$ [0.23, 0.49] $p < .001$
Executive function				

### 4.3.3 Metacognitive efficiency

To estimate metacognitive efficiency, we estimated the group meta- $d'/d'$  ratio for each task (see Figure 4.3). According to the overlap of 95% HDIs, metacognitive efficiencies were similar for the two memory tasks, which in turn were greater than both the executive function and visual perception tasks (for means and HDIs related to the difference distributions for each comparison see Table 4.3). Executive function metacognitive efficiency was also greater than visual perceptual metacognitive efficiency.

In order to evaluate domain-general contributions to metacognitive efficiency, we estimated correlations between all four task pairings within the hierarchical model.

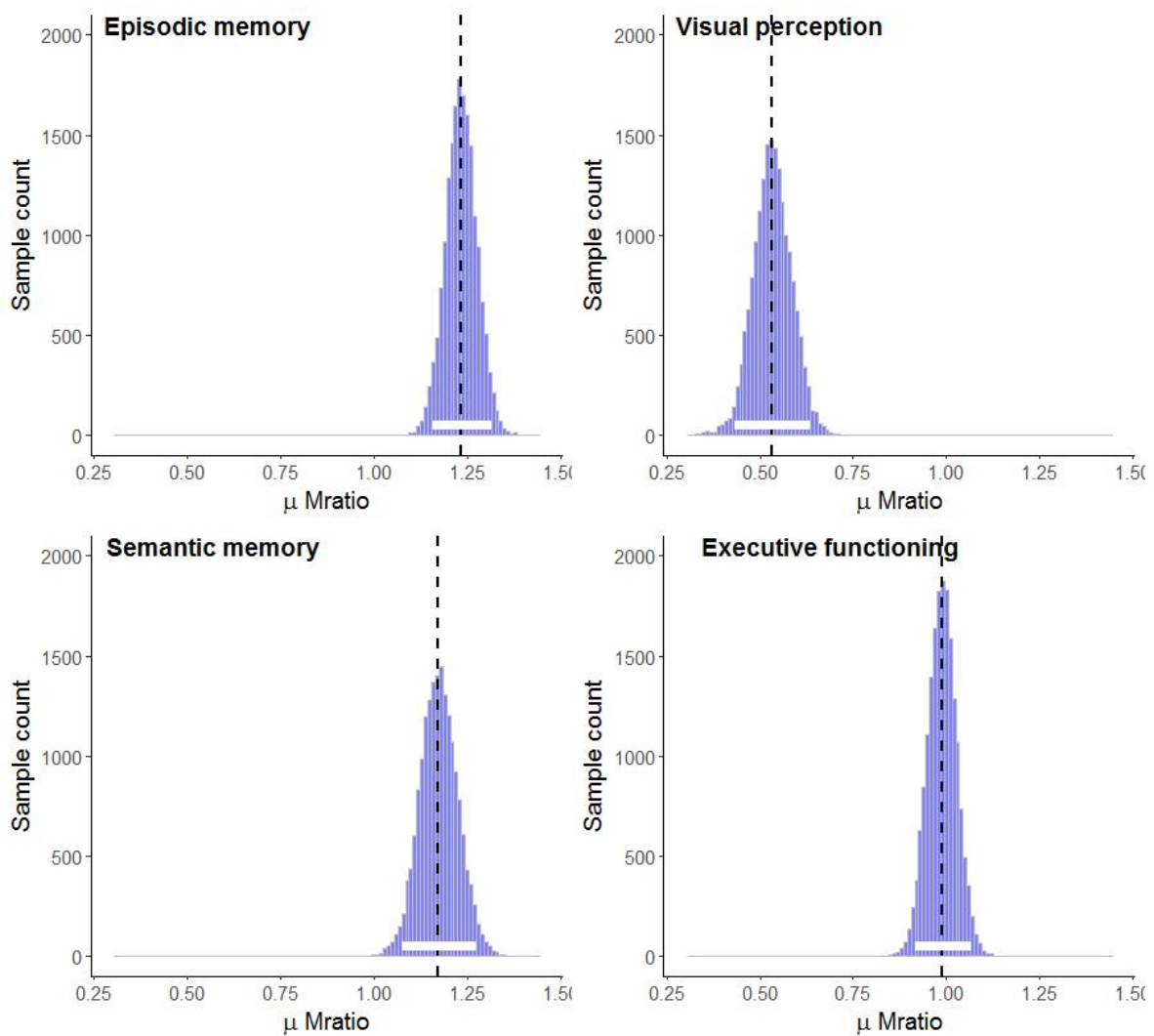


Figure 4.3: Posterior distributions over  $\mu$  Mratio (meta- $d'/d'$  ratio) for the episodic memory, visual perception, semantic memory and executive functioning tasks.

These correlations are estimated at the group-level from the variance-covariance matrix. Figure 4.4B presents posterior distributions over each cross-task correlation parameter and associated 95% HDIs are presented in Table 4.4. Figure 4.4A visualises the relationships between single-subject meta- $d'/d'$  values estimated within the hierarchical model. Critically, 95% HDIs on the posterior correlation coefficients for 5 out of 6 task pairings did not overlap zero suggesting substantial covariance in metacognitive efficiency across domains. This was also the case for task pairings for which we did not observe correlations in task performance (e.g. visual perception and semantic memory; Table 4.1), suggesting it is unlikely to be an artefact of covariance in first-order capacity. Only the HDI for the correlation between visual perception task and episodic memory

Table 4.3: Means and HDIs of the posteriors of the difference between  $\mu$  Mratio distributions for each task pairing. Only the difference distribution between episodic memory and semantic memory overlaps with 0, indicating no significant difference between tasks.

Difference distributions between group-level meta-d'/d' estimates				
	Episodic memory	Visual perception	Semantic memory	Executive function
Episodic memory		<b>0.84</b> [0.72, 0.97]	0.05 [-0.01, 0.11]	<b>0.22</b> [0.17, 0.28]
Visual perception			<b>0.79</b> [0.68, 0.91]	<b>0.62</b> [0.51, 0.75]
Semantic memory				<b>0.17</b> [0.11, 0.23]
Executive function				

task ( $\rho = 0.28$ ; HDI= [-0.03, 0.60]) overlapped zero, indicating a lack of cross-task correlation.

Table 4.4: Means and HDIs of the distribution of posteriors of the  $\rho$  value for each task pairing. Only the HDI for the correlation between episodic memory and visual perception overlap with 0, indicating a lack of cross-task correlation.

Group-level correlations in meta-d'/d'				
	Episodic memory	Visual perception	Semantic memory	Executive function
Episodic memory		0.28 [-0.03, 0.60]	<b>0.41</b> [0.14, 0.66]	<b>0.44</b> [0.24, 0.63]
Visual perception			<b>0.69</b> [0.36, 0.94]	<b>0.65</b> [0.35, 0.89]
Semantic memory				<b>0.41</b> [0.16, 0.65]
Executive function				

Although the current study has few trials per task, for completeness we nonetheless performed non-hierarchical estimation of subject-specific meta-d' in order to calculate a meta-d' / d' ratio per participant and per task. We excluded nine participants with very low performance ( $d' < 0.10$ ) in one of the four tasks. Then, we performed Pearson's correlations for metacognitive efficiency across tasks (Table 4.5). When controlling for multiple comparisons, we found positive correlations for meta-d'/d' across visual perception and semantic memory and across visual perception and executive function.

### 4.3.4 Supplementary results

#### Preregistered analyses

We estimated metacognitive bias using difference scores. Thus, the difference between mean confidence and mean performance was calculated for each participant and each



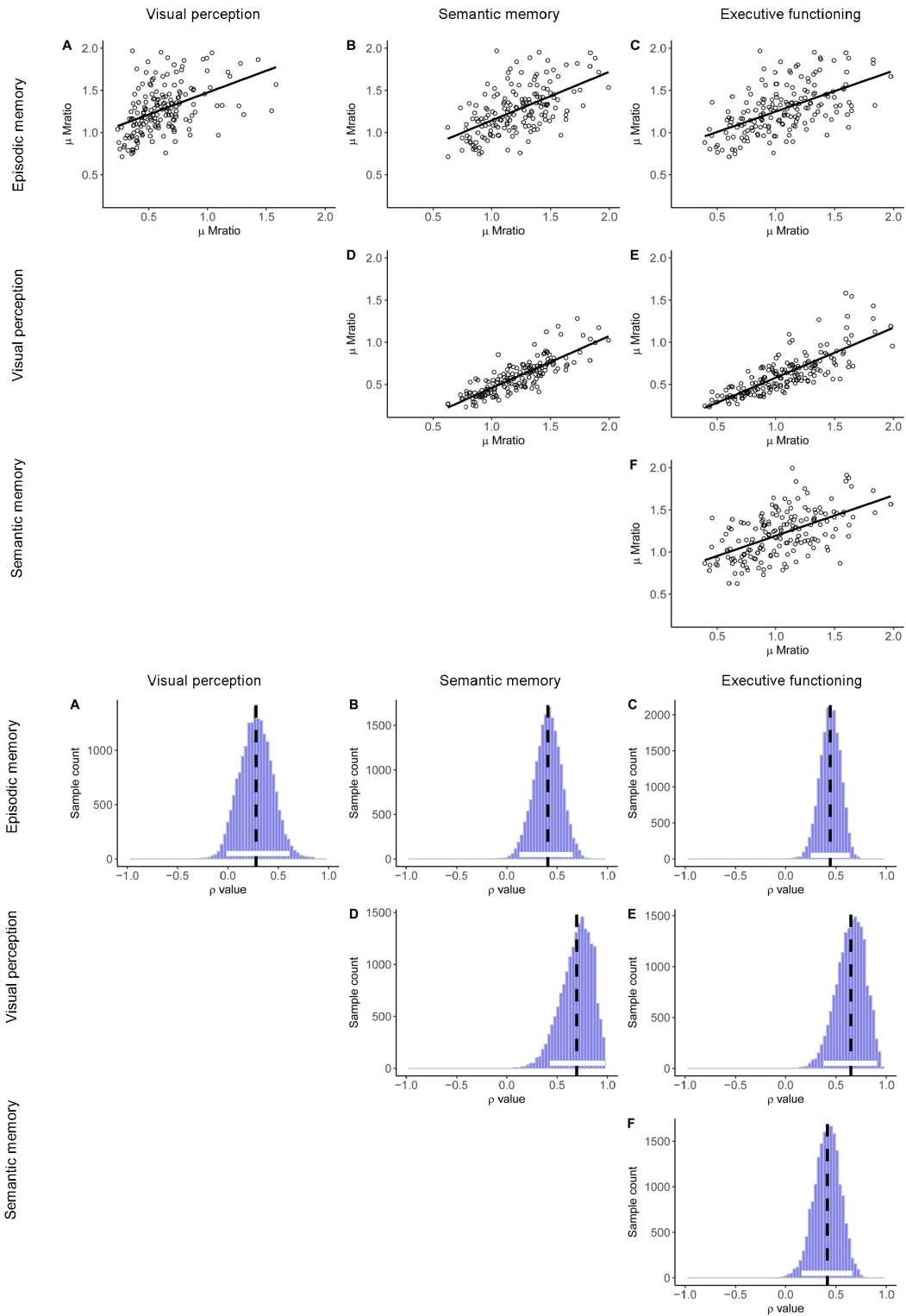


Figure 4.4: (A) Single-subject parameter estimates from the hierarchical model of meta-d'/d' and Pearson correlations between meta-d'/d' estimates across the four tasks. (B) Posterior distributions over  $\rho$  for each entry in the covariance matrix determining the correlations between meta-d'/d' across the four tasks.

Table 4.5: Pearson correlation coefficients, confidence intervals, and p values for correlations in individual meta-d'/d' between pairs of tasks. Alpha threshold is .008. N = 172.

	Correlations in individual meta-d'/d' estimates			
	Episodic memory	Visual perception	Semantic memory	Executive function
Episodic memory		$r = 0.10 [-0.05, 0.24]$ $p = .212$	$r = 0.05 [-0.10, 0.19]$ $p = .516$	$r = 0.18 [0.04, 0.33]$ $p = .015$
Visual perception			<b><math>r = 0.21 [0.06, 0.35]</math></b> <b><math>p = .006</math></b>	<b><math>r = 0.25 [0.11, 0.39]</math></b> <b><math>p &lt; .001</math></b>
Semantic memory				$r = 0.12 [-0.03, 0.26]$ $p = .106$
Executive function				

task. As the performance scale (from 50% to 100% of performance) was different from the confidence scale (0% to 100%) we transformed the confidence scale. Therefore, 0% was 50%, 10% was 55%, 20% was 60% and so on. Here, a negative score refers to underestimation of performance whereas a positive score refers to an overestimation. Bonferroni corrected single sample t-tests were used ( $\alpha = 0.05/6 = 0.008$ ); revealing that confidence tended to be higher than task performance (episodic memory:  $M = 0.015$ ,  $SD = 0.084$ ; semantic memory:  $M = 0.071$ ,  $SD = 0.089$ ; visual perception:  $M = 0.094$ ,  $SD = 0.116$ ; all  $p < .05$ ) except for the executive functioning task which was neither underestimated nor overestimated ( $M = 0.001$ ,  $SD = 0.057$ ,  $t(180) = 0.28$ ,  $p = .778$ ,  $d_z = 0.02$ ).

We also compared these difference scores across task using Bonferroni corrected paired t-tests ( $\alpha = 0.05/6 = 0.008$ ). Analyses showed that metacognitive bias was the same in the executive functioning task and the episodic memory task,  $t(180) = 2.16$ ,  $p = .032$ ,  $d_z = 0.17$ , whereas it was lower in the executive functioning task compared to the semantic memory task,  $t(180) = 10.28$ ,  $p < .001$ ,  $d_z = 0.80$ , and than visual perception task,  $t(180) = 11.77$ ,  $p < .001$ ,  $d_z = 0.92$ . The episodic memory task has also a lower bias than the semantic memory,  $t(180) = 7.83$ ,  $p < .001$ ,  $d_z = 0.61$ , and than the visual perception task,  $t(180) = 9.50$ ,  $p < .001$ ,  $d_z = 0.74$ . Finally, the bias for the visual perception task was higher than that for the semantic memory task,  $t(180) = 2.67$ ,  $p = .008$ ,  $d_z = 0.21$ . Overall, overconfidence in performance tended to be higher in tasks with lower performance.

In order to estimate domain-general and domain-specific influences on these difference scores, we computed correlations across tasks (see Table below). All scores significantly correlated with each other (all  $p < .001$  and  $r$  ranging from 0.27 to 0.42)

suggesting that the more a participant overestimates their performance in one task, the more they overestimate their performance in another task.

Table 4.6: Pearson correlation coefficients, confidence intervals, and p values for paired correlations of metacognitive bias (mean confidence - mean performance) across tasks. Alpha threshold is .008.

	Metacognitive bias			
	Episodic memory	Visual perception	Semantic memory	Executive function
Episodic memory		$r = 0.41 [0.28, 0.52]$ $p < .001$	$r = 0.37 [0.24, 0.49]$ $p < .001$	$r = 0.35 [0.21, 0.47]$ $p < .001$
Visual perception			$r = 0.42 [0.29, 0.53]$ $p < .001$	$r = 0.42 [0.29, 0.53]$ $p < .001$
Semantic memory				$r = 0.27 [0.13, 0.40]$ $p < .001$
Executive function				

Metacognitive sensitivity was estimated using AUROC2. Thus, we computed one AUROC2 for each participant and each task. The Type 2 ROC curve represents the relationship between performance (i.e., type 1 decision as correct or incorrect) and confidence. AUROC2 ranges between 0.5 and 1 where 0.5 refers to chance-level discrimination between correct and incorrect answers and 1 refers to perfect discrimination. We excluded 17 participants from these analyses because they gave 100% correct answers in one of the 4 tasks. As we were interested in the comparison of AUROC2 values across all task pairs the significance threshold for these analyses was  $\alpha = 0.05/6 = 0.008$ . Paired t-tests were conducted across tasks and showed that executive functioning AUROC2 was higher than episodic memory AUROC2,  $t(163) = 2.69$ ,  $p = .008$ ,  $d_z = 0.21$ , semantic memory AUROC2,  $t(163) = 8.42$ ,  $p < .001$ ,  $d_z = 0.66$ , and visual perception AUROC2,  $t(163) = 19.92$ ,  $p < .001$ ,  $d_z = 1.56$ . Furthermore, episodic memory AUROC2 was higher than semantic memory AUROC2,  $t(163) = 6.07$ ,  $p < .001$ ,  $d_z = 0.47$  and visual perception AUROC2,  $t(163) = 16.95$ ,  $p < .001$ ,  $d_z = 1.32$ . Finally, semantic memory AUROC2 was higher than visual perception AUROC2,  $t(163) = 12.78$ ,  $p < .001$ ,  $d_z = 1.00$ . As further discussed in the main text, these differences in metacognitive sensitivity are potentially confounded by differences in first-order performance ( $d'$ ), hence why we decided to use meta- $d'/d'$  as our primary measure of metacognitive efficiency when analyzing cross-task correlations.

For completeness, we also examined correlations in AUROC2 across tasks (see Table S7). We found a significant correlation between AUROC2 values estimated from the

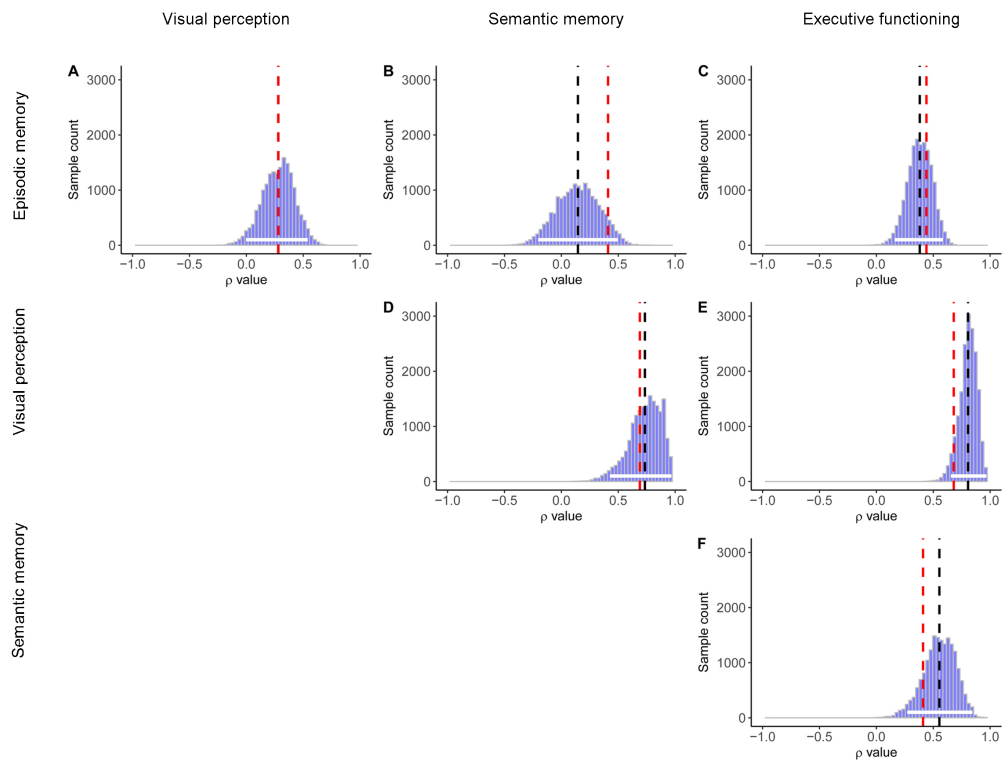


Figure 4.5: AUROC2 values per participant and per tasks.

episodic and semantic memory tasks,  $r = 0.23$ ,  $p = .003$ , but other correlations did not reach significance.

Table 4.7: Pearson correlation coefficients, confidence intervals, and p values for paired correlations of AUROC2 across tasks. Alpha threshold is .008.

	Metacognitive sensitivity (AUROC2)			
	Episodic memory	Visual perception	Semantic memory	Executive function
Episodic memory		$r = 0.14 [-0.01, 0.29]$ $p = .066$	<b><math>r = 0.23 [0.08, 0.37]</math></b> <b><math>p = .003</math></b>	$r = 0.13 [-0.03, 0.27]$ $p = .106$
Visual perception			$r = 0.13 [-0.03, 0.28]$ $p = .010$	$r = 0.07 [-0.09, 0.22]$ $p = .385$
Semantic memory				$r = 0.04 [-0.11, 0.19]$ $p = .609$
Executive function				

### Comparison between hierarchical and non hierarchical models

We carried out simulations to compare the power of hierarchical and non-hierarchical estimation procedures in recovering cross-task correlations in metacognitive efficiency.

Simulated data were generated using the variance-covariance matrix and parameters estimated from data from the current experiment. We used the `metad_sim` function from the HMeta-d toolbox to generate confidence rating data (this function simulates confidence rating data from the meta-d' model with pre-specified levels of meta-d'/d'; see Fleming, 2017, for further details).

Two types of dataset were generated: one with 40 trials per task/subject (as in the current study) and one with 400 trials per task. As in the current experiment, we simulated data for 181 participants and for 11 distinct confidence rating levels. For each participant and each task, meta-d'/d' ratios were sampled from a multivariate Gaussian distribution (using the means and covariance matrix obtained from the current study) and d' values were sampled from a normal Gaussian distribution (again using the parameters from the current study). Then, simulated confidence ratings were generated from each participant's d' and meta-d'. Decision and confidence criteria were fixed across tasks and participants. Code and generated data are available on OSF (<https://osf.io/b5ype/>). Simulation parameters were specified as follows:

- Group-d': EM = 1.84; VP = 1.19; SM = 0.92; EF = 2.58
- Type 1  $\sigma$ : EM = 0.88; VP = 0.60; SM = 0.36; EF = 0.74
- Group-Mratio: EM = 1.23; VP = 0.53; SM = 1.17; EF = 0.99
- Type 2  $\sigma$ : EM = 0.31; VP = 0.54; SM = 0.36; EF = 0.39
- $\rho$ : EM/VP = 0.28; EM/SM = 0.41; EM/EF = 0.44; SM/VP = 0.69; SM/EF = 0.68; VP/EF = 0.41
- c: 0
- c<sub>2</sub>, "S1": (-2, -1.8, -1.6, -1.4, -1.2, -1, -0.8, -0.6, -0.4, -0.2)
- c<sub>2</sub>, "S2": (0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2)

First, we estimated group-level parameters from the simulated data with 40 trials per task (mean and covariance of meta-d'/d' ratios across tasks) using the hierarchical Bayesian model (Fleming, 2017). The mean and HDIs of group-level posterior distributions over the cross-task correlation parameters are presented in the Figure 4.6. The red vertical line showed the ground-truth correlations, and the black line shows the recovered parameter. Five out of six of the recovered correlations were significantly above zero, as was the case for the actual data.

Second, we analysed the same simulated data by estimating one meta-d' / d' ratio per participant and per task using a single-subject Bayesian model (Fleming, 2017). We excluded 13 simulated participants with very low performance ( $d' < 0.10$ ) in one of

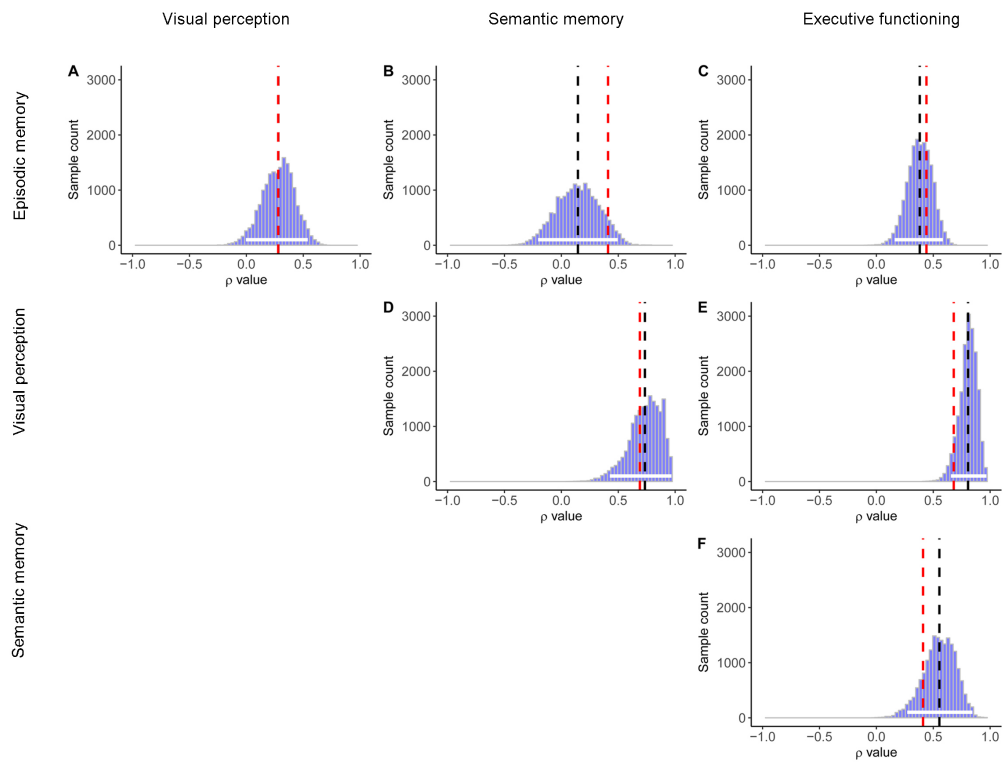


Figure 4.6: Posterior distributions over  $\rho$  for each entry in the covariance matrix determining the correlations between meta- $d'/d'$  across the four simulated tasks.  $N = 40$  trials. Red lines are ground truth correlations and black lines are mean correlations estimated by the model.

the four tasks. Then, we performed Pearson’s correlations for estimated metacognitive efficiency across tasks (see Table 4.8).

Table 4.8: Pearson correlation coefficients, confidence intervals, and p values for paired correlations of simulated meta- $d'/d'$  ratios across simulated tasks. Simulated participants with  $d'$  lower than 0.1 were excluded.  $N = 168$  subjects and  $N = 40$  trials. Alpha threshold is .008.

	Metacognitive bias			
	Episodic memory	Visual perception	Semantic memory	Executive function
Episodic memory		$r = 0.11 [-0.04, 0.26]$ $p = .147$	$r = 0.02 [-0.14, 0.17]$ $p = .838$	$r = 0.13 [-0.02, 0.28]$ $p = .083$
Visual perception			$r = 0.07 [-0.08, 0.22]$ $p = .367$	$r = 0.17 [0.02, 0.31]$ $p = .028$
Semantic memory				$r = 0.40 [0.26, 0.52]$ $p < .001$
Executive function				

When comparing correlations estimated using single-subject estimations of meta- $d'/d'$  and those obtained within the hierarchical model, the latter are closer to the ground truth correlations (see Table 4.9) except for the correlation between executive function and semantic memory. Five out of six correlations estimated within the hierarchical model were significantly above zero, compared to only 1 out of 6 correlations using the single-subject approach. Finally, correlations estimated from single-subject values tended to be lower than those estimated in the hierarchical model.

Table 4.9: Comparisons of values for “true” correlations, correlations estimated within the hierarchical model and correlations calculated from single-subject estimations.  $N = 181$  subjects and  $N = 40$  trials.

	"True" correlation	Comparisons	
		Hierarchical estimation	Single-subject estimation
EM / VP	0.28	0.28	0.11
EM / SM	0.41	0.15	0.02
EM / EF	0.44	0.38	0.13
VP / SM	0.69	0.73	0.07
VP / EF	0.68	0.81	0.17
SM / EF	0.41	0.55	0.40

We next repeated the same simulation and parameter recovery process when generating 400 simulated trials per task/subject. Results from the hierarchical estimation are presented in Figure 4.7, using the same conventions as before.

We next estimated one meta- $d' / d'$  ratio per participant and per task using a single-subject Bayesian model (Fleming, 2017). As in the 40-trial simulation, we excluded 13 simulated participants with very low performance ( $d' < 0.10$ ) in one of the four tasks were excluded. Cross-task correlations are presented in Table 4.10.

Now, in contrast to 40 trials, the comparison between correlations recovered using single-subject estimates of meta- $d'/d'$  and those recovered within the hierarchical model showed that all values were closer to the “true” correlations (see Table 4.11). Therefore, we conclude that the benefit of the hierarchical approach is likely to be particularly pronounced when low trial numbers lead to individual meta- $d'/d'$  estimates to be unreliable (see also Fleming, 2017).

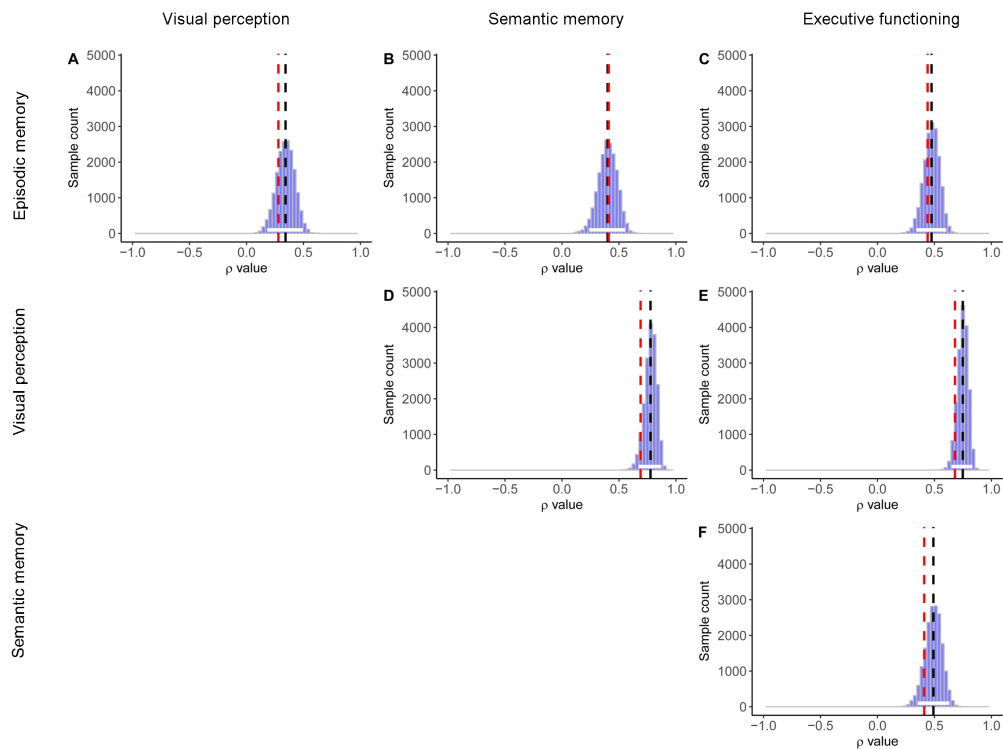


Figure 4.7: Posterior distributions over  $\rho$  for each entry in the covariance matrix determining the correlations between meta- $d'/d'$  across the four simulated tasks.  $N = 400$  trials. Red lines are ground truth correlations and black lines are mean correlations estimated by the model.

Table 4.10: Pearson correlation coefficients, confidence intervals, and  $p$  values for paired correlations of simulated meta- $d'/d'$  ratios across simulated tasks. Simulated participants with  $d'$  lower than 0.1 were excluded.  $N = 168$  subjects and  $N = 400$  trials. Alpha threshold is .008.

	Metacognitive bias			
	Episodic memory	Visual perception	Semantic memory	Executive function
Episodic memory		$r = 0.24 [0.10, 0.38]$ $p = .002$	$r = 0.30 [0.16, 0.43]$ $p < .001$	$r = 0.42 [0.29, 0.54]$ $p < .001$
Visual perception			$r = 0.59 [0.44, 0.66]$ $p < .001$	$r = 0.60 [0.50, 0.69]$ $p < .001$
Semantic memory				$r = 0.41 [0.28, 0.53]$ $p < .001$
Executive function				

## 4.4 Discussion

The present study compared RCJs across four cognitive tasks in order to quantify a potential domain-general metacognitive resource. We focused on both confidence level



Table 4.11: Comparisons of values for “true” correlations, correlations estimated within the hierarchical model and correlations calculated from single-subject estimations. N = 181 subjects and N = 400 trials.

	"True" correlation	Comparisons	
		Hierarchical estimation	Single-subject estimation
EM / VP	0.28	0.34	0.24
EM / SM	0.41	0.40	0.30
EM / EF	0.44	0.46	0.42
VP / SM	0.69	0.77	0.59
VP / EF	0.68	0.74	0.60
SM / EF	0.41	0.49	0.41

and metacognitive efficiency. Our study goes beyond previous studies by using a large sample to increase reliability, employing four distinct 2AFC tasks to avoid problems that arise when comparing different task formats, and using a hierarchical estimation of meta- $d'/d'$  (and covariance parameters) that facilitated efficient estimation of group-level correlation parameters.

We reproduced previous findings on the domain-generalty of metacognitive bias using a confidence level (e.g., [Ais et al., 2016](#)). Except for a trend between episodic memory and visual perception, we found that the tendency to report high confidence in one task is correlated with the tendency to report high confidence in another task, suggesting domain-general contributions to overall confidence level. These results are in line with judgements of confidence being biased by domain-general contextual factors such as mood (see [Ais et al., 2016](#), for influences of optimism on bias) and psychiatric symptoms (see [Rouault, Seow, et al., 2018](#), in perceptual decision-making).

Our study also allowed us to estimate the extent of across-task stability in metacognitive efficiency, by estimating the parameters of a covariance matrix governing the association between meta- $d'/d'$  values in a hierarchical framework. We found substantial shared variance in meta- $d'/d'$  across tasks, with 5 out of 6 correlation parameters deviating from zero. As the meta- $d'/d'$  measure controls for influences of task performance, this result suggests a substantial shared variance in metacognitive efficiency, and is consistent with a domain-general resource supporting metacognition. Critically, these correlations were obtained even for pairs of tasks that did not show correlations in first-order performance (i.e., for semantic memory and visual perception; for episodic

memory and executive function). This suggests that correlations in metacognitive efficiency are unlikely to be driven by covariance in task performance.

The one 95% HDI that did overlap zero, for the correlation between episodic memory and visual perception, still showed a substantial probability mass above zero, suggesting uncertainty around the proportion of shared variance, rather than an absence of correlation (HDI = [-0.03, 0.60]). Although our findings are less clear regarding these two tasks, a recent study (Lee et al., 2018) suggested a positive relationship between metacognitive sensitivity for short-term memory and visual perception when comparing 2AFC tasks using a large sample size (100 participants) and a larger number of trials (120 trials). The correlation they found was very close to the one we estimated here ( $r = 0.31$  and  $r = 0.28$ ).

Our results on shared variance in metacognitive efficiency across tasks thus suggest the involvement of a common resource in metacognitive sensitivity across domains. Nevertheless, it seems that the involvement of this common resource differed across tasks, with variation in the strength of cross-task correlations (from 0.28 to 0.69). From this perspective, general metacognition explains between 7% and 48% of the variance in cross-task meta- $d'/d'$  estimates (i.e.,  $r^2$  coefficient). As this range is large, it supports the idea that both domain-general and domain-specific processes are at play in metacognition. Recent work has indeed found common and distinct brain areas tracking confidence across recognition memory and visual perceptual metacognition tasks (Morales et al., 2018) supporting the idea that both domain-specific and domain-general processes may influence the sensitivity of metacognitive judgments. Moreover, the contribution of a putative global resource may differ according to the cognitive domain.

Comparing the highest cross-task correlation (semantic memory and visual perception) with the lowest (episodic memory and visual perception) is especially interesting. We will briefly highlight a potential reason for this difference in the use of domain-general metacognition. In the visual perception task there is an objective level of task difficulty – the difference in terms of number of dots between the two stimuli. In the semantic memory task, difficulty varies in a more subjective way: for instance, people are more likely to know a very famous actor compared to a less well-known one. However, as people share representations about general knowledge (Juslin, 1993), this variation could also create shared knowledge about task difficulty and therefore a quasi-“objective” level (or at least an inter-subjective or a consensual level; Koriat, 2008). As such, a putative domain-general metacognitive resource could reflect the ability to build meta-level representations of task difficulty to infer confidence. We would therefore

expect that the more task difficulty can be easily inferred across two tasks (e.g., from shared experimental cues, see [Barthelme & Mamassian, 2010](#)), the more metacognitive efficiency would also correlate. Conversely, if one task has an easily available difficulty signal and another task did not, we would expect a lower cross-task correlation for metacognitive efficiency: the ability to infer task difficulty is less useful for the second task. We suggest that such a lack of correlation occurs in the episodic memory task because there is less intersubject consensus regarding task difficulty, and such domain-general cues are less readily available.

Our findings are also consistent with a second-order model which proposes that a common algorithm for second-order inference may be engaged across domains ([Fleming & Daw, 2017](#)). As such, shared aspects of the state space, such as motor responses being shared across tasks ([Faivre et al., 2018](#)), can increase the prevalence of domain-general metacognition. Another driver to global metacognition would be the ability to generalize priors from one task to another, such as between two memory tasks, or two perceptual tasks ([Rouault, McWilliams, et al., 2018](#)). Although some cross-domain cues and processes influencing bias have been identified (as described above), further research should focus on identifying domain-general processes influencing metacognitive efficiency.

When analyzing our data using a non-hierarchical estimation of cross-task correlations, only two out five correlations remained significant. This is likely due to the low number of trials in this experiment and we suggest that the hierarchical model is more powerful and accurate in this context (see Methods section). To confirm this intuition, we carried out simulations to compare the power of hierarchical and non-hierarchical estimation procedures in recovering cross-task correlations in metacognitive efficiency. Simulated data were generated using the variance-covariance matrix and parameters estimated from data from the current experiment. When analyzing these data using both hierarchical and non-hierarchical estimations of cross-task correlations, we found that the hierarchical model estimations achieved a closer match to the ground truth correlations than the non-hierarchical fits for a low number of trials ( $N = 40$ ), a difference which was not seen when conducting parameter recovery simulations with a higher number of trials ( $N = 400$ ). In the present work, we opted to use a large number of participants and several cognitive tasks to study a breadth of cross-task correlations and isolate a domain-general resource. However, this approach was at the expense of having fewer trials per task. It will be important to replicate our findings with a higher number of trials in order to strengthen conclusions regarding the involvement

of a domain-general resource for metacognitive efficiency.

Finally, as in previous studies (e.g., [Morales et al., 2018](#)), we found that metacognitive efficiency was better for memory (for both episodic and semantic memory tasks in the present study) compared to visual perception. Here we consider potential explanations of this difference. One potential possibility is that the one-dimensional SDT model that underpins the modeling of metacognitive efficiency is less appropriate for memory compared to perception tasks, because memory decisions are presumably made by matching a target to a sample in a high-dimensional space. How confidence is formed in such a situation, and how the link between confidence and accuracy should be modeled relative to SDT-observer predictions, therefore remains an open question ([Van den Berg, Yoo, & Ma, 2017](#)). This may especially be the case for episodic memory decisions, which have been proposed to be influenced both by a familiarity process accommodated by classical SDT and an all-or-none recollection process (e.g., [Yonelinas, 1994](#); [Yonelinas, 2002](#)). Another possibility is that control processes exert a greater influence on confidence in the memory compared to perceptual task. In memory, metacognitive beliefs are important in regulating attempts to retrieve information: participants are more likely to engage in a search if they believe they can recall the information ([Nelson & Narens, 1994](#)). Thus, a positive feedback loop might ensue in which good metacognitive sensitivity is used to guide memory search which in turn may further increase measured metacognitive sensitivity: if one knows that she can remember the answer, she will engage a search in memory which is more likely to lead to successful remembering (compared to a situation with no active search in memory). Conversely, a belief that one cannot remember a target would lead to weaker memory search and the increased likelihood of an incorrect response. However, such a belief would be metacognitively informative for these incorrect responses (i.e., “this response should be incorrect as I did not search in memory”) therefore increasing metacognitive sensitivity. We suggest that such processes are less likely to occur in the case of visual perception which would point to a unique variance component associated with metamemory. However, as also suggested by previous work ([Morales et al., 2018](#)), both domain-specific processes and a more domain-general resource may make independent contributions to confidence formation.

To conclude, we find that contrary to previous results, both metacognitive bias (measured by confidence level) and metacognitive efficiency share common resources across domains. This observation of a domain-general signature of metacognitive efficiency was obtained after ensuring that task structures were similar across domains (2AFC

tasks), that experimental power was sufficient, and that performance-controlled measures of metacognition were employed (meta- $d'/d'$ ). The percentage of explained variance however suggests that both domain-specific and domain-general resources are involved in metacognitive efficiency which is consistent with previous neuroimaging data (Morales et al., 2018) and models of confidence formation (Fleming & Daw, 2017). It also suggests that the use of a global resource may differ according to the evaluated domain. Nevertheless, this lends support to the idea that training metacognitive efficiency in one domain can enhanced metacognitive efficiency in another domain (Carpenter et al., 2019). Such "transfer effects" on metacognition may have important implications for education and rehabilitation programs as they offer a pathway towards generalized improvements of awareness of abilities (or disabilities). Although domain-general cues have been identified for biases in confidence judgement (i.e., anchoring effects, confirmation bias), the source(s) of domain-generality in metacognitive efficiency has received less attention. Further work should focus on identifying the types of processes which influence metacognitive efficiency across domains. A second question of interest is understanding sources of variation in cross-task correlation which may indicate that some tasks are less likely than others to rely on a domain-general metacognitive resource.

## Chapter 5

# Metacognitive domain specificity in feeling-of-knowing but not retrospective confidence

Previous research has converged on the idea that metacognitive evaluations of memory dissociate between semantic and episodic memory tasks, even if the type of metacognitive judgement is held constant. This often observed difference has been the basis of much theoretical reasoning about the types of cues available when making metacognitive judgements of memory and how metacognition is altered in memory pathologies. Here we sought to revisit the difference between episodic and semantic feeling-of-knowing judgements in the light of recent research which has supported a domain general account of metacognition. One hundred participants performed classical episodic and semantic memory tasks with feeling-of-knowing judgements and confidence judgements. Using the meta- $d'$  framework, we applied a hierarchical Bayesian model to estimate metacognitive sensitivity and cross-task covariance. Results revealed a significant correlation in metacognitive efficiency ( $\text{meta-}d'/d'$ ) between the episodic memory task and the semantic memory task for confidence judgements, however no evidence was found for a cross-task correlation for FOK judgements. This supports the view that FOK judgements are based on different cues in semantic and episodic memory, whereas confidence judgements are domain general.

## 5.1 Introduction

When people fail to retrieve information from memory, they may have a feeling that they nonetheless know this information. Such FOKs are a self-evaluation of cognitive abilities and can be thought of as a metacognitive experience, akin to the TOT state (see [Brown & McNeill, 1966](#)). In experiments, FOKs refer to the predictions of future stimulus recognition when this stimulus has not been recalled. They are a special kind of judgement in that they are cued by a retrieval attempt, but pertain to future performance, and in this sense, they are prospective judgements. In the eFOK paradigm (e.g., [Schacter, 1983](#); [Souchay et al., 2000](#)), participants learn cue-target paired-words. In a subsequent recall phase, a cue word is presented and the participants are asked to recall the target word. In cases where participants are not able to remember the target, they judge whether they will be able to recognize the target amongst a set of distracter words (the FOK prediction). Finally, participants perform the recognition task. In the sFOK paradigm (e.g., [Hart, 1965](#); [Nelson, 1990](#)) instead of learning paired-words, word definitions or general knowledge questions are presented to participants (for example: 'A movement, passage, or composition marked to be performed in slow time') and they are asked to recall the answer or word referring to this definition (adagio). As in the episodic task, they make an FOK if they are not able to find the correct word, and later perform a recognition task. In the current experiment, we address the question of whether sFOK and eFOK share similar basis or resources.

The literature shows striking dissociations between the accuracy of judgements in sFOK and eFOK and especially in what is named "metacognitive sensitivity" (i.e., the ability to discriminate between correct and incorrect responses). In several neuropsychological populations, a profile of impaired eFOK despite preserved sFOK has been reported (e.g., Alzheimer's disease, [Souchay, 2007](#); schizophrenia, [Bacon, Danion, Kauffmann-Muller, & Bruant, 2001](#); [Souchay, Bacon, & Danion, 2006](#); patients with frontal lobe lesions, [Schnyer et al., 2004](#)). The same dissociation is observed in older adults (e.g., [Souchay, Moulin, Clarys, Tacconat, & Isingrini, 2007](#)). This profile is proposed to occur because sFOK and eFOK are based on different retrieval processes. [Hicks, Marsh, and Ritschel \(2002\)](#) suggest that eFOK rely on auto-noetic consciousness in order to retrieve partial information from the study phase on which to base the FOK prediction such as being able to recall what you were thinking when you first saw the pair. Such a process is exclusively involved in episodic memory. In comparison, sFOK judgements do not rely on auto-noetic consciousness but can be inferred from the retrieval of lexical or

semantic information associated to the target question such as when trying to answer the question: “word obtained by transposing the letters of another word,” one can know that the word ends with “gram”. Finally, in terms of neural regions, sFOK and eFOK seem to be based on both common and distinct regions (Elman, Klostermann, Marian, Verstaen, & Shimamura, 2012; Reggev, Zuckerman, & Maril, 2011).

This distinction is a critical one in theories of metacognition and memory, helping identify the cues used to make metacognitive evaluations of retrieval, and pointing to a domain specificity of FOK accuracy in human memory. Despite being such a robust finding in the neuropsychological and aging literatures, to our knowledge no research has examined the episodic-semantic distinction in correlational tasks in healthy participants. This is of interest for two reasons which we develop below. First, it seems that metacognitive for episodic materials only tends to occur in special populations with an episodic memory deficit (the Memory Constraint Hypothesis; Hertzog, Dunlosky, & Sinclair, 2010) limiting the generalisability of this theory. Moreover, a number of methodological issues raise questions about the nature of the episodic-semantic dissociation (for a recent review see Renault, Irish, Moscovitch, & Rugg, 2019). Secondly, research in other cognitive domains has found evidence of domain generalisable processes in metacognition (e.g., Ais et al., 2016; Lee et al., 2018; Mazancieux, Fleming, Souchay, & Moulin, 2018; McCurdy et al., 2013; Samaha & Postle, 2017), contrasting with the observed neuropsychological dissociation between semantic and episodic memory. We introduce these two areas in turn.

Firstly, several methodological issues cloud the results of studies comparing sFOK and eFOK. As shown in Table 5.1, few studies have directly compared these two types of acrsFOK and most of the dissociations related above are driven by distinct studies with different pools of participants. Moreover in the metamemory literature, metacognitive sensitivity is mainly assessed by Goodman-Kruskal gamma correlations (Kruskal & Goodman, 1954) which is a within-subject non parametric correlation between metacognitive judgements and the accuracy of the memory task. It has been shown that gamma correlations are sensitive to metacognitive bias (i.e., the tendency to be overconfident or underconfident) as well as task performance (Fleming & Lau, 2014). Task performance differences are obviously a critical issue in neuropsychological and aging populations, especially where the key theory is that eFOK judgements are less accurate due to deficits in episodic memory (e.g., in Alzheimer’s disease, Ernst et al., 2016). As such, differences in FOKs in the studies of Table 5.1 are concomitant with a deficit in either episodic recall of recognition (except in autism, see Wojcik et al., 2013). Another solution to



evaluate the potential dissociation between eFOK and sFOK is to estimate the cross-task correlations for metacognitive sensitivity in FOK. However, this has rarely been performed as sample sizes are typically small (most of the studies being with patients, see Table 5.1).

Secondly, investigating the question of whether sFOK and eFOK share common resources relates to the domain-generality of metacognition. This assumes that if metacognitive sensitivity depends on domain-general resource, sensitivity indices across two different domains will be correlated. Research to date has exclusively focused on RCJs. In contrast to FOKs, RCJs refer to a subjective evaluation of the confidence in a previous decision. As this metacognitive judgement can be performed for decisions in several cognitive domains (e.g., memory, visual perception, language), it is a good candidate to answer the question of the domain-generality of metacognition. As well as an interest in sensitivity, researchers have also examined bias, finding it is domain-general e.g., [Ais et al., 2016](#); e.g., [Baranski & Petrusic, 1995](#)) that is, people who are over confident on one task will tend to be over confident on another task. The correlation is less clear for sensitivity across tasks. However, under controlled conditions and with appropriate statistical power, it is possible to find cross-domain correlations for metacognitive sensitivity in RCJs ([Lee et al., 2018](#); [Mazancieux et al., 2018](#)). In support, a recent meta-analysis of neuroimaging studies has identified a domain-general neural network involved in RCJs for decision-making and memory tasks ([Vaccaro & Fleming, 2018](#)).

A further question arises regarding when the metacognitive evaluation is made. As metacognitive judgements can be performed prospectively and retrospectively, it is of interest whether metacognition dissociates according to this variable, especially since the above literature on domain generality tends to focus on retrospective judgements. Several findings support the idea that prospective and retrospective judgements measure distinct aspects of metacognition. They are uncorrelated (e.g., [Kelemen et al., 2000](#)), supported by separate brain regions ([Chua, Schacter, & Sperling, 2009](#)), and seem to rely on different cues and processes. In the memory field, FOKs are proposed to be based on both the access of partial information of stimuli and the familiarity of the cue ([Koriat & Levy-Sadot, 2001](#); [Metcalf et al., 1993](#)), whereas RCJs are proposed to be related to the strength of the memory trace (e.g., [Yonelinas, 1994](#)). Neuroimaging studies support the idea that common and distinct neural mechanisms underpin FOKs and RCJs in memory ([Chua et al., 2009](#)). Moreover, studies with neurological populations often exhibit a pattern of impaired sensitivity for prospective judgements and preserved sensitivity for retrospective judgements (e.g., [Pannu & Kaszniak, 2005](#)). Within the visual perception

Table 5.1: Review of studies comparing episodic FOK and semantic FOK. Using the keywords “feeling-of-knowing”, “episodic”, and “semantic” in Pubmed, we identified 16 articles. After having screened the abstracts, we selected 7 articles. Our inclusion criteria were as follows: (1) eFOK and sFOK have to be compared in the same study (2) judgements have to be related to retrospective memory (3) metacognitive sensitivity has to be measured. Two more articles were found in the references of the above studies. All the studies rely on gamma correlations for assessing metacognitive sensitivity. AD is Alzheimer’s disease. PD is Parkinson’s disease. ASD is autism spectrum disorder. Notes: eFOK = episodic Feeling of Knowing; sFOK = semantic Feeling of Knowing.

N°	Reference	Participants	Metacognitive sensitivity index	FOK sensitivity	Memory
1	<a href="#">Bacon et al. (1998)</a>	12 placebo, 12 low lorazepma dose, and 12 high lorazepma dose	Gamma	Lower eFOK for lorazepma groups. No difference for sFOK	Lower recall and recognition for both tasks in lorazepma groups
2	<a href="#">Eakin, Hertzog, and Harris (2014)</a>	50 young adults and 56 older adults	Gamma	No difference between groups in both FOKs. A trend for higher sFOK	Lower episodic recall and recognition for older adults
3	<a href="#">Morson, Moulin, and Souchay (2015)</a>	35 young adults and 16 older adults	Gamma	Lower eFOK for older adults. No difference for sFOK	Better semantic recall and recognition for older adults
4	<a href="#">Pappas et al. (1992)</a>	12 older adults and 12 patients with AD	Gamma	Lower sFOK for AD. No difference for eFOK (low for both)	Lower recall for both tasks for AD. Lower recognition for AD only in episodic memory
5	<a href="#">Perfect and Hollins (1996)</a>	46 young adults	Gamma	Lower for eFOK. No difference for sFOK	No difference in recall. Lower recognition for episodic memory
6	<a href="#">Reggev et al. (2011)</a>	23 young adults	Gamma + Hamann	No difference between eFOK and sFOK	Task performance differences not tested
7	<a href="#">Souchay and Moulin (2013)</a>	16 older adults and 16 patients with PD	Gamma	Lower sFOK and eFOK for PD	Better semantic and episodic recall and recognition for older adults
8	<a href="#">Souchay et al. (2007)</a>	20 young adults and 40 older adults	Gamma	Lower eFOK for older adults. No difference for sFOK	Lower recall and recognition for both tasks for older adults
9	<a href="#">Wojcik et al. (2013)</a>	18 children with ASD and 18 neurotypical children	Gamma	Lower eFOK for children with ASD. No difference for sFOK	No difference in task performance between groups

decision making field, RCJs are mainly influenced by reaction time and the correctness of the decision, whereas prospective judgements rely more on the judgements made for previous tasks (Fleming et al., 2016). In both literatures, prospective judgement sensitivity is lower than retrospective judgement sensitivity (Fleming et al., 2016; Perfect & Hollins, 1996). There is some variability in the procedures used and hence in the definition of 'prospective'. For instance, in perceptual tasks, the 'prospective' judgement is not a prediction before having seen the trial, but are made just before the motor response of the first-order decision, and with all the on-screen information necessary to perform the task. Even so, with such paradigms, there is still significantly lower sensitivity in prospective tasks (e.g., Siedlecka et al., 2016). In contrast to sensitivity, metacognitive bias seems consistent across judgement type (Fleming et al., 2016). FOK judgements, however, have a particular status, since they are prospective evaluations but made after a retrieval attempt, and in that way they share characteristics of both prospective and retrospective judgements.

Here, we expected to reproduce previous results regarding the domain-generalty of metacognitive efficiency for RCJs (Mazancieux et al., 2018) and to investigate the same issue in FOKs. The clearest hypothesis to formulate is that we will find cross-task (i.e. episodic/semantic) correlations for RCJs, and such a finding will help ground our studies in the work on domain general contributions to metacognition. The novel hypothesis tested here is that FOKs too will show some domain generality. Based on neuroimaging studies (Reggev et al., 2011) and on the idea that methodological issues exist in the estimation of metacognitive sensitivity in previous work on FOKs, we suggest that episodic and semantic FOKs will produce low, but significant, correlations.

As we claim for a more general view of metacognition, we here wanted to use novel methods to estimate metacognitive sensitivity stemming from the visual perception literature rather than the classic methods used in the metamemory field. Therefore, we chose to estimate metacognitive sensitivity using the meta- $d'$  framework (Maniscalco & Lau, 2012). Based on signal detection theory, this framework allows the calculation of metacognitive efficiency which is a ratio between task performance ( $d'$ ) and metacognitive performance (meta- $d'$ ). Metacognitive efficiency has the advantage of being independent from task performance and metacognitive bias which is less the case for classical correlational methods such as phi and gamma correlations (Fleming & Lau, 2014). Moreover, as we are interested in cross-task correlations, we estimated meta- $d'$  in a hierarchical Bayesian model (Fleming, 2017) which allows an estimation of metacognitive efficiency for individual participants and group level parameters (mean per task

and cross-task correlations). Preregistration of the study, raw data, model and analysis scripts are available in OSF (<https://osf.io/p5gaq/>).

## 5.2 Method

### 5.2.1 Participants

Based on a power calculation, 100 young adults ( $M_{age} = 20.50$ ,  $SD_{age} = 4.45$ ; 87% women) were included in the study recruited through an advertisement at the Grenoble-Alpes University. Because one of our aims was to reproduce the correlation for metacognitive efficiency between episodic and semantic memory in RCJs, we focused on this effect in a previous study (Mazancieux et al., 2018;  $\rho = 0.41$ , sample size estimation was made with a power of .99 at the standard .05 alpha error probability).

According to our preregistration, we excluded six participants who performed close to chance or ceiling (below 55% or above 95% correct) in the recognition of one of the two tasks. We also excluded two participants who used only one point on the metacognitive scale. Thus, we performed analyses on 92 participants.

### 5.2.2 Materials and procedure

The procedure included two tasks: an episodic task and a semantic task (Figure 5.1). In the episodic task, participants studied 40 cue-target word-pairs each presented for 1500ms. After this phase, the cue was presented, and participants attempted to recall the target with a time limit of 15 seconds. In any case (with or without a response in the recall phase), participants performed an FOK judgement. More specifically, they had to judge their confidence in recognizing the correct target between 2 words in a next phase. To do so, they used a 6-point scale ranging from 50% (I will guess the answer) to 100% (very confident in finding the answer). During the last phase, participants performed a 2AFC recognition task where the 40 cues were presented for a second time and participants had to choose between a target and a distractor by pressing either the "s" or the "l" letter. Finally, for each trial, participants had to estimate their level of confidence in their response using the same 6-point scale (ranging from "guessing" to "very confident").

In the semantic task, participants responded to 40 general-information questions. These questions were word definitions and the participants were asked to recall the word given the definition with a time limit of 15 seconds. As in the episodic task, in

either case, they judged their ability to recognize in a next phase the correct definition between two alternatives. Then in a recognition phase, the definitions were presented a second time and participants had to select the correct answer in a 2AFC task by pressing either the “s” or the “l” letter. They also estimated their level of confidence in their response using the same scale as the episodic memory task.

The stimuli were based on a previous experiment (see [Souchay et al., 2007](#)), with the exception that in order to use the meta- $d'$  framework we presented two alternatives and not four in the recognition phase. We used a feature of the [Souchay et al. \(2007\)](#) task, in that the same target words were used in the episodic and the semantic tasks in a counterbalanced fashion (see Figure 5.1). Each target has a cue for the episodic memory task and a definition for the semantic memory task. Two lists of forty targets were created such that each participant was randomly allocated to one set of targets in the episodic condition, and the other in the semantic condition. For both tasks, participants had three training trials in order to familiarise themselves with the task before the test trials. They had 15 seconds to recall the word before the performing their FOK judgement. For the recognition phase, there was no time limit. The task order was random (on the 92 non-excluded participants, 49 begun with the episodic memory task and 43 begun with the semantic memory task) for each participant and the entire procedure lasted around 45 minutes.

### 5.2.3 Data and statistical analyses

Our analyses focused on task performance, metacognitive bias, and metacognitive sensitivity. Metacognitive bias was estimated by calculating the difference between mean RCJs or FOKs and mean performance for each participant and each task. Sensitivity was estimated as metacognitive efficiency (meta- $d'/d'$ ). Meta- $d'$  is the  $d'$  that would be expected if the Type 2 ROC curve observed for a participant was his or her ideal ([Fleming, 2017](#)) using an extension of the the HMeta- $d$  toolbox (<https://github.com/metacoglab/HMeta-d>) in R software. Two models were used: one for the FOKs and one for the RCJs. Both models estimated a meta- $d'/d'$  ratio for each participant and each task as well as a group-level parameter for both the episodic and the semantic task and the cross-task correlation. To assess the significance of group-level parameters, we calculated the 95% highest density intervals (HDIs; the smallest interval containing 95% of the MCMC samples; [Kruschke, 2014](#)) on the posterior distributions and looked at its potential overlaps with zero.

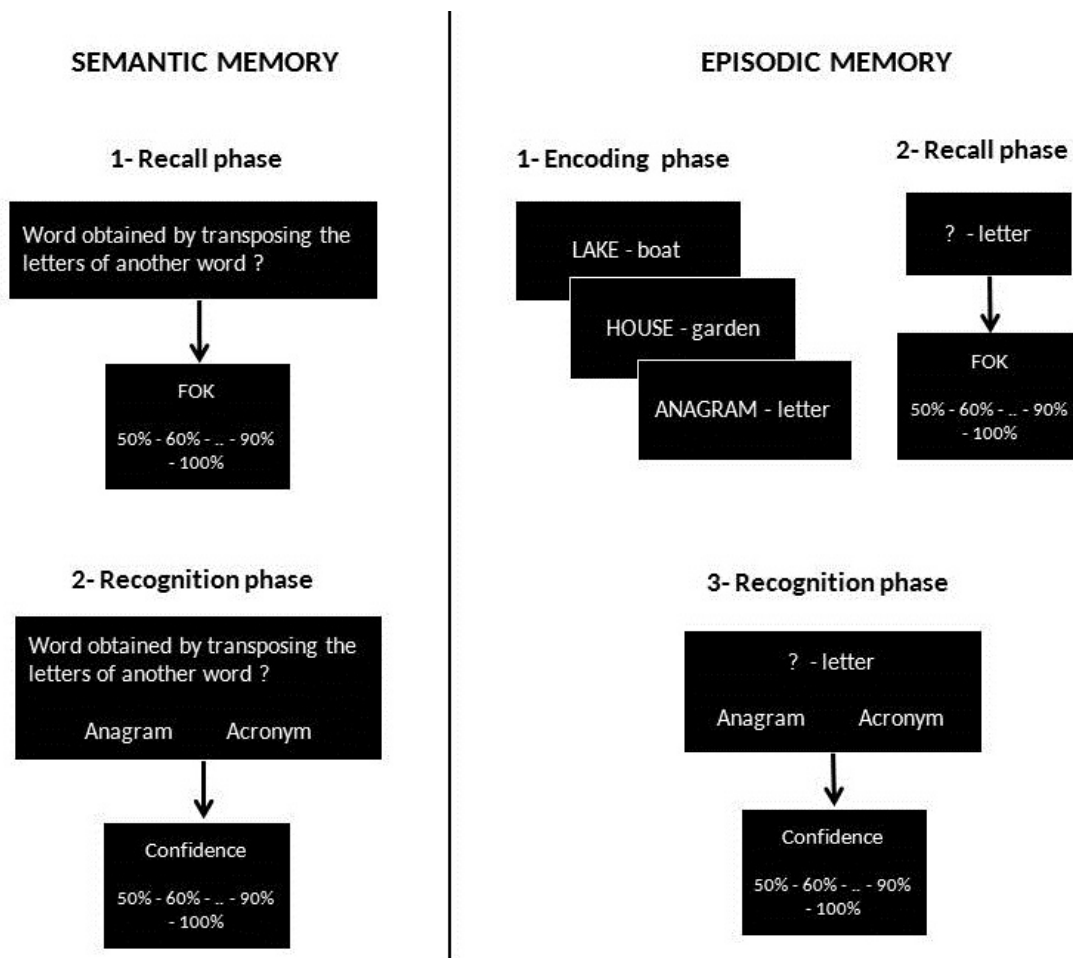


Figure 5.1: Summary of the two tasks. The semantic memory task includes 2 phases and the episodic memory includes 3 phases.

The HMeta-d toolbox uses MCMC sampling to estimate posterior distribution over model parameters using JAGS (Plummer, 2003) in R ("rjags" package). We modified the HMeta-d code to allow estimation of parameters in R using rjags. As in the HMeta-d toolbox, we discarded early samples of the posterior distributions and ran three chains in order to diagnose convergence problems. We estimated convergence of the three chains with the "potential scale reduction factor"  $\hat{R}$  (Gelman & Rubin, 1992). This approach was exactly the same as in Mazancieux et al. (2018) and was carried out according to our preregistration document. Other analyses include ANOVA, t-test, and Pearson's correlations. Outliers were detected using Leverage, RSS and Cook's distance. When necessary, Bonferroni corrections were applied.

## 5.3 Results

### 5.3.1 Type 1 performance

Task performance was estimated in two ways (Figure 5.2). First, we calculated the proportion of correct recall for episodic recall and semantic recall. A recalled item was judged as correct when the exact word was retrieved (we judged as correct those that were recalled in the plural form, even though all targets were singular). Second, we calculated a Type 1  $d'$  for performance in the recognition task. Paired t-tests showed that performance on the semantic recall ( $M = 0.27$ ;  $SD = 0.13$ ) was better than the episodic recall ( $M = 0.15$ ;  $SD = 0.09$ ),  $t(91) = 9.68$ ,  $p < .001$ ,  $d_z = 1.01$ . However, the reverse pattern was observed regarding recognition for all items: the episodic memory recognition ( $M = 1.48$ ;  $SD = 0.68$ ) was better performed than the semantic memory recognition ( $M = 1.23$ ;  $SD = 0.56$ ),  $t(91) = 3.10$ ,  $p = .003$ ,  $d_z = 0.32$ . The pattern of results was the same when comparing proportion of correct recognition only for unrecalled items,  $t(91) = 5.98$ ,  $p < .001$ ,  $d_z = 0.62$ , episodic recognition being higher ( $M = 0.73$ ;  $SD = 0.11$ ) than semantic recognition ( $M = 0.65$ ;  $SD = 0.10$ ).

We were also interested in the intersubject correlations in first-order performance across the two tasks. These analyses revealed positive correlations between episodic and semantic memory for both recall,  $r = 0.50$  [0.32; 0.64],  $p < .001$ , and recognition for all items,  $r = 0.26$  [0.05; 0.44],  $p = .013$ , suggesting that participants who performed well on one task also performed well on the other.

### 5.3.2 Metacognitive bias

Metacognitive bias (mean confidence minus mean performance) was calculated for each participant, each task, and for both FOKs and RCJs (Figure 5.3). We performed an ANOVA with judgement types and task as factors on metacognitive bias score. It revealed a main effect of task,  $t(91) = 8.99$ ,  $p < .001$ ,  $d_z = 0.94$ , with participants being more overconfident in the semantic memory task compared to the episodic memory task. We also found a main effect of judgment,  $t(91) = 18.91$ ,  $p < .001$ ,  $d_z = 1.91$ , with participants being more overconfident for RCJs compared to FOKs. Finally, the interaction,  $t(91) = 2.41$ ,  $p = .018$ ,  $d_z = 0.25$ , showed that the difference between the semantic memory task and the episodic memory task was smaller for FOKs,  $t(91) = 7.64$ ,  $p < .001$ ,  $d_z = 0.80$ , compared to RCJs,  $t(91) = 9.80$ ,  $p < .001$ ,  $d_z = 1.02$ .

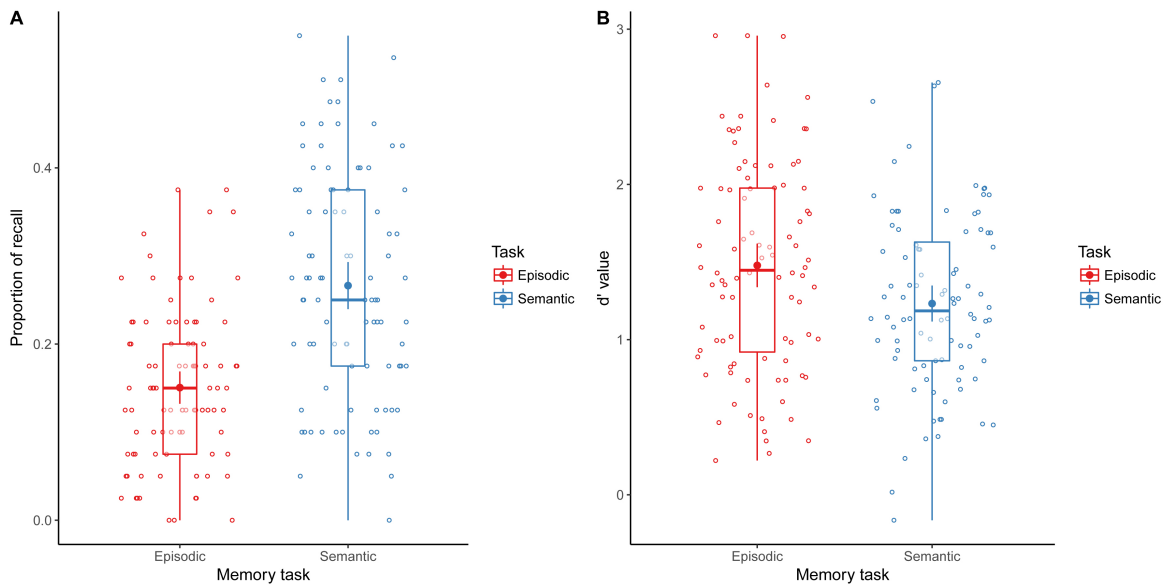


Figure 5.2: (A) Proportion of recall for the episodic and the semantic task. (B) Recognition performance ( $d'$  value) for the episodic and the semantic task.

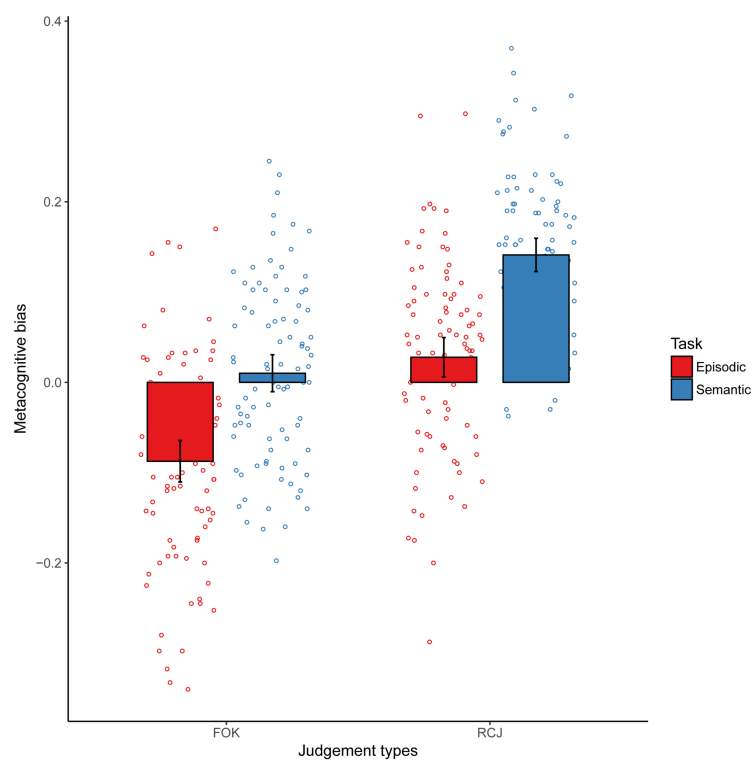


Figure 5.3: Metacognitive bias (mean judgements – mean performance) for FOKs and RCJs in the episodic memory task and semantic memory task

We were also interested in the intersubject correlations in metacognitive bias across the two tasks for both FOKs and RCJs. These analyses revealed positive correlations



between eFOK and sFOK,  $r = 0.32$  [0.13; 0.50],  $p = .002$ , and eRCJ and sRCJ,  $r = 0.36$  [0.16; 0.52],  $p < .001$ .

### 5.3.3 Metacognitive efficiency

We estimated the group meta- $d'/d'$  ratio for each task and for both FOKs and RCJs (see Figure 5.4). To test difference across the 4 distributions, we performed the difference distribution for each two-by-two comparisons (Table 5.2). According to overlaps of 95% HDIs with 0, metacognitive efficiency was better for eFOK compared to sFOK. Efficiency for eFOK was however lower than efficiency for both eRCJ and sRCJ. The same pattern was observed for sFOK. Finally, metacognitive efficiencies were the same for eRCJ and sRCJ.

Table 5.2: Means and HDIs of the posteriors of the difference between  $\mu$  Mratio distributions for each task pairing. Only the difference distribution between episodic memory and semantic memory overlaps with 0.

	Difference distributions			
	Episodic memory	Visual perception	Semantic memory	Executive function
Episodic memory		0.20 [0.06, 0.34]	0.51 [0.39, 0.64]	.52 [0.39, 0.63]
Visual perception			0.71 [0.57, 0.84]	0.71 [0.58, 0.84]
Semantic memory				0.01 [-0.11, 0.09]
Executive function				

Hierarchical models allow us to estimate correlations between metacognitive efficiency eFOK and sFOK on the one hand and eRCJ and sRCJ on the other hand (Figure 5.5). According to overlaps of 95% HDIs with 0, we found no cross-task correlation (0.22 [-0.89; 0.99]) for FOKs but a positive correlation ( $\rho = 0.47$  [0.13; 0.78]) for RCJs.

### 5.3.4 Exploratory analyses

Note that the following analyses were not preregistered.

#### Logistic regressions

Following the idea that FOK can be related to the quality of memory (the memory constraint hypothesis), we were also interested in the relationship between FOK and prior recall. As FOK occurs after a recall attempt, the idea here is to calculate the retrospective sensitivity and bias of the FOK. Therefore we performed exploratory analyses to see

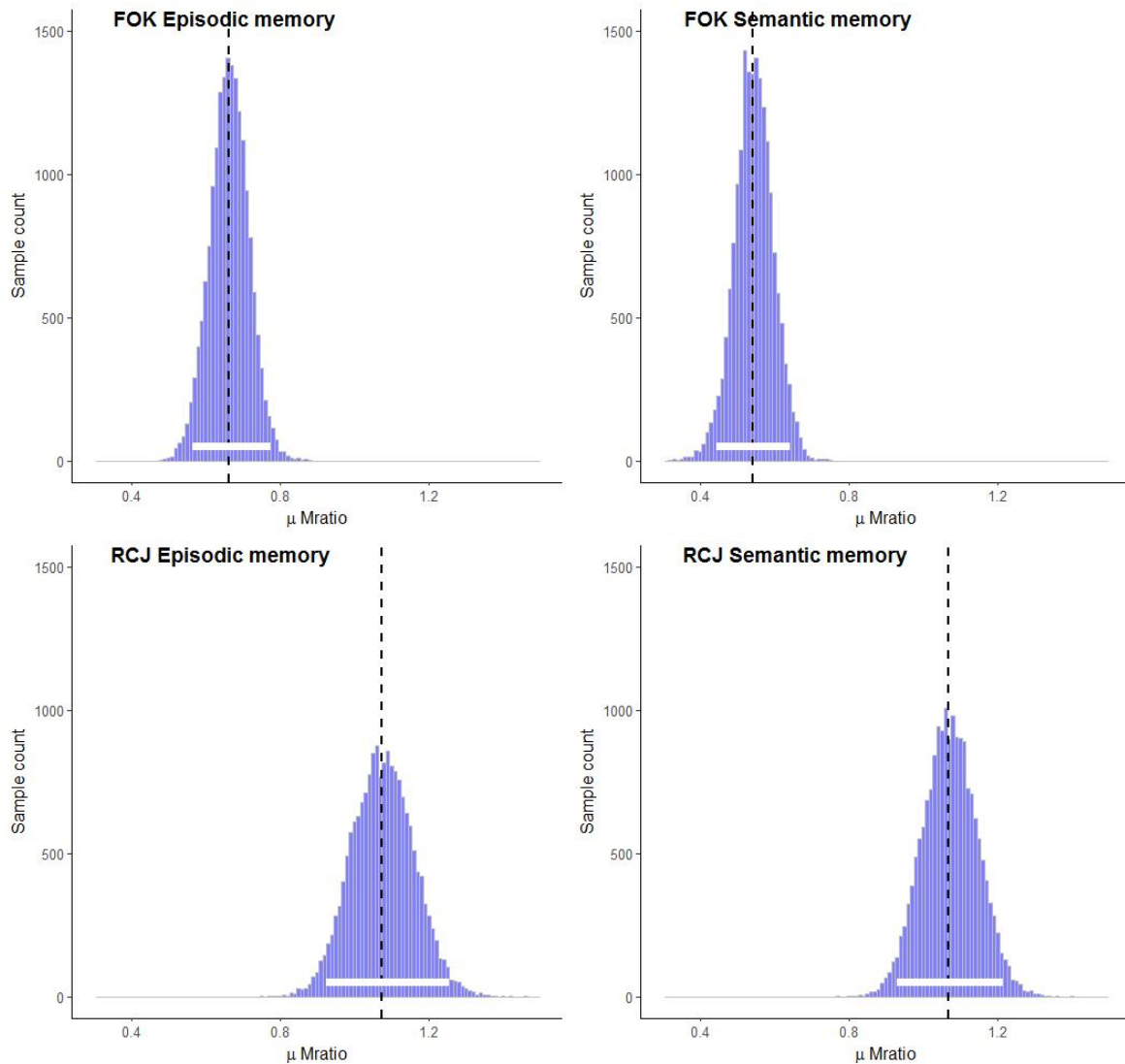


Figure 5.4: Posterior  $\mu$  Mratio (meta- $d'/d'$  ratio) distributions for FOKs and RCJs for both the episodic memory task and the semantic memory task.

if FOK judgements are related to prior recall. We quantified metacognitive sensitivity using mixed effect logistic regressions, rather than meta  $d'$ , as meta- $d'$  assumes that target and distractor distributions are Gaussian with equal variance. Although this is the case for 2AFC recognition, we suggest that this is less applicable for remembered and forgotten words, which limits the use of SDT in this context. Moreover, the estimation of the Type 2 ROC curve from the Type 1 parameters in the meta- $d'$  model as based on the average or response-specific Type 2 ROC curves (i.e., the construction of one curve for “S1” responses for “S2” responses, [Maniscalco & Lau, 2014](#)).

Consequently, we performed mixed effect logistic regressions to see if FOKs track task performance accuracy for recognition on the one hand and recall on the other

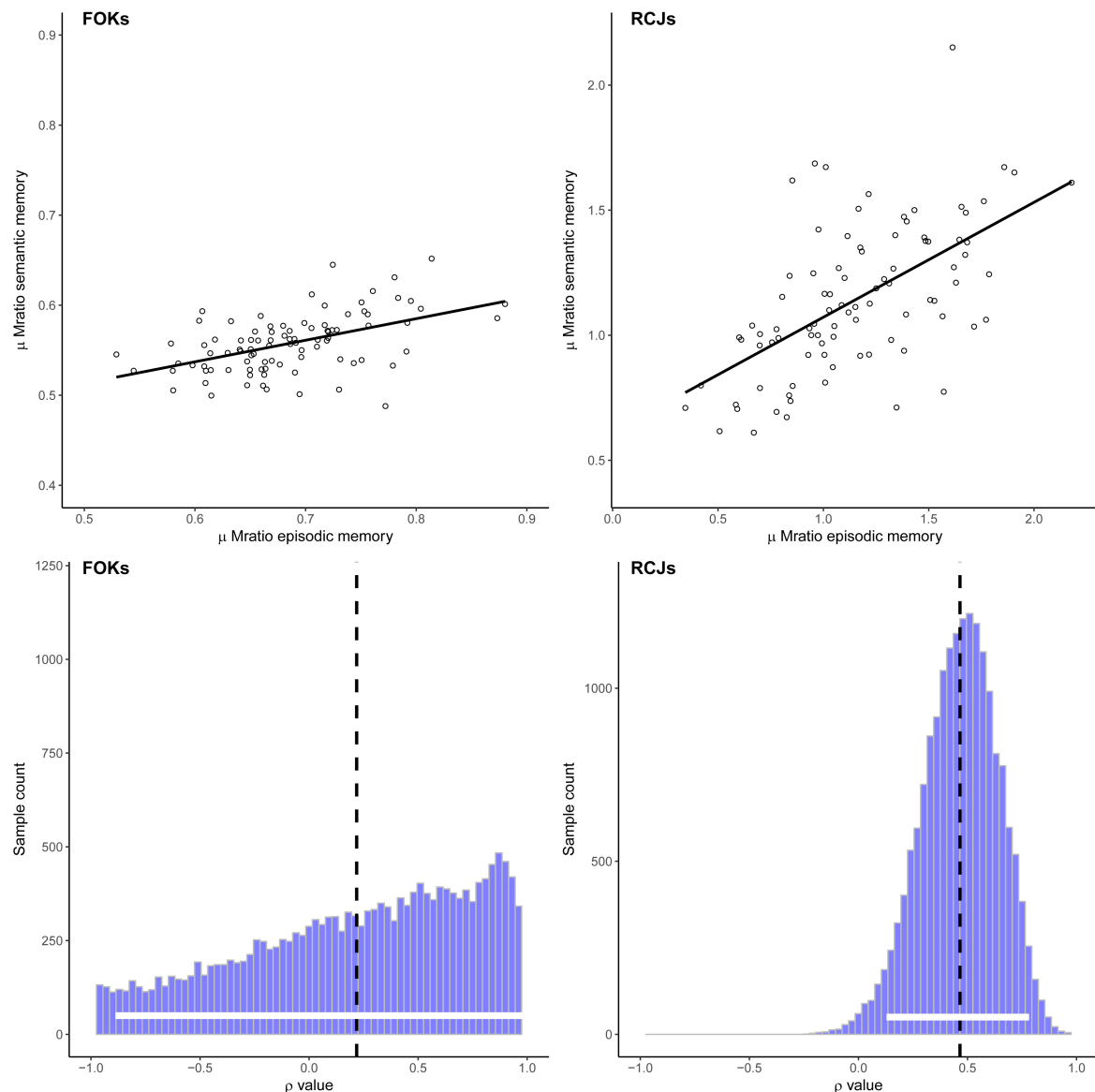


Figure 5.5: Single-subject parameter estimates from the hierarchical model of meta- $d'/d'$  and posterior distributions over  $\rho$  in the two covariance matrix determining the correlations between meta- $d'/d'$  across FOKs and RCJs. Distribution of  $\rho$  values overlaps with 0 for FOKs (0.22 [-0.89; 0.99]) which is not the case for RCJs (0.47 [0.13; 0.78]).

hand. This method has been already used as a quantification of metacognitive sensitivity because it is independent from metacognitive bias although not independent from task performance (e.g., [Faivre et al., 2018](#)). We created two models per task: one model in which FOK ratings explain first order accuracy for the recognition task and a second in which FOK ratings explain first order accuracy for the recall task. Each model includes confidence as a fixed effect and the estimation of an intercept per participant. For the episodic memory task, the models reveal that accuracy of recognition was predicted

by FOK ( $estimate = 0.37, z = 11.70, p < .001, OR = 1.45$ ) which was also the case for the accuracy of recall ( $estimate = 1.65, z = 22.26, p < .001, OR = 5.22$ ). Odds ratio comparison revealed that the effect size was higher for the model in which FOKs explain the accuracy of recall compared to the model in which FOKs explain the accuracy of recognition (odds ratio are 3.6 times larger in the recall model). For the semantic memory task, models reveal that accuracy of recognition was predicted by FOK ( $estimate = 0.23, z = 9.52, p < .001, OR = 1.30$ ) as well as the accuracy of recall ( $estimate = 0.97, z = 24.61, p < .001, OR = 2.63$ ). As for eFOK the odds ratio comparison shows that the effect size was higher for the model in which FOK explain the accuracy of recall compared to the model in which FOK explain the accuracy of recognition (odds ratio are 2.02 times larger in the recall model).

### Correlational analyses

Exploratory analyses revealed that eFOK metacognitive efficiency was correlated with episodic recall,  $r = 0.32, p = .002$ , which was not the case for eRCJ,  $r = 0.03, p = .781$  (these correlations being significantly different,  $z = 2.01, p = .036$ ). Similarly, while sFOK metacognitive efficiency was correlated with semantic recall,  $r = 0.33, p = .002$ , this correlation was significantly not different from the non significant correlation between semantic recall and sRCJ metacognitive efficiency,  $r = 0.09, p = .406$  (difference,  $z = 1.68, p = .093$ ).

## 5.4 Discussion

The present study focused on the domain-generality of metacognition in two dimensions. First, we wanted to investigate cross-task correlations in episodic and semantic memory indicating a potential general metacognitive resource as in previous works (Lee et al., 2018; Mazancieux et al., 2018; Morales et al., 2018). Second, we wanted to take into account the multifaceted nature of metacognition by comparing two types of metacognitive judgements: FOKs and RCJs. We computed metacognitive efficiency in both the episodic memory task and the semantic memory task separately for FOKs and RCJs. As our tasks contained relatively few trials due to the FOK procedure which requires recall and so cannot use too many items, we estimated metacognitive efficiency in a Bayesian manner (Fleming, 2017) and used hierarchical models to estimate cross-task correlations.

First, we reproduced previous findings (Mazancieux et al., 2018) regarding the group-level estimation of the correlation between metacognitive efficiency for eRCJ and sRCJ ( $\rho = 0.47$  [0.13; 0.78] vs.  $\rho = 0.41$  [0.14, 0.66]), although the estimation of the current correlation is less precise (with a larger HDI), which can be explained by having roughly half the number of participants in this study ( $N = 92$  vs.  $N = 181$ ). This nonetheless suggests that there is a common resource in judging our confidence across episodic and semantic memory. As RCJs are performed after the first order decision, a suitable candidate for this, especially in memory, is response fluency. That is, we suggest that participants use a common cue from the ease of answering as a basis for gauging the correctness of their response. In support, retrieval or answer fluency (shown as shorter response times) in both general knowledge tasks (Kelley & Lindsay, 1993) and episodic recognition tasks (e.g., Benjamin, Bjork, & Schwartz, 1998) is associated with higher confidence. Hence, we suggest that retrieval fluency is a diagnostic cue allowing the discrimination between correct and incorrect responses. This cue could even be used beyond the memory domain as answer fluency has been shown to influence confidence in reasoning (Thompson et al., 2013) and response time is negatively correlated with confidence in visual perception decision making (e.g., Grimaldi et al., 2015).

Second, and most importantly, we found no evidence for a cross-task correlation in eFOK and sFOK. One possible explanation is that even if FOKs imply a prospection of a future recognition task, these judgements are made after a recall test. People can use the output of the retrieval attempt to perform the FOK and therefore use recall as a heuristic for performing FOKs (Schwartz, Boduroglu, & Tekcan, 2016). In contrast with semantic recall, episodic recall involves auto-noetic consciousness as a re-experience of the remembered information. Exploratory analyses revealed that for the episodic memory task FOK metacognitive efficiency was correlated with episodic recall which was not the case for RCJ metacognitive efficiency. This is consistent with the FOK literature showing that eFOK judgments are partly based on the recollection process (e.g., Hicks et al., 2002) and especially the retrieval of the encoding context; retrieving information or details about the original encoding context (see the non-criterial recollection hypothesis of eFOK; Hertzog, Fulton, Sinclair, & Dunlosky, 2014; Isingrini et al., 2016). This is also consistent with the classical discrepancy found between impaired eFOK and preserved sFOK in older adults (e.g., Castel, Middlebrooks, & McGillivray, 2016) as recollection is the most impaired process with age (e.g., Clarys, Isingrini, & Gana, 2002), as well as several neurological diseases involving episodic memory impairment (e.g., Alzheimer's disease, Souchay, Isingrini, & Gil, 2002; Multiple Sclerosis; Beatty &

Monson, 1991; patients with frontal lobe lesions, Schnyer et al., 2004). This difference in terms of processes involved in both retrieval mechanisms can also explain why we found a better metacognitive efficiency for eFOK compared to sFOK: partial retrieved information can be used as a cue for performing the FOK judgement eFOK but not sFOK.

A classic distinction in the basis of metacognitive judgements is between experience-based metacognition and information-based metacognition (Koriat & Levy-Sadot, 1999). Experience-based metacognition is based on heuristics that are used automatically and that give rise to feeling (as epistemic feelings, see Moulin & Souchay, 2014. As an example mentioned above, the fluency heuristic has been shown to influence FOKs (e.g., the cue-familiarity heuristic, Metcalfe et al., 1993). On the contrary, information-based metacognition is based on the application of explicit beliefs or naïve theories. Although it is possible that heuristics influence metacognitive judgements across tasks (e.g., the fluency heuristic), we suggest here that beliefs are different in semantic and episodic memory tasks. Semantic memory refers to general knowledge which is shared across people who tend to have an accurate appreciation of what others know (e.g., Juslin, 1993). Thus, Koriat, 2008 has shown that metacognitive judgements correlate more with the consensual response (the one which is the most chosen by participants) compared to the actual correct response. In episodic memory tasks, such consensus is less likely to occur as retrieval abilities are closely related to the self (e.g., the self-reference effect; Symons & Johnson, 1997 and can be thought of as idiosyncratic (Klatzky, 1984) and therefore differs across people. Self-referencing has been shown to improve eFOK accuracy (Boduroglu, Pehlivanoglu, Tekcan, & Kapucu, 2015).

Going toward the idea that beliefs used for FOKs differ across tasks, Perfect and Hollins (1996) have found a between-subject correlation between FOKs and task performance for semantic memory despite no such relation for episodic memory. We also suggest that explains why sFOK are rarely impaired in patients with memory problems (e.g., Pannu & Kaszniak, 2005): information-based cues in sFOK rely more on consensus and do not necessitate autoevidence to be accurate. Semantic FOKs rely more on the inference of what people know or should know based on the activation of a network of related information, whereas eFOKs depend on self-knowledge and outputs from the retrieval attempt. That is, if you do not know anything about capitals of African countries, for instance, the sFOK will not be able to retrieve any partial information on which to make an accurate FOK judgement: the process is somewhat all-or-nothing. However, in episodic memory, the integration of autoevidence and the results of the deliberative memory search will likely yield information which is pertinent to eFOK judgements.

Whilst these inferential processes can occur as well in RCJs, we suggest here that they are less important, as RCJs rely more on the evidence driven by the given response and experience based metacognition (e.g., answer fluency) therefore explaining the cross-task correlation observed in the present study. Complementing this hypothesis, the exploratory analyses showed that FOK ratings better explains the accuracy of the recall task compared to the accuracy of the recognition. In support, recall was correlated with FOK metacognitive efficiency (albeit in both episodic and semantic tasks).

We therefore here suggest that FOKs are more based on the output of retrieval attempt during the recall (a kind of retrospective recall metacognition) rather than actual prospection of future performance. Thus, the two types of judgement are not correlated because they rely on different types of retrieval process. Future research should experimentally manipulate the semantic and episodic information available at different phases of the task, in order to test the hypothesis that eFOK and sFOK differ according to the cues used, especially because other variables not controlled in this experiment, familiarity, fluency, etc. may bring to bear on the metacognitive decision. Also, here we matched the target word in the two conditions, but it would be of interest to run a task with identical cues (rather than targets) used in the semantic and episodic conditions, such as asking someone to define a word, or retrieve something that was associated with it. It is perhaps possible that our particular pattern of recall and recognition scores have lead to the pattern of FOK correlations shown here. Of note we have a higher level of semantic recall, but a lower level of semantic recognition, and there is a large variance in recall scores (see Figure 5.2). Whilst the pattern of cross-task correlations in recall but not in FOK, suggests that recall is not a particular concern here, an interesting future experiment would be to manipulate difficulty levels in recall and recognition across the two tasks. Also, FOK sensitivity in the two tasks was correlated to recall.

Finally and regarding metacognitive bias, our results revealed that participants had a lower magnitude of judgement for FOKs compared to RCJs. This is consistent with the fact that less sensory evidence is available for prospective judgements because the task is not yet performed. Therefore participants are less confident in the FOKs reported here compared to RCJs. We also found that participants tend to give a higher metacognitive judgement for the semantic memory task resulting in more accurate FOKs but an overestimation for RCJs. In accordance with previous work in retrospective judgments (Ais et al., 2016; Mazancieux et al., 2018), we found a cross-task correlation for metacognitive bias in FOKs and RCJs. This is congruent with the fact that metacognitive bias is domain-general but also consistent across judgements types (Fleming et al., 2016).

To conclude, this study revealed a cross-task correlation for RCJs in episodic memory and semantic memory suggesting a common resource for metacognitive efficiency in these two tasks. However, no correlation was found across eFOK and sFOK. According to the dual-process view of metacognition ([Koriat & Levy-Sadot, 1999](#)), we suggest that FOKs and RCJ rely on both experience-based and information-based cues although the amounts of each process differ across judgements. We propose that these processes differ across tasks: sFOK uses inferences based on simply what we know about a subject (noesis) whereas eFOK relies more on inferences based on self-knowledge and an access to the personal past (autonoesis).



## Chapter 6

# Investigate ease-of-processing in metacognition for memory and visual perception

After having found cross-task correlations for both metacognitive bias and efficiency for retrospective judgements in the preceding chapters, the current experimental chapter focuses on the investigation of a potential domain-general resource that could explain such correlations. In this chapter, we focus on a commonly researched experience-based metacognitive cue which is fluency (as also suggested in Chapter 4 and Chapter 5). As briefly described earlier, fluency is the subjective experience of processing information easily. [Oppenheimer \(2008\)](#) suggests that “[...] fluency isn’t the process itself but, rather, information about how efficient or easy that process feels. Thus, fluency effects can be generated by nearly any form of cognitive processing, which makes it a difficult construct to pin down.” (p.238). As such, it is a good candidate for a domain-general cue ([Alter & Oppenheimer, 2009](#)).

The influence of fluency on metacognitive judgements has been mainly investigated in memory tasks (especially episodic memory) through different manipulations. Table 6.1 presents a non-exhaustive list of studies that have evaluated the influence of processing fluency on metacognitive judgements. We classified these effects into four types: perceptual fluency, answer fluency, conceptual fluency, and motor/body fluency. As we aim to have a more general view of metacognition and because most of this prior work has been focused on metamemory, relabeled some types of fluency that were too related to memory processes. Encoding fluency refers to the ease of encoding a stimulus has therefore been classified as perceptual fluency or answer fluency depending on the manipulation used in these studies. Similarly, retrieval fluency which is the ease of retrieving information from memory has been put in the broader category of answer

fluency. As presented in Table 6.1, the type of fluency that has been the most used across different cognitive domains is answer fluency. Defined as the ease in which information comes to mind, Benjamin and Bjork (1996) define three relevant characteristics for this accessibility: latency, persistence, and amount. Latency refers to the speed that information comes to mind. It is the most common subjective source of answer fluency and is often measured by reaction time. In metacognition, it has been shown that decision or answer time is negatively correlated with confidence: the shorter the reaction time, the higher the metacognitive judgements in episodic recall (Hertzog, Dunlosky, Robinson, & Kidder, 2003), reasoning (Thompson et al., 2013), problem solving (Ackerman & Zalmanov, 2012), and decision making (De Martino et al., 2013). Persistence is a characteristic similar to accessibility as it refers to a concept that comes to mind the most frequently. For instance, people are more likely to generate London compared to Dhaka when they have to think about a capital city. Finally, amount refers to the amount of information that comes to mind. The subjective ease of processing would be higher for a concept associated with multiple others compared to another with a low amount of related concepts. Although it is theoretically possible to dissociate these three characteristics, they are often confounded empirically (Benjamin & Bjork, 1996).

The other types of fluency have been exclusively studied in memory tasks and typically in episodic memory. Here, we will detail two of the most studied fluency effects on metamemory: the font-size effect and the relatedness effect. A classical manipulation of perceptual fluency on JOL is the font-size effect (Rhodes & Castel, 2008). In their original experiment, participants performed a standard JOL task on pairs of words in which half of the words are written in a small font-size (18-point font) and the other half in a large font-size (48-point font). The authors observed what they termed a ‘metacognitive illusion’: higher JOLs were given to the words written in a large font-size despite there being no difference in recall. This effect has been widely replicated (for a meta-analysis see Luna, Martín-Luengo, & Albuquerque, 2018) although the processes thought to underpin it are often debated (Mueller, Dunlosky, Tauber, & Rhodes, 2014). One major alternative explanation of the font-size effect specifies that people have beliefs regarding the size of the stimuli: they explicitly believe that words written in a large font-size will be better recalled because (e.g., because they are thought to be more important, Luna, Nogueira, & Albuquerque, 2019). Mueller et al. (2014) found that the font-size effect was also found when participants did not experience the stimuli but when they only had to predict if people would better remember words in a large font size compared to words in a small font size, supporting the belief hypothesis. A couple of their experiments did

Table 6.1: Non-exhaustive review of fluency effects on metacognitive judgements (including only studies with young adult population). \* Many others have focused on the font-size effect. We report here only the very first article (see details in the main text). \*\* Here again, we only report one of the first articles focusing on the relatedness effects (see details in the main text).

Type of fluency	Reference	Type of judgement	Domain	Effect on magnitude of judgements
Perceptual fluency	Rhodes and Castel (2008)*	JOL	Episodic memory	The font-size effect: words written in bigger have higher JOLs
	Yue, Castel, and Bjork (2013)	JOL	Episodic memory	Clear words have higher JOLs compared to blurred words
	Besken and Mulligan (2014)	JOL	Episodic memory	Auditory intact words have higher JOLs
	Rhodes and Castel (2009)	JOL	Episodic memory	Words presented in a higher loud volume have higher JOLs
Conceptual fluency	Dunlosky and Matvey (2001)**	JOL	Episodic memory	The relatedness effect: higher JOLs for highly related pair-words
	Benjamin (2003)	JOL	Episodic memory	Higher predictions for high frequency words compared to low frequency words
Answer fluency	Benjamin et al. (1998)	Prediction of free recall	Episodic memory (from trivia questions)	Negative correlation between the latency to respond to the trivia question and predictions
	Kelley and Lindsay (1993)	RCJ	Semantic memory	Primed responses easily retrieved (correct and incorrect) are judged with higher confidence
	Matvey, Dunlosky, and Guttentag (2001)	JOL	Episodic memory	Correlation with latency to generate target
	Hertzog et al. (2003)	JOL	Episodic memory	Correlation with speed of generating an image
	Koriat and Ma'ayan (2005)	JOL	Episodic memory	Shorter retrieval latency associated with higher JOLs
	Thompson et al. (2013)	FOR	Reasoning	Higher FOR when the first answer is produce quickly
	Kiani, Corthell, and Shadlen (2014)	RCJ	Perceptual decision making	Longer decision times are associated with lower confidence
	Ackerman and Zalmanov (2012)	RCJ	Problem solving	Higher confidence for solutions provided quickly
Motor/body fluency	De Martino et al. (2013)	RCJ	Value-based decision making	Lower decision time for high confidence compared to low confidence
	Susser and Mulligan (2015)	JOL	Episodic memory	Fluently written words have higher JOLs
	Susser, Panitz, Buchin, and Mulligan (2017)	JOL	Episodic memory	Words written with the dominant hand have higher JOLs
	Alban and Kelley (2013)	JOL	Episodic memory	Words studied with a lower clipboard weight have higher JOLs

not account for the fluency hypothesis and they suggested that only beliefs underpinned the font size effect. Moreover, [Hu et al. \(2015\)](#) asked both learners and observers to performed JOLs on small font size paired words and big font size paired words and found that the effect was bigger on observers compared to learners. This result put the emphasis on the importance of belief: when items are not directly experienced the effect is stronger. Nonetheless, it seems that both fluency and beliefs underpin the font-size effect ([Su et al., 2018](#)). As such, [Yang, Huang, and Shanks \(2018\)](#) provided experimental evidence that processing fluency influences JOLs as well as beliefs by directly measuring fluency using a continuous identification task.

A second fluency effect in metamemory is the relatedness effect. On the contrary to the font-size effect, this effect rather pertains to conceptual fluency as it refers to the semantic relationship between the target and the cue at encoding. [Mueller, Tauber, and Dunlosky \(2013\)](#) proposed a review of this effect: there is a substantial effect where higher JOLs are given for related words (water – lake) compared to unrelated words (cow – chair). As for the font-size effect, both processing fluency and beliefs have been used to explain such difference. When using a questionnaire, participants predict that they will be more able to recall related items compared to unrelated items without actually experiencing these stimuli. This suggests that people have beliefs about the relative memorability of such stimuli [Mueller et al. \(2013\)](#). When fluency is measured by self-paced study time (i.e., stimuli that are less studied are thought to be more fluent), related words are indeed studied less than unrelated words. However this difference did not mediate the effect on JOL magnitude ([Mueller, Dunlosky, & Tauber, 2016](#)). Even though evidence for the involvement of processing fluency is weak, other works highlighted nonetheless the influence of fluency especially when study-test experience was increased using repeated presentations ([Undorf & Erdfelder, 2015](#); Experiment 3). Overall, both the font-size effect and the relatedness effect seem to be driven by experience-based metacognition as a result of fluency and information-based metacognition in terms of beliefs.

Across these different ways of generating fluency, it seems that its effect on metacognitive judgments is consistent: a greater experience of fluency increases the magnitude of judgements. In most of cases, this effect results in a metacognitive illusion where fluency has an effect on subjective judgements without affecting the first-order performance. Hence, it modifies metacognitive bias (without affecting metacognitive sensitivity). However, for few fluency effects, the cues used to make the judgement are also pertinent for the processing goals of the task. This is the case for the relatedness effect

where related words are indeed more recalled than unrelated words. These metacognitive cues (e.g., relatedness) are named diagnostic as they allow correct predictions of performance. The effect of diagnostic cues can be explained in two different ways in the context of prospective judgements. It is possible that participants are either aware of factors that influence their task performance or that the process of making JOLs itself affects future recall (Mitchum, Kelley, & Fox, 2016; Soderstrom, Clark, Halamish, & Bjork, 2015). In favor of the latter hypothesis, Janes, Rivers, and Dunlosky (2018) have shown that the recall difference between related paired words and unrelated paired words was higher when participants had to make JOLs although recent work failed to replicate the influence of JOLs on task performance (Dougherty, Robey, & Buttaccio, 2018). It is possible that JOLs strengthen metacognitive cues and the inferential processes involved and can therefore change the way words are encoded. In this context, there is no metacognitive illusion as the cues are diagnostic. Critically, the relatedness effect is found only in prospective judgements; recall is increased only when participants perform JOLs which is not the case when they make RCJs (Dougherty, Scheck, Nelson, & Narens, 2005). This suggests that prospective and retrospective judgements are not based on the exact same processes .

In sum, across a number of different manipulations there is evidence that metamemory judgements, particularly JOLs, can be influenced by factors which are proposed to be related to fluency. These effects may even be dissociated from actual performance, generating a metacognitive illusion. The pattern of results is relatively complex, with some cues being thought to be diagnostic, and others illusory, and with some differences between types of judgement used, but it is safe to conclude that we should anticipate that fluency is a cue that is used to modulate metacognitive judgements. Whether it is a cue which can be extended to tasks other than memory tasks is something yet to be tested, apart from in the case of answer fluency, where fast reaction times have been shown to be related to the magnitude of metacognitive judgements across several first-order tasks.

In this chapter, we propose three experiments which investigate the role of fluency in metacognitive judgments. Overall, the idea of this chapter is to isolate a domain-general metacognitive cue by creating metacognitive illusion as in metamemory. Starting from the metamemory literature where fluency has been largely studied and manipulated experimentally, we suggest that fluency can be a cue influencing metacognition irrespective of the cognitive domain. Answer fluency is the type of fluency that has been the most studied as a cue for metacognition across different domains (episodic memory,

semantic memory, reasoning, problem solving, and visual perception, see Table 6.1), and so we wanted to extend this by investigating the role of both perceptual and conceptual fluency in visual perception. Our reasoning is to manipulate perceptual fluency (Experiment 1) and conceptual fluency (Experiment 2) as is done in metamemory in order to observe potential effects on metacognition in a visual perception task. We hypothesise that if fluency is a general cue available across domains then we should observe similar effects in perception tasks as is seen in memory tasks.

The majority of the studies on the effect of fluency on metamemory have focused on JOLs, a type of prospective judgement which does not have an immediate equivalent in visual perception. We thus chose to compare the impact of fluency on prospective and retrospective judgements, using a standard RCJ in the retrospective condition, and developing a prospective judgement which is made based on a cue, akin to a FOK judgement (see preceding chapter). This prospective-retrospective comparison is also of interest when adopting a more domain-general view of metacognition.

We chose to narrow our experimental design on the domains of visual perception and episodic memory for two main reasons. First and as related before, research focusing on the question of the domain-generality of metacognition predominantly compares these two domains. We therefore aim to fall within this existing literature. Second, Experiment 1 showed that these two tasks had the lowest correlations for both metacognitive bias and metacognitive efficiency. Thus, we suggest that if we find domain-general cues generate metacognitive bias across these two domains, it would be easier to generalize for the other domains as these cross-tasks correlations are stronger.

## **6.1 Experiment 1: Perceptual fluency induces by prior exposure effects on metacognition for visual perception**

### **6.1.1 Overview**

This first experiment of this section manipulates fluency and especially perceptual fluency in a comparable fashion to the font-size effect (Rhodes & Castel, 2008; Mueller et al., 2014). To do so, we chose a standard procedure used in the memory literature: a prior exposure of stimuli, something which has a long precedent in cognition and which alters judgements in numerous ways (e.g. Zajonc, 1968). The idea that fluency can be

induced by prior exposure is particularly interesting because it suggests that retrieval of information from memory implicitly or explicitly affects current metacognitive judgements (Koriat & Levy-Sadot, 1999; Koriat, 2008). Here we suggest that if participants have prior experience with task-related materials or stimuli, this information will affect both fluency and metacognitive judgements, even if it is not directly pertinent to the processing goals of the task.

Specifically, we tested the effect of pre-exposure of stimuli on metacognitive judgements for a visual perception task. First, we presented to participants different pictures during an "encoding" phase. In a second phase, participants had to perform a visual perception task (a task which did not need prior study – there was no memory component). During the perceptual task, participants reported their level of confidence. Critically, half of the stimuli were the same as those in the encoding phase, and we hypothesized that prior exposure to these stimuli would modify the metacognitive judgements, due to the greater fluency experienced when encountering the previously experienced images. Interestingly, in experiments on memory, such "cue priming" when suprathreshold, has been shown to lead to an increase in positive FOK judgements (i.e. a positive bias; Reder, 1987; Reder, 1988), leading to what is termed "spurious feelings of knowing". In contrast, subliminal priming seems to influence recall but not FOK magnitude (for an overview, see Narens, Jameson, & Lee, 1994). Our prediction was that prior exposure to the items in a study phase would lead to a higher confidence value.

We were also interested in the difference between prospective confidence judgements and retrospective confidence judgments. Because retrospective judgments are typically more accurate than prospective judgments (Fleming et al., 2016), we further hypothesized that the fluency effect would be greater for prospective judgments: more information about performance is available after task completion (such as trial difficulty and response time), and this could overcome the bias induced by the prior exposure of the cue. In comparison, before task completion people may rely more heavily on other available cues, such as our manipulation past experience with the item. As the font size effect in metamemory seems to rely on both perceptual fluency and the application of beliefs by the participants, we also measured explicit memory of the encoded stimuli at the end of the experiment. This allows us to estimate a potential relationship between the explicit remembering of having seen previously a stimuli and the metacognitive bias for seen item. We suggested here that participants who had better performance for the recognition task would also give higher confidence for previously seen items.

## 6.1.2 Method

### Participants

Participants were recruited through Prolific Academic and were paid £3 for their participation in the experiment. To estimate the sample size we focused on the intra-individual pre-exposure effect which could be related to a fluency effect as the font-size effect in metamemory judgement. Therefore, we ran a power analysis using the effect size ( $d_z = 0.36$  [0.3; 0.42]) from a recent meta-analysis (Luna et al., 2018). As recommended by Perugini, Gallucci, and Costantini (2014), we considered the lower boundary of the confidence interval. Therefore, we included ninety participants in each group and the total sample size was 180 participants. As we wanted to have around 180 participants included in our analyses and as we had several exclusion criteria, we tested 266 participants.

According to our preregistered criteria (<https://osf.io/n4t8p/>) we excluded 45 participants with performance below 55% and above 95%. Because we aimed to compare metacognition for items previously seen and not seen, participants were excluded if they had performance out side the range of our criteria in one (or both) of the two types of items. Eight other participants were excluded because they used fewer than three points on the metacognitive scale. One additional participant was excluded because he or she did not perform all the trials. Finally, as in Sherman et al. (2018), we excluded 34 participants with  $\text{meta-}d' < |0.1|$  to avoid dividing by very small values. Thus, we excluded 33.08% of our sample in total which is consistent with online studies (Chandler, Mueller, & Paolacci, 2014). Analyses were therefore conducted on 178 participants (83 in the "prospective" condition and 95 in the "retrospective" condition).

### Material and procedure

The experiment included three phases: an "encoding" phase, a visual perception task, and recognition task (Figure 6.1). All participants carried out these tasks. During the encoding phase, participants saw 26 abstract pictures for 5000ms per picture. They were not instructed to learn the items, nor told of the upcoming task. During the visual perception task, participants were represented the images again and had to select which one of the two presented small pictures was a detail of the larger picture (2AFC task). Critically, half of the larger pictures were presented during the encoding phase, such that pre-exposure to items was manipulated within-subject.



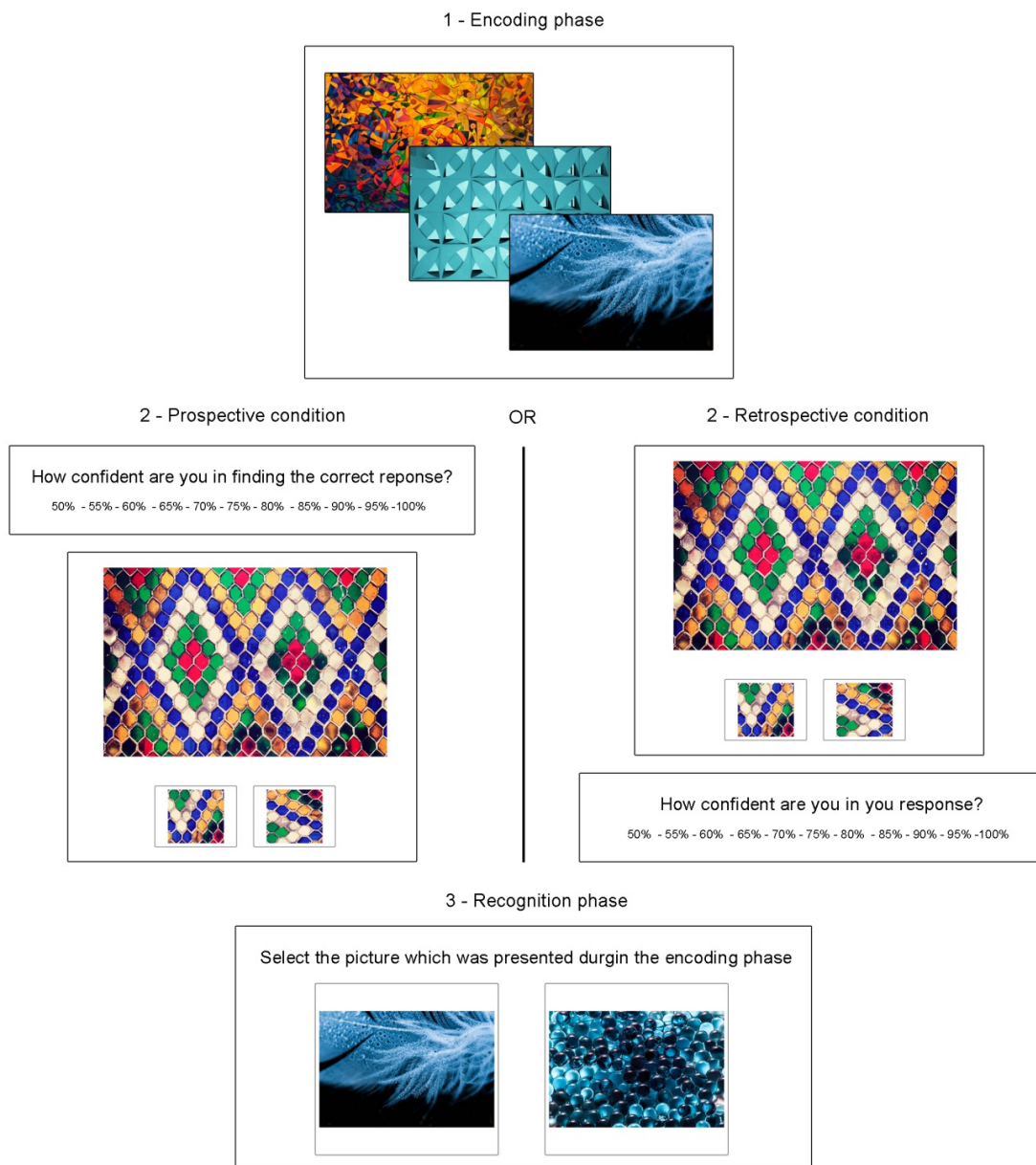


Figure 6.1: Summary of the procedure. This included 3 phase: the encoding phase, the test phase (either prospective condition or retrospective condition), and the recognition phase.

Prospective and retrospective judgements were manipulated between subject. Half of the participants were randomly assigned to the prospective judgment condition and half to the retrospective judgment condition. In the prospective condition, participants saw only the larger picture before performing the visual task. While the picture remained on the screen, they then had to judge how confident they thought they were in finding the correct detail for this picture on an 11-point-scale (ranging from 50% to 100% of confidence). The picture stayed on the screen until a confidence judgement was

made. This judgement is a prospective confidence judgement as it was made before the task response. After this subjective rating, the two alternatives appeared and the participants chose the detail which corresponded with the larger image. In the retrospective condition, participants saw first the larger picture, as above. When they were ready, they pressed a button for the two responses to appear and had to select the correct detail. Then, they had to report their level of confidence in the given response using the same 11-point-scale. To avoid ceiling performance, the big picture and the two responses only appeared together for 1500ms in both conditions. Finally, during a final recognition phase, participants had to identify the picture they had seen during the encoding phase. Specifically, they had to select the correct response between two (2AFC task).

Stimuli were abstract pictures selected in the Pexel website which is a free of rights photography database. Details from these pictures were created using the Photoshop software. Three type of detail were initially created differing in terms of size and location in the picture. Then, each picture was associated with a type of detail. Detail distractors were created from an axial symmetry of the correct detail. The experiment was programmed in the Gorilla website and was preregistered on the OSF (<https://osf.io/n4t8p/>).

### Data and statistical analyses

We examined the effects of fluency on the two facets of metacognitive accuracy: metacognitive bias and metacognitive efficiency. We estimated a meta- $d'$  per participants and per item type (previously seen or not seen) using a Bayesian framework (Fleming & Daw, 2017). A  $d'$  value was also calculated per participant and item type. We also estimated one  $m$ -distance (Sherman et al., 2018) for each participant and per item type as a quantification of metacognitive bias. All analyses were linear regressions and effect sizes were calculated as  $d_s$  and  $d_z$  according to Lakens (2013). Outliers were detected using three tests: Leverage, RSS and Cook's distance. Non-preregistered analysis for metacognitive bias was linear mixed-effect model computed using "lmerTest" and "lme4" packages (Kuznetsova, Brockhoff, & Christensen, 2017) in R software with participants and stimuli as random effects (estimation of one intercept per participant and per stimuli). Exploratory analyses included logistic mixed-effect models with participants as random effects. Because these effects are not the main focus here, we only report fixed-effects. These models included an estimation of an intercept for each participant. As there is no consensus in the calculation of effect sizes for mixed-effect models especially when models have several predictors, we decided to compute them from the t

value for each fixed-effect as for regular t-tests (Lakens, 2013) and to report odds ratios for exploratory analyses models.

### 6.1.3 Results

The following section is organized according to task performance (first-order decision), metacognitive efficiency, and metacognitive bias. The last part focuses on the final recognition task and its relationship with metacognition.

#### First-order performance

We estimated task performance using the type 1  $d'$ . As we were interested in the difference between items previously seen and not, we calculated two  $d'$  per participant: one for seen items and a second for not seen items (Figure 6.2A). Then, we compared  $d'$  values for item type (seen vs not seen) and judgement type (prospective vs retrospective). The analyses revealed a trend main effect of judgement type,  $t(176) = 1.91$ ,  $p = .058$ ,  $d_z = 0.14$ ,  $d'$  tending to be higher for items not seen ( $M = 1.37$ ,  $SD = 0.56$ ) compared to stimuli previously seen ( $M = 1.26$ ,  $SD = 0.53$ ). There were neither an effect of judgement type,  $t(176) = 0.30$ ,  $p = .767$ ,  $d_s = 0.05$ , and no interaction,  $t(176) = 1.27$ ,  $p = .207$ ,  $d_s = 0.19$ .

#### Metacognitive efficiency

To estimate metacognitive efficiency, we calculated meta- $d'/d'$  ratio for each participant and each item type (seen and not seen; (Figure 6.2B)). We compared meta- $d'/d'$  ratio for item type and judgement type. Four participants were excluded from the following analysis because they had a higher  $M$ ratio compared to the rest of the group ( $M$ ratio  $< 3.5$ ). The analyses showed a significant effect of judgement type,  $t(172) = 15.08$ ,  $p < .001$ ,  $d_s = 0.02$ , metacognitive efficiency being better for participant in the retrospective condition ( $M = 0.79$ ,  $SD = 0.70$ ) compared to the prospective condition ( $M = -0.34$ ,  $SD = 0.48$ ).  $M$ ratio in the prospective condition were significantly different from 0,  $t(172) = -6.24$ ,  $p < .001$ ,  $d_z = 0.47$ , but were negative values meaning that if any discrimination between correct and incorrect responses was possible, they did it in an opposite way (i.e., confident for future incorrect responses and non confident for future correct responses). On the contrary, participants performed correctly in the retrospective condition,  $t(172) = 15.30$ ,  $p < .001$ ,  $d_z = 1.16$ . There were neither an effect

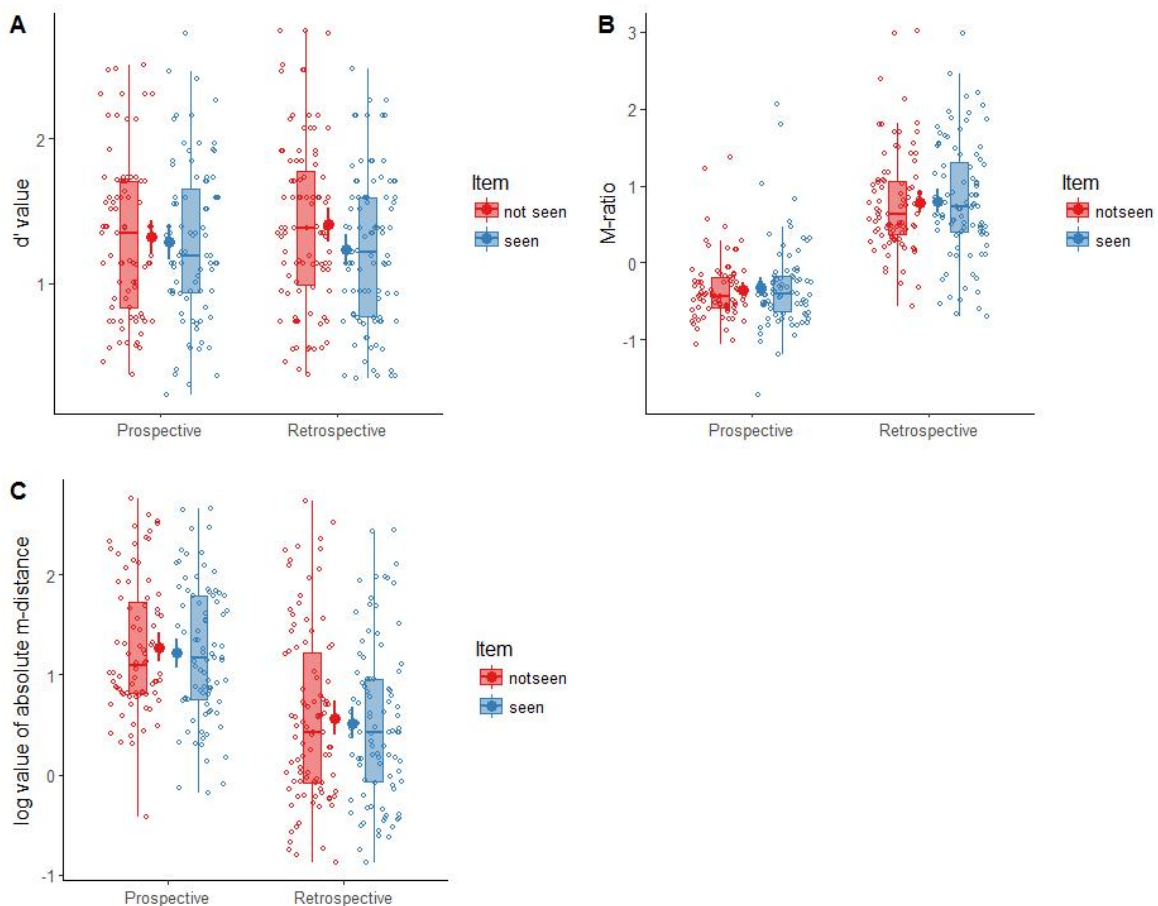


Figure 6.2: (A) Task performance measured by  $d'$  values for seen and not seen items according to the metacognitive judgement condition. (B) Mratio (meta- $d'/d'$ ) values for seen and not seen items according to the metacognitive judgement condition. Errors bars are confidence intervals. (C) Logarithmic absolute m-distance values for seen and not seen items according to the metacognitive judgement condition.

of item type,  $t(172) = 0.47$ ,  $p = .637$ ,  $d_z = 0.03$ , nor interaction,  $t(172) = -0.11$ ,  $p = .914$ ,  $d_s = 0.02$ .

### Metacognitive bias

We decided to measure metacognitive bias in two ways. As preregistered, we calculated a value of m-distance for each participant and for seen item and not seen items (Figure 6.2C). As metacognitive efficiency in the prospective condition did not differ from 0, several participants had a negative value of meta- $d'$ . Therefore, because it is the distance between confidence and response thresholds which is of interest here, we analyzed m-distance in absolute value to avoid negative scores. We compared m-distance values for

item type and judgement type. The analyses showed a significant effect of judgement type,  $t(176) = 7.32$ ,  $p < .001$ ,  $d_s = 1.11$ , participant allowing higher confidence in the retrospective condition ( $M = 0.54$ ,  $SD = 0.82$ ) compared to the prospective condition ( $M = 1.24$ ,  $SD = 0.67$ ). There were neither an effect of item type,  $t(176) = -0.92$ ,  $p = .357$ ,  $d_z = 0.07$ , nor interaction,  $t(176) = 0.03$ ,  $p = .979$ ,  $d_s < 0.01$ .

As we suggest that the perceptual fluency induce an erroneous sense of knowing, it is possible that the effect of bias might exist only for incorrect responses. As we cannot calculate m-distance for correct and incorrect answers separately because the calculation of m-distance needs meta-d', something we overlooked in our pre-registration, we conducted mixed-effect model analyses on the magnitude of confidence (Figure 6.9). Three predictors were included in the model: response accuracy (correct and incorrect), stimuli type (words and non-words), and type of judgement (prospective or retrospective). This analysis revealed main effects of the three factors. An higher confidence was given for items seen compared to item not seen  $t(6706.00) = 3.60$ ,  $p < .001$ ,  $d_z = 0.27$ , and an higher confidence was also given for retrospective judgements compared to prospective judgement,  $t(183.00) = 8.27$ ,  $p < .001$ ,  $d_s = 1.24$ . Moreover, correct responses had a higher confidence rate compared to incorrect responses,  $t(140.00) = 7.35$ ,  $p < .001$ ,  $d_z = 0.55$ . We also found a interaction between accuracy and judgements type,  $t(9024.00) = 18.59$ ,  $p < .001$ ,  $d_z = 1.39$ , showing that an higher confidence was given for correct responses in the retrospective condition,  $t(172.00) = 13.20$ ,  $p < .001$ ,  $d_z = 0.99$ , but not in the prospective condition,  $t(181.00) = 0.75$ ,  $p = .456$ ,  $d_z = 0.06$ . This result is therefore consistent with Mratio analyses. Finally, we also found a interaction between item type and judgements type,  $t(9013.00) = -2.53$ ,  $p = .011$ ,  $d_s = 0.38$ , showing that an higher confidence was given for seen item compared to not seen item for prospective judgements,  $t(8279.00) = 4.28$ ,  $p < .001$ ,  $d_z = 0.32$ , but not for retrospective judgements,  $t(8090.00) = 1.25$ ,  $p = .211$ ,  $d_z = 0.09$ .

### Relationship with recognition memory

We calculated a d' value per participant for the final recognition task. Participants performed the task as the same level in the prospective ( $M = 1.92$ ,  $SD = 0.89$ ) and retrospective condition ( $M = 1.93$ ,  $SD = 0.82$ ),  $t(176) = 0.07$ ,  $p = .942$ ,  $d_s = 0.01$ . Then, we wanted to see if the ability the retrieve the previously seen items could predict m-distance for seen item in the two judgements conditions. The model revealed that consistently with m-distance analyses that a lower confidence was given for prospective judgements,  $t(176)$

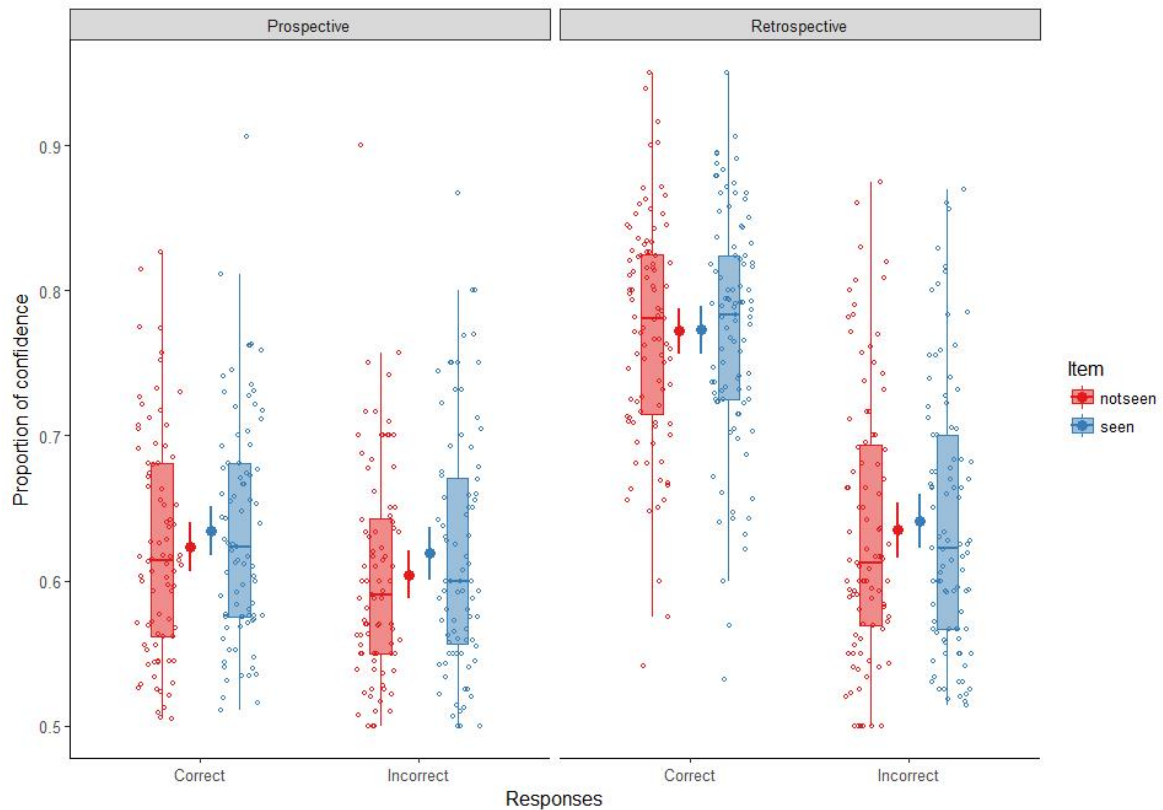


Figure 6.3: Magnitude of confidence according to accuracy of the response and metacognitive judgement condition for words and nonwords items.

$= 2.58, p = .011, d_s = 0.39$ , however there were neither an effect of memory performance,  $t(176) = -0.81, p = .417, d_z = 0.06$ , nor an interaction,  $t(176) = -0.02, p = .981, d_s < 0.01$ .

### Exploratory analyses

As metacognitive efficiency was extremely low in the prospective condition, we were interested in the possibility that participants performed prospective judgements as retrospective judgements regarding the previous trial. Thus, we performed mixed effect logistic regressions to see if metacognitive judgements track task performance accuracy. This method has been already used as a quantification of metacognitive sensitivity because it is independent from metacognitive bias although not independent from task performance (e.g., [Faivre et al., 2018](#)). However, our aim here is to compare different models where different metacognitive judgements (i.e., for trial N and trial N-1) explain the same task performance. Therefore, we suggest that variation in task performance is less problematic here. As the following analyses are performed only on prospective judgements, they only include 52 participants.

We created two models: one model in which confidence track first order accuracy for the same trial and a second in which confidence for the trial N-1 track first order accuracy. For the latest, the first trial was excluded because the matching with an N-1 judgement was not possible. However, because we want to compare the two models we also excluded the first trial for the first model to have the same number of observation per participants. Each model includes confidence as fixed effects and the estimation of an intercept per participant. Models revealed that accuracy was predicted by confidence for the same trial ( $estimate = 0.08, z = 4.85, p < .001, OR = 1.08$ ) as well as by confidence for the trial N-1 ( $estimate = 0.04, z = 2.63, p = .009, OR = 1.04$ ). However, model comparison revealed that the two models explained the same amount of variance,  $X^2 < 0.01, p > .999$ .

Finally, we focused on how the reaction time for the first-order decision can predict the magnitude of confidence for the two types of metacognitive judgements. A mixed-effect linear model revealed a main effect of reaction time showing that faster reaction times were associated with higher confidence judgements,  $t(9252.00) = -12.80, p < .001, d_z = 0.96$ . However, the interaction between reaction time and judgement type was also significant,  $t(9252.00) = -14.58, p < .001, d_z = 1.09$ . Faster reaction time predicted higher confidence judgements in the retrospective condition,  $t(9235.00) = -24.78, p < .001, d_z = 1.85$ , but this was not the case in the prospective condition,  $t(9250.00) = 1.06, p = .287, d_z = 0.08$ .

#### 6.1.4 Brief discussion

The aim of this experiment was to manipulate perceptual fluency by using prior exposure of stimuli. Our reasoning here was that if fluency is a general metacognitive cue, we should observe an effect on metacognitive judgements even if this is not directly pertinent for the first-order task, therefore creating a metacognitive illusion in visual perception (akin to the font-size effect in metamemory). We also wanted to compare prospective and retrospective judgements. Our results revealed no effect of perceptual fluency on our measure of metacognitive bias. Consistently, we found no evidence for a relationship between episodic memory and m-distance for seen items suggesting that the actual remembering of the pre-exposure to the stimuli did not affect metacognitive bias.

However we found an effect when we compared the raw magnitude of metacognitive judgements: a higher confidence was given for the previously seen items, but only in the prospective condition. This is consistent with the idea that people may use perceptual

fluency as a cue for metacognition especially when information related to the task is not sufficient (for prospective judgements). In the case of retrospective judgements, we suggest that people rely more on other cues that are more diagnostic for confidence such as answer fluency. As such, we found that higher confidence judgements were predicted by faster reaction times in the first-order decision only for retrospective judgements. It is nonetheless important to note that our first measure of interest did show any difference in metacognitive bias. A likely explanation of this could be in the quantification of metacognitive bias itself. As the estimation of  $m$ -distance needs parameters from the meta- $d'$  model (both meta- $d'$  and meta- $c_2$ ) and as the number of trial per item type is here very low ( $N = 26$ ), we suggest that estimations of  $m$ -distance were not very accurate. As the effect size for the effect of higher judgements for seen item compared to not see item was relatively small ( $d_z = 0.32$ ), it can explain that we did not detect it using  $m$ -distance.

Although participants were able to perform the first-order task in both prospective and retrospective condition, this was not the case for metacognitive efficiency.  $M$ ratio were negatively different from 0 for the prospective condition, meaning that if any discrimination between correct and incorrect responses was possible, participants did it in an opposite way (i.e., confident for future incorrect responses and non confident for future correct responses). However, mean meta- $d'$  for both seen and not seen item for prospective judgements was -0.46 suggesting that metacognitive sensitivity for these judgements was very low. Thus, participants were not able to do perform prospective confidence judgements for this visual task: the visual cue was not informative enough to create accurate judgements. This support the idea that there is a qualitative difference between metacognitive judgements for perception and memory as prospective metamemory judgements are often accurate (JOL and FOK). We will further detail this point in the general discussion of this chapter.

As exploratory analysis, we were interested in the possibility that participants performed prospective judgements as retrospective judgements regarding the previous trial as we did in Chapter 5 for FOK judgements. Prospective judgements indeed predicted the accuracy of both the current trial and the previous trial. However model for trial  $N-1$  did not predict more variance than model for trial  $N$ . This means that people use information from the current trial and previous trials to perform their prospective metacognitive judgements.

Overall, we identified a low effect for higher prospective metacognitive judgements meaning that participant use prior exposure of stimuli (and likely perceptual fluency)



as a metacognitive cue. However, as we found no effect on our measure of metacognitive bias this outcome stay to be confirmed. We suggest that this is due to the low number of trials in this experiment.

## 6.2 Experiment 2: The effect of conceptual fluency on metacognitive judgements for visual perception

### 6.2.1 Overview

Following the first experiment of this chapter, we investigated the role of fluency as a cue used to form confidence in visual perceptual decisions, in line with analogous models of metamemory judgments (e.g. [Koriat, 1997](#)). As the results of the previous experiment were ambiguous regarding the influence of perceptual fluency on metacognitive judgements, we aim here to reproduce these findings using a procedure with more trials. We were also interested in the fact that “fluency exerts the same influence on judgements independently of how it is generated” (p. 220, [Alter & Oppenheimer, 2009](#)). Thus, we here focused on conceptual fluency. As reviewed preciously, in episodic memory, manipulations of conceptual fluency have been shown to affect metacognitive judgements (e.g., the relatedness effect on JOL, [Dunlosky & Matvey, 2001](#); [Mueller et al., 2013](#); [Undorf & Erdfelder, 2015](#)). Similarly, participants give higher metacognitive judgements for high frequency words than low frequency words ([Benjamin, 2003](#); [Jia et al., 2016](#)). Here, we proposed to participants the same visual perception task as in the first experiment albeit using different stimuli. These stimuli were visual patterns for which the global shape was three letters. For half of the stimuli these three letters were a word and they were a non-word for the other half. Akin to the frequency effect in metamemory, we suggest that if fluency is a common metacognitive cue across domains, metacognitive judgements of participants will be modified by the conceptual fluency arising from the words (compared to the non-words) even if it is not directly pertinent to the processing goals of the task. Regarding the frequency effect, if this cue is a diagnostic cue for memory (participants indeed recall more frequent words compared to non-frequent words), using this cue would create a metacognitive illusion for visual perception tasks. As previously, we were also interested in the difference between prospective confidence judgements and retrospective confidence judgments and we further hypothesized that the fluency effect would be greater for prospective judgments in keeping with the results of Experiment 1.

## 6.2.2 Method

### Participants

Participants were recruited at Grenoble Alpes University. To estimate the sample size we focused on the within-subject difference between words and non-words stimuli. We based our power calculation on the magnitude of the word frequency effect on metamemory judgements. Using the first experiment of [Jia et al. \(2016\)](#), we calculated an effect size and confidence interval for Judgement-Of-Learning magnitude difference between low frequency words and high frequency words ( $d_z = 0.80$  [0.40; 1.23]). As recommended by [Perugini et al. \(2014\)](#), we considered the lower boundary of the confidence interval. Using a power of .80 we determined a sample size of 51 participants. Because our design included a between-subject comparison we included 102 participants in total. Because we have several exclusion criteria, we therefore tested 116 participants.

According to our preregistered criteria we excluded 4 participants with performance below 55% and above 95% in one of the two types of items (i.e., words or non words). One other participant was excluded because he or she used fewer than 3 points on the confidence scale. Finally, as previously, we excluded 13 participants with  $\text{meta-}d' < |0.1|$  to avoid dividing by very small values. Thus, we excluded 15.52% of our sample and analyses were conducted on 98 participants (44 in the "prospective" condition and 54 in the "retrospective" condition).

### Material and procedure

Participants performed the task individually in a quiet experimental box. As previously, the experiment was a visual perception task in which participants had to select which one of the two presented small pictures is a detail of a larger picture (2AFC task). Stimuli were visual pictures forming three letters (see Figure 6.1). To manipulate conceptual fluency, half of these larger pictures were French words and the other half were non-words. Each stimuli condition contained 40 trials. Words were selected from the Lexique database ([New et al., 2004](#)) and were the most frequent three-letters words in the database. Non-word creation was performed using the same letters as the real words and randomly mixing them. Each word and non-word was associated with visual patterns which were abstract pictures (a part was the same as those used in the previous experiment). To control for the effect of the visual pattern, the same picture was associated with both a word and a non-word, and their presentation was counterbalanced

across participants. Details from the picture were created using the same procedure as the previous experiment.

As previously, half of the participants were randomly assigned to a prospective judgment condition and half to a retrospective judgment condition. Therefore, participants had to either estimate prospectively their confidence in finding the correct detail for each trial or to estimate level of confidence in their response retrospectively. The metacognitive scale was also the same. In both conditions, each trial began by the presentation of the larger picture for 4000ms. Then, participants in the prospective condition had to give their metacognitive judgement. After that, all participants had to perform the visual task and select the correct detail between two answers. During the response selection, the big picture and the two responses appeared together for 2000ms. This time was increased compared to the first experiment (1500ms) as task performance was lower than when we pretested the task. After this decision, participants in the retrospective condition had to judge their level of confidence in their answer (see Figure 6.4 for a summary of the procedure and stimuli examples). The experiment was programmed using E-prime software and was preregistered on the OSF (<https://osf.io/mp6qk/>).

### Data and statistical analyses

As in Experiment 1, we examined the effects of fluency on metacognitive bias and metacognitive efficiency. We also estimated a meta- $d'$  per participants and per item type (previously seen or not seen) using a Bayesian framework (Fleming & Daw, 2017) and one  $m$ -distance (Sherman et al., 2018) per participants and per item type as a quantification of metacognitive bias. All analyses were linear regressions and effect sizes were calculating as  $d_s$  and  $d_z$  according to Lakens (2013). Exploratory analyses included logistic mixed-effect models with participants as random effects. Effect sizes for mixed-effect models were here odds ratios.

### 6.2.3 Results

As with the previous experiment, the result section is organized according to task performance (first-order decision), metacognition efficiency, and metacognitive bias. The last part focuses on the final recognition task and its relationship with metacognition.

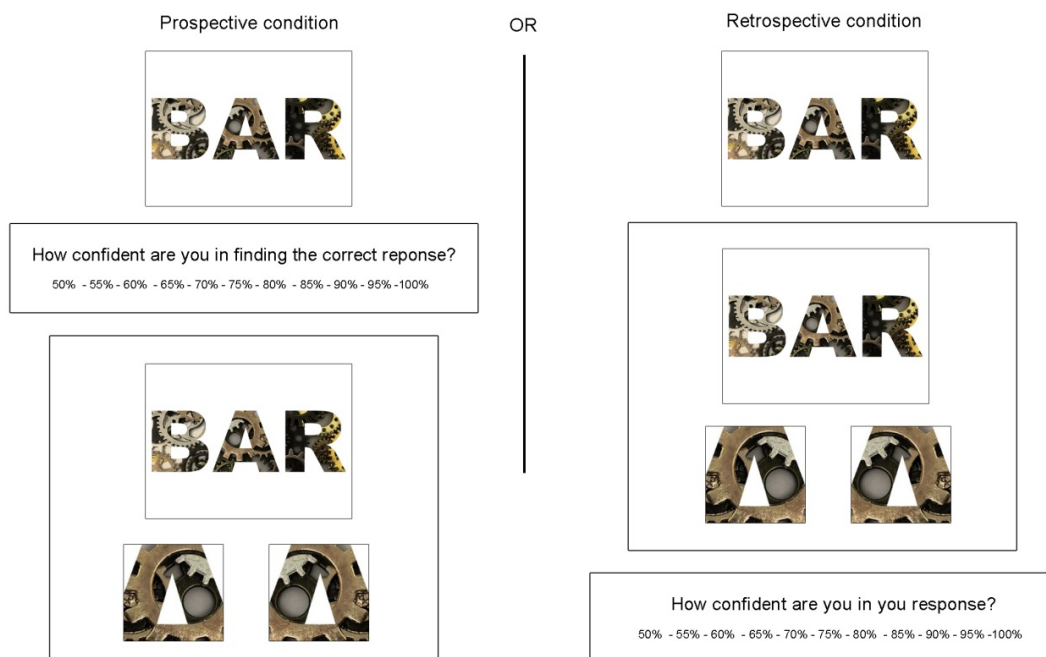


Figure 6.4: Summary of the procedure. Participants were either in the prospective condition or in the retrospective condition. For each participant, half of the stimuli were words and the other half were non words.

### First-order performance

We estimated task performance using one  $d'$  per stimuli condition (words and non-words; see Figure 6.5A). We compared  $d'$  values for item type (words vs non-words) and judgement type (prospective vs retrospective). The analyses revealed no main effect of either judgement type,  $t(96) = 0.80$ ,  $p = .429$ ,  $d_s = 0.14$ , or type of judgement,  $t(96) = 1.35$ ,  $p = .179$ ,  $d_z = 0.14$ . Moreover, there was no interaction between the two factors,  $t(96) = 0.19$ ,  $p = .851$ ,  $d_s = 0.04$ . Our manipulated variables had no effect on performance.

### Metacognitive efficiency

To estimate metacognitive efficiency, we calculated meta- $d'/d'$  ratio for each participant and each stimuli type (words and non-words; Figure 6.5B). We compared meta- $d'/d'$  ratio for item type and judgement type. We excluded two participants from these analyses who had a very high M-ratio in one of the stimuli condition (7.11 and 4.31 respectively). The analyses showed a significant effect of judgement type,  $t(94) = 13.55$ ,  $p < .001$ ,  $d_s$

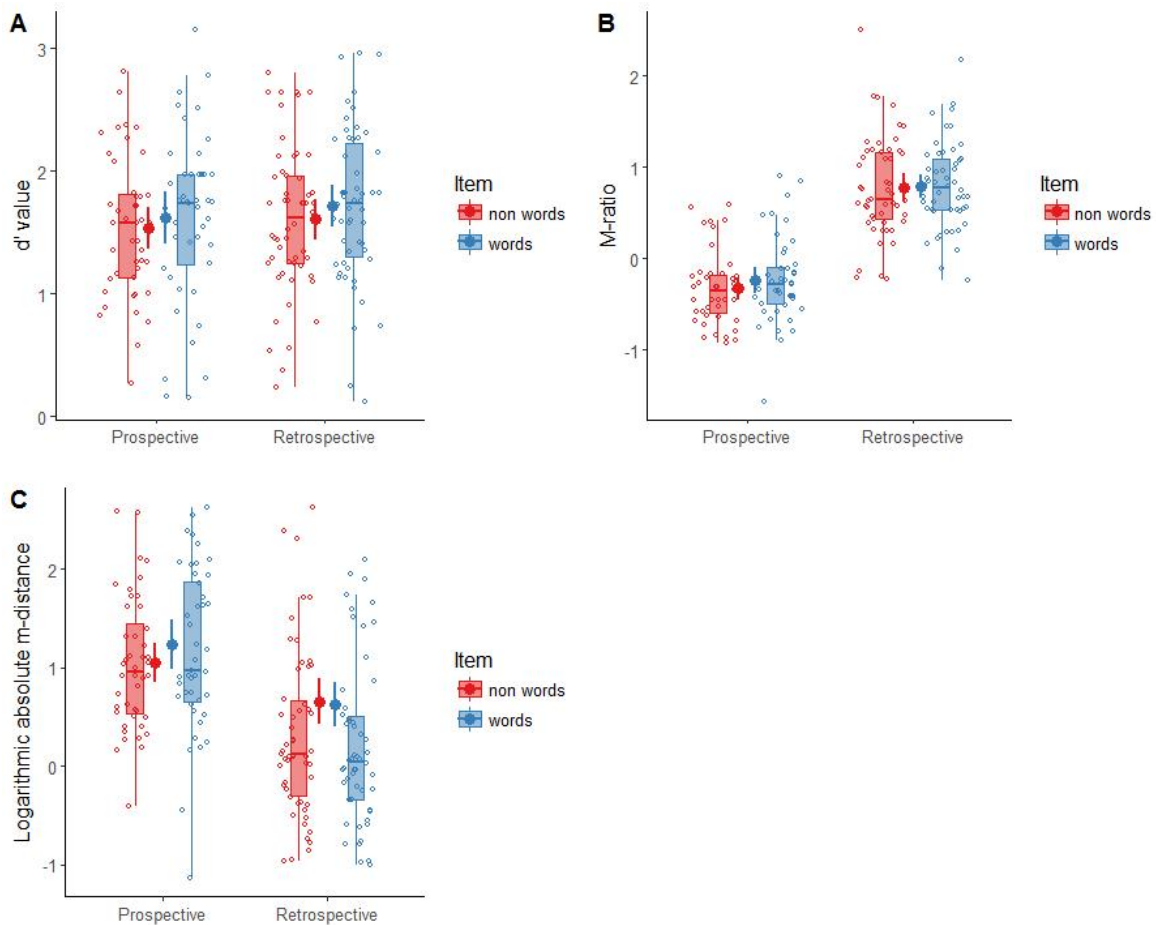


Figure 6.5: (A) Task performance measured by  $d'$  values for words and non-words items according to the metacognitive judgement condition. (B) Mratio (meta- $d'/d'$ ) values for words and non-words items according to the metacognitive judgement condition. Errors bars are confidence intervals. (C) Logarithmic absolute m-distance values for words and non-words items according to the metacognitive judgement condition. Errors bars are confidence intervals.

= 2.75, metacognitive efficiency being better for participants in the retrospective condition ( $M = 0.80$ ,  $SD = 0.77$ ) compared to the prospective condition ( $M = -0.49$ ,  $SD = 0.58$ ), as with Experiment 1. Mratios in the prospective condition were significantly different from 0,  $t(94) = -4.92$ ,  $p < .001$ ,  $d_z = 0.50$ , but were negative meaning that if any discrimination between correct and incorrect responses was possible, they did it in the opposite direction (i.e., confident for future incorrect responses and non confident for future correct responses). On the contrary, participants performed correctly in the retrospective condition,  $t(94) = 14.66$ ,  $p < .001$ ,  $d_z = 1.50$ . There were neither an effect of item type,  $t(94) = 0.93$ ,  $p = .357$ ,  $d_z = 0.09$ , nor interaction,  $t(94) = -0.57$ ,  $p = .566$ ,  $d_s = 0.12$ .

### Metacognitive bias

A value of *m*-distance was calculated for each participant and for words and not non-words (Figure 6.5C). As metacognitive efficiency in the prospective condition did not differ from 0, we analyzed *m*-distance in absolute values as we did in the previous experiment. We compared *m*-distance values for item type and judgement type. We used a logarithmic transformation to fit to the normality assumptions of linear models. The analysis showed a significant effect of judgement type,  $t(96) = 5.82, p < .001, d_s = 1.18$ , participants having a higher tendency to report higher confidence in the retrospective condition ( $M = 0.26, SD = 0.77$ ) compared to the prospective condition ( $M = 1.09, SD = 0.75$ ). There was neither an effect of item type,  $t(96) = 0.22, p = .829, d_z = 0.02$ , nor an interaction,  $t(96) = 1.54, p = .127, d_s = 0.31$ .

As preregistered, and as we carried out in Experiment 1, we were also interested in the comparison of the magnitude of confidence for response accuracy (correct and incorrect), stimuli type (words and non-words), and type of judgement (prospective or retrospective). We conducted mixed-effect model analyses showing main effects of response accuracy and type of judgement (Figure X). Higher confidence was given for correct responses compared to incorrect responses  $t(7833.00) = 15.04, p < .001, d_s = 3.04$ , and higher confidence was also given for retrospective judgements compared to prospective judgements,  $t(100.00) = 3.61, p < .001, d_z = 0.36$ . Both of these are in support of the metacognitive efficiency and *m*-distance measures above. However, in this experiment, no difference was found between words and non-words,  $t(186.00) = 0.29, p = .770, d_z = 0.03$ : our critical manipulation of fluency. We did find an interaction between type of judgement and response accuracy,  $t(7721.00) = 20.20, p < .001, d_z = 4.09$ : higher confidence was given for correct responses compared to incorrect responses in the retrospective condition,  $t(4881.00) = 25.89, p < .001, d_z = 2.60$ , and the reverse pattern was found in the prospective condition  $t(5285.00) = -3.03, p = .002, d_z = -0.30$ ; again consistent with the *M*ratio analysis.

### Exploratory analyses

As in Experiment 1, we were interested in the possibility that participants performed prospective judgements as retrospective judgements regarding the previous trial. We performed mixed effect logistic regressions to see if metacognitive judgements track task performance accuracy. Here again, we also only included the participants in the prospective condition ( $N = 44$ ). The analyses revealed that accuracy was not predicted

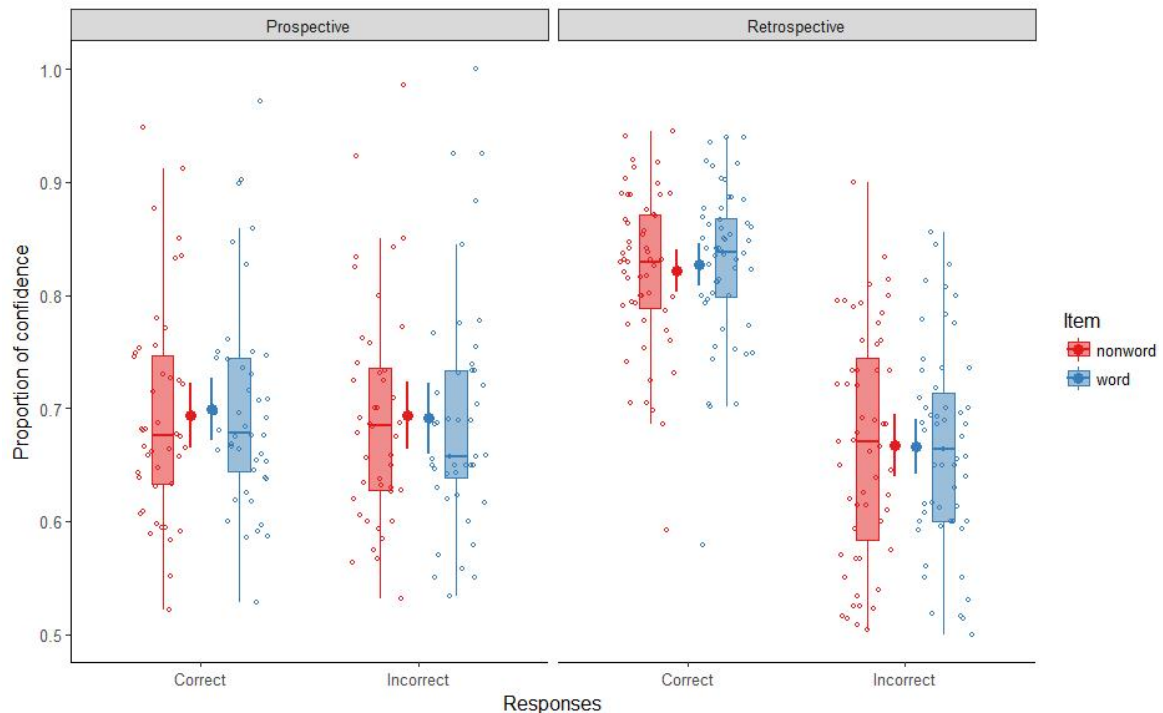


Figure 6.6: Magnitude of confidence according to accuracy of the response and metacognitive judgement condition for seen and not seen items.

by confidence for the same trial ( $estimate = 0.03$ ,  $z = 2.89$ ,  $p = .093$ ,  $OR = 1.03$ ) although it was the case by confidence for the trial N-1 ( $estimate = 0.12$ ,  $z = 5.50$ ,  $p < .001$ ,  $OR = 1.12$ ). Model comparison revealed that the model where accuracy is predicted by the judgements of the same trial explains less variance compared to the model where accuracy is predicted by the judgements of the trial before,  $X^2 = 28.72$ ,  $p < .001$ .

Finally, as with the previous experiment, we focused on how the reaction time for the first-order decision can predict the magnitude of confidence for the two types of metacognitive judgements. A mixed-effect linear model revealed a main effect of reaction time showing that faster reaction times were associated with higher confidence judgements,  $t(7890.00) = -19.05$ ,  $p < .001$ ,  $d_z = 1.91$ . However, the interaction between reaction time and judgement type was also significant,  $t(7890.00) = -21.93$ ,  $p < .001$ ,  $d_z = 2.20$ . Faster reaction time predicted higher confidence judgements in the retrospective condition,  $t(7899.00) = -32.89$ ,  $p < .001$ ,  $d_z = 3.31$ , but this was not the case in the prospective condition,  $t(7882.00) = 1.83$ ,  $p = .066$ ,  $d_z = 0.18$ , repeating the same pattern as in Experiment 1.

## 6.2.4 Brief discussion

The second experiment of this chapter focused on conceptual fluency as a potential metacognitive cue in a visual perception task. In line with our first experiment, we suggested that if fluency is a general metacognitive cue, we should observe an effect on metacognitive judgements even if this is not directly pertinent for the processing goal of the first-order task. This experiment broadly replicated the findings of Experiment 1 yielding a series of results in accordance with our predictions: notably, retrospective judgements are more accurate than prospective judgements, faster first order reaction times are associated with higher levels of confidence, and prospective judgements is more related to the performance of the trial before than the trial in hand<sup>1</sup>. However, our critical analysis did not reveal an effect of the word/non word manipulation on metacognitive bias either with m-distance or magnitude of judgements. In short, our manipulation of conceptual fluency did not influence second order measures. Before discussing these results more fully, we concentrate on Experiment 3 where we aimed to show that our manipulation of conceptual fluency was at least pertinent in a memory task

## 6.3 Experiment 3: The word heuristic as conceptual fluency effect on metacognition for episodic memory

### 6.3.1 Overview

In the two previous experiments, we investigated the role of fluency as a cue used for metacognitive judgements in visual perceptual decisions. The idea of these experiments was to extend findings of the metamemory field to perceptual decision making tasks as we suggest that fluency may be a common cue for metacognition across different domains. However, our results so far revealed a small effect of perceptual fluency (induced by prior exposure of stimuli) and showed no effect of conceptual fluency (induced by a word/non-word comparison).

In episodic memory tasks, manipulations of both perceptual fluency (e.g., the font size effect on JOLs, [Rhodes & Castel, 2008](#)) and conceptual fluency (e.g., the relatedness effect on JOLs, [Dunlosky & Matvey, 2001](#)) have been shown to influence metacognitive judgements. As it is possible that our absence of findings in visual perception is due

---

<sup>1</sup>Note that in the previous experiment, performance on trial N and trial N-1 explained the same amount of variance of prospective judgements.



to our manipulation of conceptual fluency, we aim to reproduce the effect of fluency in metamemory with the same manipulation and type of stimuli we used in the previous experiment. Moreover, as in the relatedness effect the metacognitive cue is a diagnostic cue, it is possible that differences on metacognitive judgements based on conceptual fluency arise only when the cues are pertinent to the processing goal of the task. Therefore, we aim to a) reproduce the effect of conceptual fluency on metamemory, b) observe a difference in task performance with our procedure if the conceptual fluency effect on metacognition relies on diagnostic cues.

Metacognitive efficiency was very poor in the previous experiments for prospective judgements suggesting that participants are completely unable to predict their future performance for visual perception. This could be a major difference between metacognition for visual perception and memory as prospective judgements are – although weakly – typically accurate for these tasks (see JOL studies in Table 6.1), possibly because cues intrinsic to the stimuli themselves are diagnostic (e.g. pronounceability, frequency, imageability). We also aimed to reproduce these metamemory findings in the current experiment. Moreover, this experiment allows a comparison of conceptual fluency effects between prospective and retrospective judgements in the memory field which has never been proposed before. As mentioned previously, we suggest that the fluency effect mainly affects prospective judgements compared to retrospective judgements as more pertinent information about performance is available after task completion. To the best of our knowledge, only one study focused on perceptual fluency effect in JOLs and RCJs (Frank & Kuhlmann, 2017). In this study, participants had to learn and recognize words presented in auditory format. Critically, half of the words were presented in a high volume whereas the other half were presented in a low volume. The authors founds that participants gave higher JOLs for loud words however no effect was found for RCJs.

## 6.3.2 Method

### Participants

In accordance with our preregistration documents, we collected data from 130 participants through Prolific Academic. As we were interested in the same effect as the previous experiment we also we wanted to include around 102 participants in the study. We excluded 23 participants with performance below 55% and above 95% in one of the two types of items (i.e., words or non words). One other participant was excluded because he or she used fewer than 3 points on the confidence scale. Finally, we excluded

14 participants with  $\text{meta-d}' < |0.1|$  to avoid dividing by very small values. Thus, we excluded 29.23% of our sample which is consistent with online studies (Chandler et al., 2014). Analyses were therefore conducted on 92 participants (42 in the "prospective" condition and 50 in the "retrospective" condition).

### **Material and procedure**

The main task of the experiment was an episodic memory test using the same type of stimuli (but in English) as the previous experiment. Firstly, participants studied sets of 80 three-letter word or non-word visual patterns for 3000ms each and were asked to remember them. Secondly, they were presented the same set of three-letters without the pattern and had to find which pattern each three-letters word was made of between 2 propositions (2AFC task). As in the two first experiments, half of the participants were randomly assigned to a prospective judgment condition and half to a retrospective judgment condition.

As in the previous experiment, each test trial began by the presentation of the larger picture for 4000ms. Then, participants in the prospective condition had to perform their metacognitive judgements. More specifically, they had to estimate prospectively their confidence in finding the pattern in which a particular three-letter stimulus was presented during the first phase. In the next step, all participants had to perform the visual task and select the correct detail between two answers. They had no time limit to make their choice. After this decision, participants in the retrospective condition had to judge their level of confidence in their answer (see Figure 6.7 for a summary of the procedure). As previously, half of these three-letter stimuli were English words and the other half will be non-words. As previously, the same picture was associated with both a word and a non-word to control for the effect of patterns, and targets and distractors were counterbalanced across participants. The experiment was programmed using Gorilla website and was preregistered on the OSF (<https://osf.io/tkqzw/>).

### **Data and statistical analyses**

As in the two previous experiments, we examined the effects of fluency on metacognitive bias and metacognitive efficiency. We also estimated a  $\text{meta-d}'$  per participants and per item type (previously seen or not seen) using a Bayesian framework (Fleming & Daw, 2017) and one  $m$ -distance (Sherman et al., 2018) per participant and per item type as a quantification of metacognitive bias. All analyses were linear regressions and effect

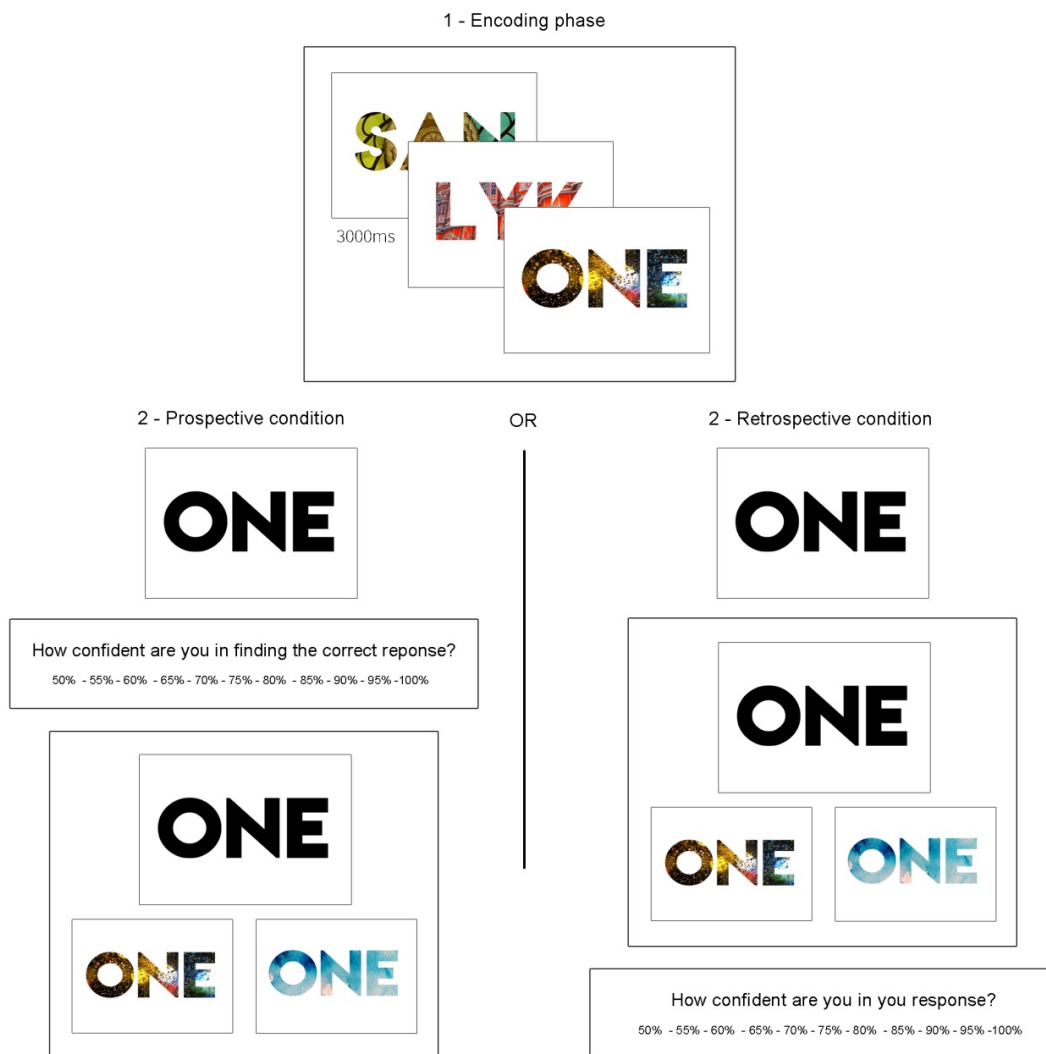


Figure 6.7: Summary of the procedure. Participants were either in the prospective condition or in the retrospective condition. For each participant, half of the stimuli were words and the other half were non-words.

sizes were calculating as  $d_s$  and  $d_z$  according to [Lakens \(2013\)](#). Exploratory analyses included logistic mixed-effect models with participants as random effects. Effect sizes for mixed-effect models were here classical odds ratio.

### 6.3.3 Results

As previously, the result section is organized according to task performance (first-order decision), metacognition efficiency, and metacognitive bias. The last part focuses on the

final recognition task and its relationship with metacognition.

### **First-order performance**

We estimated first-order performance by calculating one  $d'$  per participant and per stimuli condition (words and non-words; see Figure 6.8A). We compared  $d'$  values for item type (words vs non-words) and judgement type (prospective vs retrospective). The analyses revealed a main effect of item type,  $t(90) = -11.08$ ,  $p < .001$ ,  $d_z = 1.15$ , trials with non-words ( $M = 1.86$ ,  $SD = 0.43$ ) being better performed than trials with words ( $M = 1.13$ ,  $SD = 0.78$ ). There were no effect of judgement type,  $t(90) = 0.75$ ,  $p = .457$ , and no interaction between the two factors,  $t(90) = -1.02$ ,  $p = .310$ . Although this seems counter-intuitive, bear in mind that there is no reason why the memory performance should actually be better for words than non-words, since the memory task is purely based on abstract visual information.

### **Metacognitive efficiency**

To estimate metacognitive efficiency, we calculated meta- $d'/d'$  ratio for each participant and each stimuli type (words and non-words; Figure 6.8B). We compared meta- $d'/d'$  ratio for item type and judgement type. Two participants were excluded from the following analysis because they had a higher  $M_{ratio}$  compared to the rest of the group ( $M_{ratio} < 2.5$ ). The analyses showed a significant effect of judgement type,  $t(88) = 9.69$ ,  $p < .001$ ,  $d_s = 2.05$ , with metacognitive efficiency again being higher for participants in the retrospective condition ( $M = 0.86$ ,  $SD = 0.77$ ) compared to the prospective condition ( $M = -0.12$ ,  $SD = 0.53$ ). Participants in the prospective condition did not discriminate between correct and incorrect responses:  $M_{ratio}$  was not different from 0,  $t(88) = -1.57$ ,  $p = .120$ ,  $d_z = 0.17$ . However participants performed correctly in the retrospective condition,  $t(88) = 12.79$ ,  $p < .001$ ,  $d_z = 1.35$ . We also found a significant main effect of item type,  $t(88) = 5.56$ ,  $p < .001$ ,  $d_z = 0.58$ . Word stimuli ( $M = 0.62$ ,  $SD = 0.87$ ) were associated with a better metacognitive efficiency compared to non-words ( $M = 0.23$ ,  $SD = 0.61$ ). Finally, there was an interaction between the two factors,  $t(88) = 2.81$ ,  $p = .006$ ,  $d_s = 0.59$ . Participants had a higher  $M_{ratio}$  for words compared to non-words in the retrospective condition,  $t(88) = 6.27$ ,  $p < .001$ ,  $d_z = 0.66$ , which was not the case in the prospective condition,  $t(88) = 1.85$ ,  $p = .068$ ,  $d_z = 0.20$ .

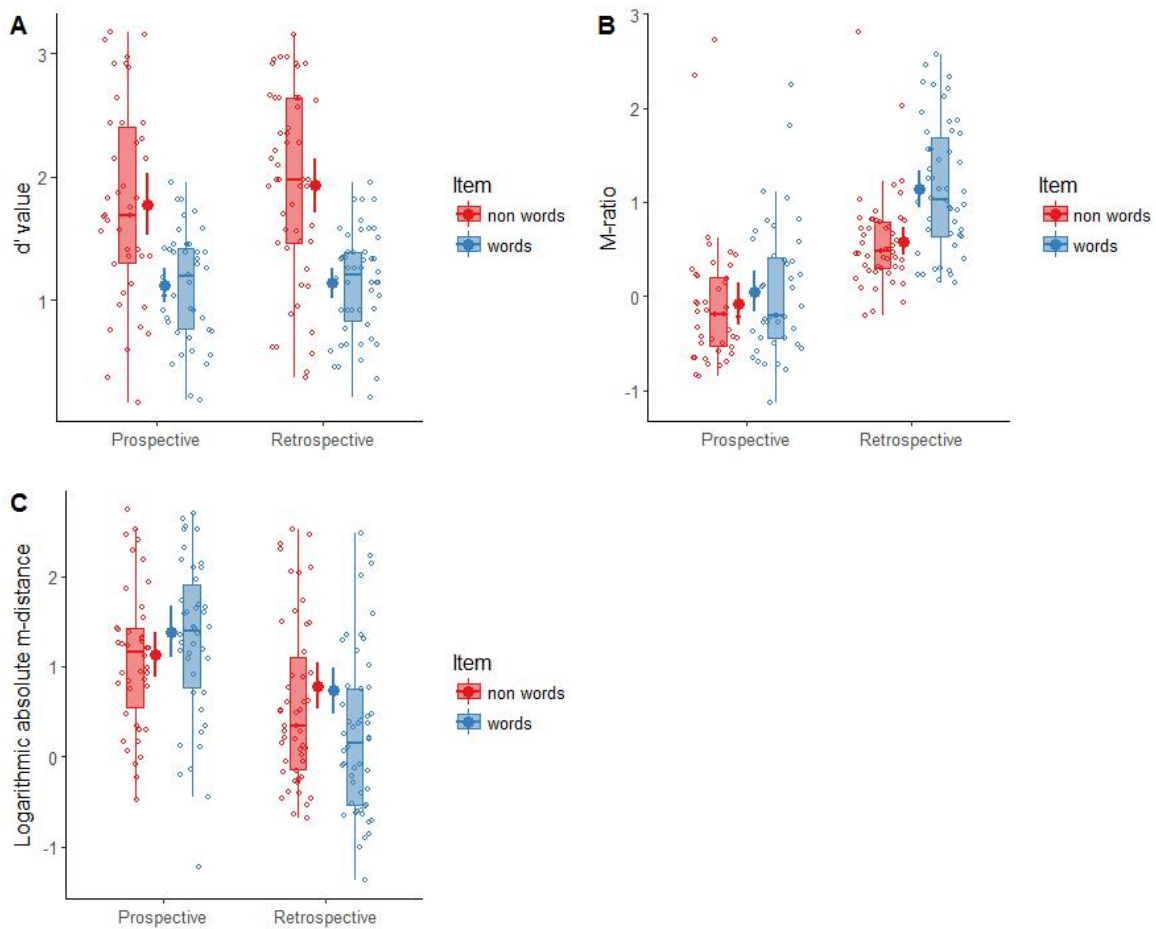


Figure 6.8: (A) Task performance measured by  $d'$  values for words and non-words items according to the metacognitive judgement condition. (B) M-ratio (meta- $d'/d'$ ) values for words and non-words items according to the metacognitive judgement condition. Errors bars are confidence intervals. (C) Logarithmic absolute m-distance values for words and non-words items according to the metacognitive judgement condition. Errors bars are confidence intervals.

### Metacognitive bias

A value of m-distance was calculated for each participant and for words and not non-words (Figure 6.8C). As metacognitive efficiency in the prospective condition did not differ from 0, we analyzed m-distance in absolute values as we did in the previous experiment. We compared m-distance values for item type and judgement type. As previously, we used a logarithmic transformation to fit to the normality assumptions of linear models. The analyses showed a significant effect of judgement type,  $t(90) = 4.88$ ,  $p < .001$ ,  $d_s = 1.04$ , participants again having a higher tendency to report high confidence in the retrospective condition ( $M = 0.41$ ,  $SD = 0.41$ ) compared to the prospective condition

( $M = 1.19$ ,  $SD = 0.85$ ). There was no effect of item type,  $t(90) = 0.74$ ,  $p = .463$ ,  $d_z = 0.08$ , however the interaction between the two factors was significant,  $t(90) = -2.83$ ,  $p = .006$ ,  $d_s = 0.60$ . Participants reported higher confidence for words compared to non-words in the retrospective condition,  $t(90) = 2.64$ ,  $p = .010$ ,  $d_z = 0.28$ , but this was not the case in the prospective condition,  $t(90) = -1.42$ ,  $p = .159$ ,  $d_z = 0.15$ .

As previously, we compare the magnitude of judgements for response accuracy (correct and incorrect), stimuli type (words and non-words), and type of judgement (prospective or retrospective). We conducted mixed-effect model analyses showing main effects of the three factors (Figure 6.9). A higher confidence was given for words than non-words  $t(186.00) = 5.29$ ,  $p < .001$ ,  $d_z = 0.55$ , and a higher confidence was also given for retrospective judgements compared to prospective judgements,  $t(94.00) = 5.38$ ,  $p < .001$ ,  $d_s = 1.13$ . Moreover, correct responses had a higher confidence rate compared to incorrect responses,  $t(6946.00) = 16.17$ ,  $p < .001$ ,  $d_z = 1.69$ . We also found an interaction between type of judgement and response accuracy,  $t(7110.00) = 9.48$ ,  $p < .001$ ,  $d_s = 1.99$ . A higher confidence was given for correct responses compared to incorrect responses in the retrospective condition,  $t(7192.00) = 19.00$ ,  $p < .001$ ,  $d_z = 2.60$ , but this difference was smaller in the prospective condition  $t(7240.00) = 5.32$ ,  $p < .001$ ,  $d_z = 1.98$ . This result is again consistent with Mratio analyses, and with the previous experiments. Finally, we also found a interaction between item type and judgement type,  $t(7085.00) = 4.25$ ,  $p < .001$ ,  $d_s = 0.89$ , showing that a higher confidence was given for words compared to non-words for retrospective judgements,  $t(270.00) = 6.73$ ,  $p < .001$ ,  $d_s = 0.70$ , but this difference was smaller for prospective judgements,  $t(307.00) = 2.80$ ,  $p = .005$ ,  $d_s = 0.29$ , again consistent with the m-distance analysis above. The three-way interaction was also significant,  $t(7092.00) = -2.37$ ,  $p = .018$ ,  $d_s = -0.46$ , however our two two-way interactions of interest did not reach significance, therefore we did not look at the simple main effects (i.e., interaction between the type of item (word and non-words) and the accuracy (correct and incorrect responses) in the retrospective condition,  $t(7180.00) = -1.71$ ,  $p = .088$ ,  $d_z = -0.18$ , and in the prospective condition  $t(7092.00) = 1.51$ ,  $p = .131$ ,  $d_z = 0.16$ ).

### Exploratory analyses

As in the two previous experiments, we were interested in the possibility that participants performed prospective judgements as retrospective judgements regarding the previous trial. We performed mixed effect logistic regressions to see if metacognitive judgements track task performance accuracy for trial N and trial N-1. As the following

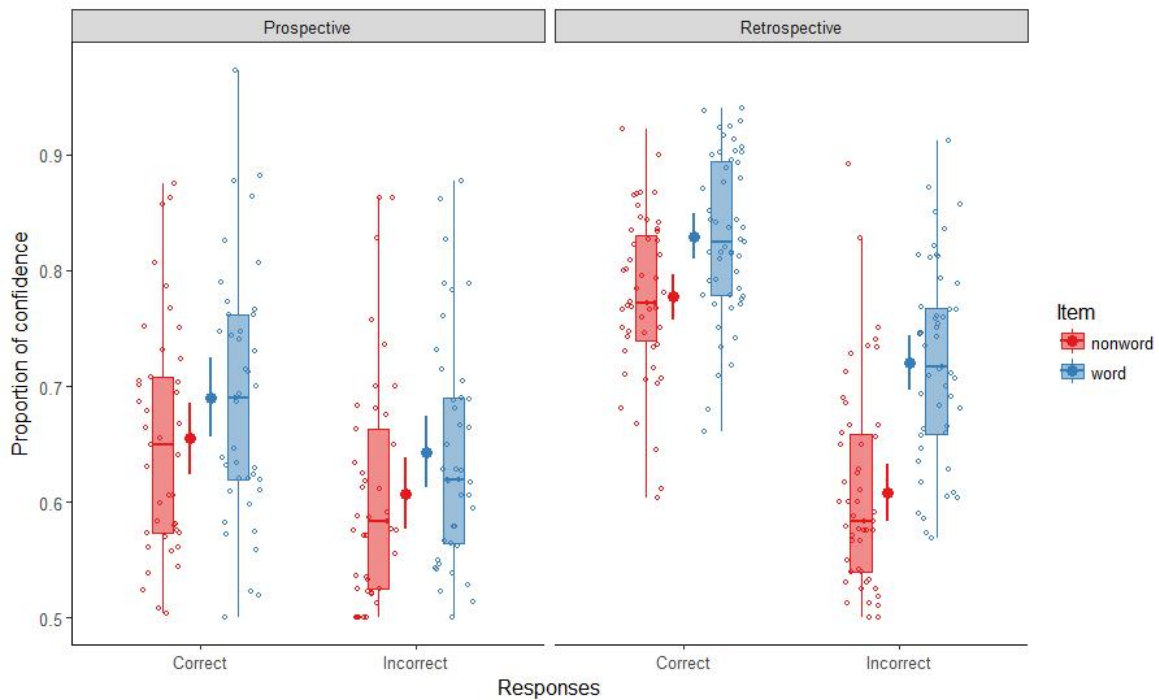


Figure 6.9: Magnitude of confidence according to accuracy of the response and metacognitive judgement condition for words and non-words items.

analyses are performed only on prospective judgements, they only include 52 participants. Each model includes confidence as fixed effects and the estimation of an intercept per participant. They reveal that accuracy was predicted by confidence for the same trial ( $estimate = 0.10, z = 6.64, p < .001, OR = 1.11$ ) as well as for the trial N-1 ( $estimate = 0.09, z = 5.57, p < .001, OR = 1.09$ ). However, model comparison revealed that the two models explained the same amount of variance,  $\chi^2 < 0.01, p > .999$ .

Finally as in the two previous experiments, we focused on how the reaction time for the first-order decision can predict the magnitude of confidence for the two types of metacognitive judgements. A mixed-effect linear model revealed a main effect of reaction time showing that faster reaction times were associated with higher confidence judgements,  $t(7296.00) = -9.44, p < .001, d_z = 0.98$ . However, the interaction between reaction time and judgement type was also significant,  $t(7296.00) = 3.07, p = .002, d_z = 0.32$ . Faster reaction time predicted higher confidence judgements in the retrospective condition,  $t(7260.00) = -11.97, p < .001, d_z = 1.25$ , as well as in the prospective condition,  $t(7297.00) = -6.49, p < .001, d_z = 0.68$ .

### 6.3.4 Brief discussion

This third experiment aimed to reproduce the effect of conceptual fluency on metamemory observed in the literature (e.g., the font-size effect and the relatedness effect). We were also interested in the comparison between prospective and retrospective judgements, suggesting that fluency effect mainly affects prospective judgements as more pertinent information about performance is available after task completion which is also in line with previous results (Frank & Kuhlmann, 2017). Our results highlight a lower *m*-distance for words compared to non-words (i.e., a higher tendency to report high confidence) in the retrospective condition as well as a higher magnitude of judgements in the same condition. However, and contrary to previous studies such as in Frank and Kuhlmann (2017), the effect of conceptual fluency was either absent (for *m*-distance) or less strong (for magnitude of judgements) for prospective judgements compared to retrospective judgements. We will turn to comparison with Experiment 1 and Experiment 2 in the General Discussion, nonetheless, here we want to highlight here that our conceptual fluency manipulation increased the magnitude of judgements in a non-verbal memory task. This is interesting as studies have exclusively measured the effect of fluency on verbal stimuli (when fluency is experimentally manipulated, see Table 6.1).

Interestingly, prospective judgements in metamemory have been shown to be relatively accurate (e.g., gamma correlation being above 0; Kelemen et al., 2000) which was not the case here (metacognitive efficiency being non different from 0). As such, it suggests that there is a major difference between verbal and non-verbal stimuli rather than processes involved in the first-order task (e.g., episodic memory) as possible intrinsic diagnostic cues (e.g. pronounceability, frequency, imageability) mainly pertain to verbal stimuli. In contrast to what we expected we found better visual recognition for non-words compared to words. A likely explanation here is that processing a word can alter the quality of encoding despite the fact that the presence of a word is not directly pertinent for the memory task. This echoes the overshadowing effect which is the phenomenon that describing a previously seen face impairs its recognition (Dodson, Johnson, & Schooler, 1997). In any case, as first-order performance is better for non-words and metacognitive judgements are higher for words, this strengthens the idea of the metacognitive illusion: conceptual fluency creates a false feeling of confidence both prospectively and retrospectively. Finally, as in the two first experiments, we replicated a series of results where retrospective judgements are more accurate than prospective judgements, faster reaction times are associated with higher levels of confidence, and



prospective judgements seem to be related to both the subsequent performance of the trial before and the trial in hand.

## 6.4 General discussion

This set of three experiments investigated the role of fluency in metacognitive judgements. We suggested that processing fluency can be a domain-general metacognitive cue since it has been proposed to influence a large variety of judgements (Alter & Oppenheimer, 2009). Our reasoning was to create metacognitive illusions akin to those in metamemory by manipulating both perceptual and conceptual fluency in a visual perception task. Because we did not find evidence for the involvement of fluency in metacognition for visual perception in our initial measure of metacognitive bias, we aimed to replicate previous findings on memory using one of our experimental manipulations. We also focused on both prospective and retrospective judgements by proposing that the influence of fluency would be lower for retrospective judgements since more pertinent information about performance is available after task completion and as most of metacognitive illusions have been shown on prospective JOLs.

A small effect of perceptual fluency on the magnitude of prospective judgements emerged in Experiment 1 (although not on our initial measure). This effect was in absence of task performance difference as is the case in the font-size illusion, which suggests that perceptual fluency can create a metacognitive illusion in both episodic memory and visual perception. This finding would benefit from being replicated, especially using a protocol that compares the two first-order task. It would however be crucial to find an experimental manipulation pertinent for both visual perception and memory. The disadvantage of using a previous exposure manipulation in episodic memory is that it modifies the memory process itself. In return, using a manipulation for the visual perception task that is often used in the metamemory field (e.g., the size of the font) would probably modify the perception process. In this respect, the comparison between Experiment 2 and Experiment 3 may be more reliable as both experiments use the same material and the same manipulation of fluency.

Conceptual fluency generated by a word/non-word manipulation showed no effect on metacognitive bias or magnitude of judgments for visual perception (Experiment 2). On the contrary, we found a higher tendency to report high confidence in the retrospective condition for words as well as a higher magnitude of judgements in both

prospective and retrospective condition for episodic memory. In the metamemory literature, the conceptual fluency effect on metamemory appears when there is also a difference in task performance. As the word/non-word manipulation did not affect task performance in Experiment 2, this could explain why we did not find a difference in metacognitive bias. Conversely, Experiment 3 created performance differences between words and non-words (however these effects are in the opposite direction compared to the metacognitive effect). It is therefore possible that conceptual fluency cues affect metacognitive judgments only when these cues are pertinent of task performance that is have a influence on the first-order performance regardless of the direction, and idea developed in Chapter 7. Importantly, our conclusion is based on the comparison on two different studies and one should again consider the comparison between visual perception and episodic memory in the same experiment in order to strengthen this claim.

This chapter generated a range of consistent and inconsistent findings. Whereas perceptual fluency seems to influence the magnitude of prospective metacognitive judgements in visual perception (Experiment 1)<sup>2</sup> as it has been shown in episodic memory (e.g., [Rhodes & Castel, 2008](#)), conceptual fluency seems to influence metacognitive judgements only in episodic memory (Experiment 3) and not in visual perception (Experiment 2). This leads to the idea that the effect of fluency on metacognitive judgements is not consistent across different means of generating fluency. Another possibility is that the contribution of fluency in previous manipulations (e.g., font size, relatedness between the target and the cue, etc.) and ours (previous exposure and word/non-word) is limited, and that beliefs are also important as previously suggested ([Mueller et al., 2013](#); [Mueller et al., 2014](#)). This is also highlighted by [Alter and Oppenheimer \(2009\)](#): “classifying studies as ‘fluency effects’ requires an important caveat: The independent variables in those studies may not have explicitly manipulated processing fluency ease, so we cannot conclude absolutely that those effects were driven by differential fluency” (p. 220). The three experiments proposed here have no direct measurement of fluency and therefore conclusions exclusively with respect to fluency are limited. Thus, it is possible that beliefs influence judgements in the opposite direction from the fluency effect in some situations leading to an absence of effect. Here it is proposed that beliefs are situation-dependent and possibly different for visual perception and episodic memory

---

<sup>2</sup>Note however that for magnitude of judgement, if this effect is strong in Experiment 3 ( $d_z$  for prospective and retrospective judgements are both above 1.5) this effect is way smaller in Experiment 1 ( $d_z = 0.23$ ). Moreover, we again found not effect on our initial measure of metacognitive bias (m-distance) largely limiting our conclusion.

as well as for perceptual fluency and conceptual fluency manipulation. We will develop further this idea in the general discussion.

Because few works have focused on prospective judgements in visual perception and in line with our results in Experiment 1 and Experiment 2 (very low or even negative metacognitive efficiency), we suggest that people can only predict their future memory compared to their visual perception. Interestingly, Experiment 3 showed a similar pattern of results in a memory task. Thus, it is possible that the effect of accurate prospective judgements largely found in the metamemory literature is due to the use of verbal materials: with memory, people might have prospectively access to a verbal content (e.g., TOT phenomenon). Because in Experiment 3 we proposed a visual recognition task (i.e., the same material as Experiment 2) such access was not possible and metacognitive efficiency was not different from zero. The ability to prospectively evaluate ourselves seems to be related to the modality that is used, which suggests a qualitative difference between prospective and retrospective judgements (see Chapter 7 for more discussion of this point). Finally, other findings were also consistently found across the three experiments, notably the fact that faster reaction times were associated with higher levels of confidence and that prospective judgements seem to be related to both the subsequent performance of the trial before and the trial in hand. These outcomes support the idea that similarities can be found in metacognition for visual perception and episodic memory.

## Chapter 7

# General discussion and perspectives

Research into metacognition to date has had a diversity of methods, analyses and even research goals (Chapter 1). There has been a long tradition of studying metacognitive judgements within fields for different purposes such as understanding processes at play in self-evaluation, assessing mechanisms that can enhance learning, or evaluating the state of consciousness in perception or memory. Nonetheless, similarities in methodological and theoretical frameworks can be observed suggesting the idea that metacognition could be domain-general. The comparison between the metamemory and the metaperception field (Chapter 2) generated two questions: whether people use a common resource in their metacognitive judgements across different types of first-order tasks and whether this resource is also shared across different metacognitive judgements (Chapter 3).

We identified the cue-utilisation framework (stemming from the metamemory literature) as a good theoretical candidate for studying potential processes involved the generality of metacognition. We also used the meta-d' framework (stemming from the metaperception literature) to quantify metacognition as it allows control of diverse biases. To investigate our two main questions, we proposed to measure cross-task correlations for both metacognitive bias and metacognitive efficiency (Chapter 4) that we then compared across judgement types (Chapter 5). Finally, we focused on the cue-utilization and especially experience-based processes to investigate the potential domain-general role of fluency (Chapter 6) in visual perception and episodic memory.

### 7.1 Summary of results

We investigated the breadth of metacognition using two means of assessing the specificity or the generality of a cognitive process (Chapter 3). Therefore, our empirical sections include two correlational chapters (Chapter 4 and Chapter 5) and one chapter

based on functional independence (Chapter 6). Table 7.1 proposes a summary our main results. Correlational studies address the domain-generality of metacognition whereas Chapter 6 summarises effects of fluency.

Our first novel contribution was to measure RCJs across four different domains as has been done recently for memory and visual perception (Baird et al., 2013; Baird et al., 2015; McCurdy et al., 2013; Morales et al., 2018; Lee et al., 2018). Our rationale was to compute cross-task correlations for both metacognitive bias and metacognitive efficiency using a large sample size. If a general resource underpins the formation of metacognitive judgements, people who are accurate in their evaluation in one task should also be accurate for another task. Our results revealed that both metacognitive bias and metacognitive efficiency share common resources across first-order tasks. If the generality of bias has often been found in previous works (Ais et al., 2016; Baranski & Petrusic, 1994; Baranski & Petrusic, 1995; Lichtenstein & Fischhoff, 1977; Song et al., 2011) this study provides for the first time evidence for a domain-general resource involved in metacognitive efficiency across four tasks. The percentage of explained variance however suggests that both domain-specific and domain-general resources are involved in metacognitive efficiency and that the involvement of this general resource might vary across tasks.

As we found cross-task correlations for RCJs, we then wondered whether such correlations also occur for other metacognitive judgements in Chapter 5. Using similar methods that we used in Chapter 4 and which allow a strict control of different variations in metacognition (see Chapter 3), we proposed to revisit the classical distinction between eFOK and sFOK. This distinction is particularly pertinent to test the breadth of metacognition as these judgements have been shown to be dissociated (see Introduction section in Chapter 5). Moreover, we were interested in the reproduction of our previous findings in RCJs. Regarding metacognitive efficiency, we reproduced the cross-task correlations between eRCJ and sRCJ found in Chapter 4 but crucially this relationship did not emerge for FOKs. On the contrary, cross-task correlations for metacognitive bias were present for both FOKs and RCJs meaning that people's general level of confidence was similar across the two tasks.

Finally, we were interested in the experimental manipulation of metacognitive cues. As hypothesised in Chapter 2, experience-based metacognition could occur across different types of first-order tasks and judgements through the fluency heuristic. In both correlational chapters, we indeed suggested that fluency was possibly a common cue for gauging the correctness of responses that could explain the cross-task correlations

Table 7.1: Summary of our main results regarding the domain generality of both metacognitive bias and metacognitive efficiency according to the type of study, the type of metacognitive judgements that have been measured, and the nature of first-order tasks. EM = episodic memory; VP = visual perception; SM = semantic memory; EF = executive function; PCJ = prospective confidence judgment; RCJ = retrospective confidence judgement.

Study	Type	Metacognitive judgement	First-order tasks	Metacognitive bias	Metacognitive efficiency	Comments
Chapter 4	Correlational	RCJ	EM / VP / SM / EF	Domain-general	Domain-general	Trend correlation for EM/VP bias. HDI for cross-task correlation EM/VP efficiency overlaps with 0.
Chapter 5	Correlational	FOK	EM / SM	Domain-general	Domain-specific	-
		RCJ		Domain-general	Domain-general	-
Chapter 6 - Exp1	Seen / not seen manipulation	PCJ	VP	Higher magnitude for seen items	No difference	Small effect size $d_z = 0.23$ and no effect on m-distance
		RCJ		No difference	No difference	-
Chapter 6 - Exp2	Word / non-word manipulation	PCJ	VP	No difference	No difference	-
		RCJ		No difference	No difference	-
Chapter 6 - Exp3	Word / non-word manipulation	PCJ	EM	Higher magnitude for words	No difference	Higher task performance for words and non-words
		RCJ		Higher magnitude for words and lower m-distance	Higher Mratio for words	

found for RCJs. As is the case in the literature [Alter and Oppenheimer \(2009\)](#), we distinguished between several types of fluency. The metamemory literature has shown that a greater experience of fluency increases the magnitude of judgements regardless of the type of fluency. Thus, Chapter 6 examined both perceptual fluency (such as induced by the size of the font; [Rhodes & Castel, 2008](#)) and conceptual fluency (such as induced by a higher relationship between the target and the cue; [Dunlosky & Matvey, 2001](#)). Based on the cue-utilization view of metamemory, we investigated these types of fluency in a visual perception task (Experiment 1 and 2) and in a memory task (Experiment 3) in both prospective and retrospective judgements. Here the focus was on metacognitive bias rather than metacognitive sensitivity because these two types of fluency are not supposed to be diagnostic heuristics (in contrast with answer fluency). First, in perceptual tasks, we found that perceptual fluency induced by pre-exposure had a small effect on the magnitude of prospective judgements (however without affecting our measure of metacognitive bias that is m-distance, Chapter 3) despite no effect on retrospective judgements. When trying to reproduce this effect with conceptual fluency induced by a word/non-word manipulation, no effect was found either on metacognitive

bias (m-distance) or magnitude of judgements. Finally, the same manipulation used in an episodic memory task revealed an effect on retrospective judgements only for metacognitive bias (m-distance) and an effect on both prospective and retrospective judgements for magnitude. As suggested in Chapter 6, inconsistencies between effects detected in the magnitude of confidence but not detected using m-distance are likely due to the quantification of metacognitive bias itself. As the estimation of m-distance needs parameters from the meta-d' model (both meta-d' and meta-c2) and as the number of trials per item type is low in Experiment 1 ( $N = 26$ ), it is possible that estimations of m-distance were not very accurate. Another likely explanation is that m-distance for prospective judgements is not precise because prospective metacognitive sensitivity is inexistent for these judgements.

Finally, a pattern of findings was consistently found across experiments in Chapter 6. Particularly, we found that metacognitive efficiency was higher for retrospective judgements compared to prospective judgements, the latter being non-different from zero. Prospective judgements were actually found to be related to both the subsequent performance of the trial before and the trial in hand. Faster reaction times were associated with higher levels of confidence only for retrospective judgements. We will now discuss these results in light of theoretical frameworks with putting the emphasis on implications for research on metacognition but also for memory research, clinical research, and research on consciousness more broadly

## 7.2 Implications for research on metacognition

The core question of this thesis is the domain-generalty of metacognition. We proposed to investigate how wide are metacognitive processes in terms of first-order tasks and types of judgements. Thus, we have highlighted the following questions to which we will briefly answer here in lights of our results and that we will re-consider in the next sections.

- Is there a common resource in metacognition in terms of first-order tasks?

Chapter 4 revealed cross-task correlations for metacognitive efficiency for RCJs across 4 tasks suggesting the existence of a domain-general resource. It is important to note that HDIs for the episodic memory task and the visual perception task overlapped with zero

(still showed a substantial probability mass above zero). We proposed that experience-based metacognition through the answer fluency heuristic was common across first-order tasks and is therefore a good candidate for a domain-general process of metacognitive efficiency. However, we proposed that the involvement of this general resource might vary across tasks. We identified that tasks having the highest correlations were those for which one can infer task difficulty easily (i.e., easily available difficulty signal such as for visual perception and semantic memory). Here we suggested that the perceived difficulty also pertains to the fluency heuristic as it could be inferred from RT. We also reproduced previous findings regarding the domain-generality of metacognitive bias using cross-task correlation for confidence level. Here again we note that there was a trend correlation for metacognitive bias across the episodic memory task and the visual perception task. These findings were replicated in Chapter 5 for an episodic memory task and a semantic memory task. In sum, our results suggest the involvement of a domain-general resource in both metacognitive bias and metacognitive efficiency for RCJs.

- Is this common resource involved in both prospective and retrospective judgements?

Regarding metacognitive efficiency, Chapter 5 revealed cross-task correlations for episodic and semantic memory tasks in RCJs despite an absence of cross-task correlations for the same tasks in FOKs. It supports the idea that metacognitive efficiency depends on the type of judgement that is made. Specifically, processes involved in prospective judgements such as FOKs seem to vary across the nature of the first-order task which is not the case for retrospective judgements. On the contrary, cross-task correlations for metacognitive bias occurred in both FOKs and RCJs suggesting that bias is the same regardless of the type of judgement.

- Is experience-based metacognition a common process used across episodic memory and visual perception for both prospective and retrospective judgements?

In Chapter 6, perceptual fluency (induced by pre-exposure of stimuli) slightly increased the magnitude of prospective judgements despite no difference in first-order performance. This effect is akin to the perceptual fluency effects on prospective metamemory judgements. This suggests again the generality of metacognitive bias regardless of the type of first-order task. However, this effect did not emerge for conceptual fluency in visual perception suggesting either an experimental manipulation failure or that fluency is



less unitary than previously proposed (Alter & Oppenheimer, 2009). The same manipulation affected metacognitive bias in an episodic memory task however this effect emerged mainly on retrospective judgements contrary to Experiment 1. Thus, a complex pattern emerges, nonetheless we can first conclude that fluency influences prospective and retrospective judgements differently. Moreover, this influence seems also different according to the first-order task. We will discuss the potential generality of fluency and experience-based metacognition further in the next section.

In light of our results and our answers to the proposed above questions, we will now extend our conclusion to the cue-utilization framework and to metamemory and metaperception more broadly. Finally, we will re-consider our initial definition of metacognition according to the three dimensions proposed in Chapter 1.

### 7.2.1 Experience-based and information-based metacognition

A prominent theory suggests that separable information-based processes and experience-based processes are at play in the formation of metacognitive judgements (Koriat, 1997; Koriat & Bjork, 2006; Koriat et al., n.d.; Koriat & Levy-Sadot, 1999). Information-based metacognition (high level metacognition see Chapter 1) involves inferential processes from explicit theories or beliefs. On the contrary, experience-based metacognition (low level metacognition see Chapter 1) refers to the influence of heuristics that are automatic and non-analytic. We will here detail our results in light of this standard distinction.

#### Information-based metacognition

The contribution of fluency as experience-based metacognition and beliefs as information-based metacognition in judgements is difficult to disentangle. Through our experimental chapters, we have mainly highlighted the involvement of beliefs, essentially in the discussion of Chapter 6 as there is no direct measure of processing fluency in our three experiments. Based on Dunlosky, Mueller, and Tauber (2014)'s distinction defined in the context of memory, we propose here that there are two types of beliefs, namely online beliefs or low-level beliefs and higher-order beliefs. Higher-order beliefs are prior beliefs about memory that can have been learnt either by associative learning or by explicit encoding. Such an idea echoes the literature of naïve theories in social cognition referring to “what make it easy or difficult to think of things or to process new information” (p.332, Schwarz, 2004). Naïve theories are involved in a model of fluency

on judgments and decision making (Unkelbach & Greifeneder, 2013). They allow a link between what people experience (e.g., ease) and how these experiences are interpreted through inferential processes (Schwarz, 2004). As such, these theories allow a use of fluency according to context, and explain that ease-of-processing has a different effect for different judgements. They are a priori explanations of the experience of fluency as they are supposed to exist before the experience of a particular task.

On the other hand, online beliefs or low-level beliefs can be created online during a particular task. Dunlosky et al. (2014) distinguish two kinds of task-dependent beliefs involved in JOLs: they can be either created following performance feedback or be created online in response to task demands. In the former case, the presence of feedback or outcomes allows participants to update their beliefs. For instance, people might not be aware of the benefit of an encoding strategy during the first test therefore predicting no difference between item repetition and mental imagery in terms of JOLs (Dunlosky & Hertzog, 2000). However, such a difference appears during a second test (for a similar effect with massed and spaced encoding see Logan et al., 2012). These kind of online beliefs learnt by feedback are therefore more likely to influence metacognitive sensitivity which has been shown in RCJs for a general knowledge task (Sharp, Cutler, & Penrod, 1988). In the second case, when giving JOLs, participants might be engaged in an analytic process (Dunlosky et al., 2014) because of the explicit instruction of the task. Thus, they look for any kind of variability across items that can be used as a metacognitive cue which would rather affect metacognitive bias. This analytic mode creating online beliefs has been used to explain the font-size effect (Mueller et al., 2014) because “focusing people on evaluating memory per se triggers an analytic mode of processing where people search for cues” (p.9).

Even if the involvement of fluency is limited in effects such as the font-size effect, the question as to whether people have general beliefs irrespective of the cognitive task is of interest. Experiment 1 in Chapter 6 may have generated the belief that the task would be easier for preciously seen pictures as the stimuli have been already processed compared to other pictures explaining the difference in magnitude of judgements. However, it is difficult to claim for the use of general beliefs across memory and visual perception even though these effects (font-size effect and Experiment 1) are in the same direction because the means of generating perceptual fluency are different (size of the font and pre-exposure of stimuli). It is indeed possible that the use of two different task-dependent beliefs modify metacognitive judgements in the same direction. However,

because Experiment 2 in Chapter 6 revealed no effect of conceptual fluency on metacognitive judgements for visual perception but Experiment 3, for memory, did show an effect, we suggest that participants are more likely to have specific beliefs regarding words/non-words manipulation that would exclusively be applied to episodic memory. These beliefs influence metacognitive bias, therefore we propose that beliefs used in Chapter 6 are online beliefs based on an analytic mode and are domain-specific (i.e., can differ across cognitive tasks). In other words, performing the task cues participants to use variability in the task in their judgements when they think that is this variability is relevant. The word/non-word manipulation likely created the belief that remembering the word would help in recognising the visual pattern (probably because performance is usually better for remembering a word compared to a non-word). This belief was only applied during the episodic memory task as it is meaningless in a visual perception task.

Here, we also suggest that online beliefs and higher-order beliefs have a hierarchical structure. Pertinent online beliefs (mainly online beliefs based on feedback as there are more likely to be accurate) might be generalised over time to other similar contexts to create higher-order beliefs. We propose that these beliefs are those we referred to in Chapter 5: they account for metacognitive sensitivity in a domain-specific manner. In eFOK formation, these beliefs refer to both the integration of auto-noetic consciousness as a result of the deliberative memory search and the efficacy of memory (Hertzog, Dixon, & Hulstsch, 1990) and the self (Boduroglu et al., 2015). In sFOK, these beliefs are shared across people (Juslin, 1993) as metacognitive sensitivity is better when calculated using responses which are the most chosen by participants (consensual responses) compared to actual correct responses (Koriat, 2008). These beliefs have been generalised because they are diagnostic of task performance. If there are less situation-specific, they remain domain-specific. Thus, they can vary across episodic memory, visual perception, or semantic memory.

### **Experience-based metacognition**

If metacognitive sensitivity is influenced by such inferential domain-specific processes, metacognitive judgements are likely not correlated and one can be metacognitively efficient in a task despite being inefficient in another task (see FOKs in Chapter 5). So, how can we explain the cross-task correlations for metacognitive efficiency found in Chapter 4? In Chapter 5, we have suggested that retrospective judgements rely more on the evidence driven by the given response and experience-based metacognition (e.g., answer fluency) whereas prospective judgements (such as FOKs) are more based on analytic

processes and information-based metacognition. This supports the idea that there is a massive difference between prospective and retrospective judgements. Similarly [Fleming et al. \(2016\)](#) found no correlations between prospective judgements and retrospective judgements for AUROC2 in a visual perception task. In the field of metamemory, similar absences of correlations for gamma have also been highlighted within prospective judgements ([Kelemen et al., 2000](#)) which suggest that different processes are at play in these judgements. Here we argue that this is likely due to the main use of information-based metacognition which is domain-specific (as suggested above) and that can also vary across prospective metacognitive judgements as task demands are different. As such, eFOK and sFOK metacognitive efficiency did not correlate in Chapter 5 and there was also an absence of correlation between JOL, EOL, and FOK sensitivity in [Kelemen et al. \(2000\)](#)'s study (2000).

On the contrary, we suggest that retrospective judgements are more based on experience-based metacognition using heuristics that are likely more task-general (see Chapter 4 and Chapter 5). For instance, we suggest that answer fluency is a common cue for gauging the correctness of their responses and is therefore a diagnostic cue across first-order tasks (Chapter 4 and Chapter 5). This heuristic can be enhanced by shared aspects of the state space such as motor responses being shared across tasks. Thus, several works have found the motor response of the first-order task improves metacognitive efficiency in a memory task ([Siedlecka, Skóra, et al., 2019](#)) and similar results have been found in a visual perception task ([Siedlecka, Hobot, et al., 2019](#)). When this information is not available ([Wokke, Achoui, & Cleeremans, 2019](#)) or altered ([Favre et al., 2020](#)) during the second-order task, metacognitive efficiency is reduced. We suggest here that this sensory-motor information pertains to the fluency heuristic: it allows an evaluation of the speed of the first-order task answer. When this information is disrupted, participants are less able to rely of the diagnostic answer fluency heuristic to discriminate between correct and incorrect. To test this hypothesis, it would be relatively straightforward to examine cross-task correlations as we did in Chapter 4 in relationships between situations where participants have no time pressure to perform the first-order decision and situations where participants have to wait a certain amount of time before performing their decision. If participants base their confidence on the RT difference between correct and incorrect responses (as in the second situation) we would observe lower cross-task correlations when response time is controlled.

Although naïve theories or beliefs can change the interpretation of fluency with respect to the context, an important feature of model of fluency on judgments ([Unkelbach](#)

& Greifeneder, 2013) is that it has been proposed that it does not change the experience of fluency that “exerts the same influence on judgements independently of how it is generated” (p.220; Alter & Oppenheimer, 2009). Thus, we suggest that fluency is a domain-general heuristic. When fluency is diagnostic of task performance (e.g., answer fluency), it influences metacognitive sensitivity. When it applies to stimuli regardless of first-order accuracy (e.g., perceptual fluency), it influences metacognitive bias. This knowledge is essentially implicit. When people can attribute the effect of fluency to a source independent from the judgements, the effect of metacognitive bias disappears. In perceptual fluency attribution model (Bornstein & D’agostino, 1992), the effect on judgements of pre-exposure is stronger when people are not aware that they have been pre-exposed to these stimuli. Fluency is a proxy for judgements as long as this effect is not discounting by other information that are more diagnostic. Similarly, perceptual fluency manipulations induced by cue priming have been shown to influence the magnitude of FOKs only when this priming is suprathreshold (for an overview, see Narens et al., 1994).

### **A proposal for metacognitive judgement formation**

In essence, we suggest that prospective metacognitive judgments are more based on information-based metacognition using domain-specific beliefs. These beliefs can either modify metacognitive bias when they are created online using an analytic mode or affect metacognitive sensitivity when they are learnt from feedback. Here we propose a hierarchical structure of beliefs whereby online beliefs are integrated over time by associative learning or explicit learning becoming higher-order beliefs. In return, these types of beliefs can also influence metacognitive judgements. Figure X is a summary of our proposal including examples of types of beliefs and fluency. Chapter 6 also consistently showed that the magnitude of prospective judgements in one trial is equally (Experiment 1 and Experiment 3) or even more (Experiment 2) explained by the first-order accuracy of the previous trial. When participants are not able to apply pertinent beliefs in order to perform the prospective task (i.e., when metacognitive efficiency is absent as in our three experiments in Chapter 6), participants rely on readily available internal feedback from the previous trial.

Interestingly, a recent study evaluated the domain-generality of metacognition in children for strategy selection in an episodic memory task and an arithmetic task (Geurten, Meulemans, & Lemaire, 2018). They found that cross-task metacognitive sensitivity for RCJs was absent at 8-9 year olds, small to medium at 10-11 year olds, and medium to

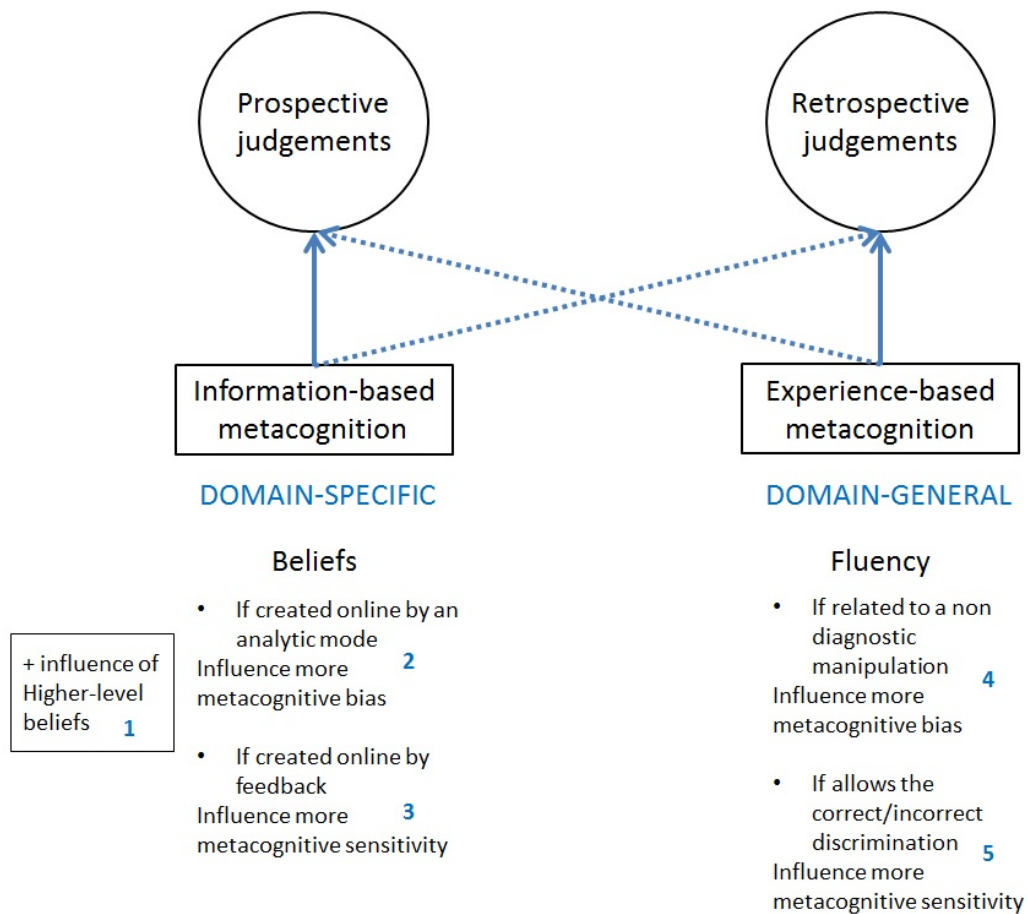


Figure 7.1: Proposed metacognitive judgement formation. Prospective judgements are more underpinned by information-based metacognition (solid arrow) rather than experience-based metacognition (dashed arrow) and the reverse is proposed for retrospective judgements. (1) Example of higher-level beliefs (context-dependent beliefs and global self-beliefs): “I should retrieve this information as most people know this”. (2) Example of online beliefs created by an analytic mode: “In this experiment, words written in a large font are easier to remember”. (3) Example of online beliefs created by feedback: “In this experiment, it seems that I can better remember high frequency words compared to low frequency words (4) Answer fluency is an example of diagnostic fluency. (5) Perceptual fluency is an example of non diagnostic fluency.

large at 10-11 year olds. Moreover, they proposed that “as information-based processes are conscious and effortful, they are probably more likely to be generalized to other domains than experience-based processes” (p.76). Although we do not disagree with this idea, we suggest that higher-order beliefs are more likely to be domain-general rather than low-level. Thus, it is possible that young children mainly use low-level online beliefs that are situation specific whereas older children are more likely to use less specific higher-level beliefs that have emerged during development. As also suggested

in Chapter 4, cross-task correlations for metacognitive efficiency in RCJs could also reflect the ability to generalise priors from one task to another an idea akin to higher-level belief. Another explanation pertains to the acquisition of the fluency heuristic. It suggests the idea that the cognitive system is able to detect variations in terms of RT and that it uses it through the implicit knowledge that longer RTs are more likely to be associated with incorrect decisions. It is possible that this knowledge has been integrated over time (with repeated experiences) and therefore one can suggest that children would rely less on the answer fluency heuristic as it would have been less automatise.

Finally, our proposal implies that that prospective judgements require a higher amount of time to be performed as information-based metacognition is proposed to be a slow process related to “System 2” whereas experience-based metacognition is faster and related to “System 1” (see Chapter 1). Complementing this hypothesis, we performed post-hoc analyses on RT for prospective and retrospective judgements in Experiment 2 of Chapter 6. This analysis (using mixed-effect models with participants as random factor) showed that RT for performing prospective judgements were indeed longer than those for retrospective judgements ( $t(96.99) = -3.85, p < .001, d_s = 0.78$ ).

Overall, we propose that prospective metacognitive judgments are more based on information-based metacognition using domain-specific beliefs, and that retrospective judgements are more based on experience-based metacognition that is more likely to be domain-general (see Table 7.1). This proposal and several of our findings have also implication for the fields of metamemory and metaperception .

## 7.2.2 Metamemory and metaperception

### Qualitative differences

As Chapter 2 proposed a comparison of metamemory and metaperception, we will here compare the two fields in lights of our results in Chapter 4 and Chapter 6 in order to propose a more unified view of metacognitive research. In Chapter 4, we found a better metacognitive efficiency for episodic memory and semantic memory compared to visual perception. Even if this hypothesis has not been directly tested in the literature, qualitatively it appears that metacognitive efficiency is usually better for memory tasks than visual perception tasks (see Table X for a summary including our Chapter 4). To the best of our knowledge, only two studies directly tested this, with statistical analysis. [McCurdy et al. \(2013\)](#) found a non significant difference between sensitivity for the two

tasks ( $p = .152$ ;  $N = 34$ ). [Morales et al. \(2018\)](#)'s analysis revealed a better sensitivity for memory than visual perception ( $N = 30$ ; no overlap between the two HDIs). According to the distribution plots of individual values in other studies ([Baird et al., 2013](#); [Baird et al., 2015](#); [Lee et al., 2018](#)), sensitivity for memory seems higher than sensitivity for visual perception, however mean data are not presented and the differences between the two was not tested. Our findings in Chapter 4 bring another piece of evidence in this direction. Although this comparison is just descriptive and includes different experiments, both in Experiment 1 and 2 (perception) in Chapter 6 we found a lower metacognitive efficiency than in Experiment 3 (memory; in retrospective condition only).

We propose here that this is due to higher-order metacognitive beliefs that are qualitatively different across memory and visual perception. Influence of beliefs on metacognitive judgements has been mainly studied in metamemory. These beliefs have been identified using questionnaires such as the Memory in Adulthood questionnaire ([Dixon & Hultsch, 1983](#)) or the Everyday Memory Questionnaire ([Sunderland, Harris, & Baddeley, 1983](#)). In this context, the notion of memory self-efficacy, "a generalized judgment that is abstracted from specific task and situation characteristics" (p. 212; [Beaudoin & Desrichard, 2011](#)) has been developed. It is seen as a set of beliefs that are organised in a hierarchical manner from local to global beliefs ([Hertzog & Dixon, 1994](#)). In self-evaluation, these beliefs are used according to the situation: if a situation is very familiar, beliefs that are used are more likely to be low-level beliefs. On the contrary, any new situation would rather imply the use of higher-level beliefs ([Hertzog et al., 1990](#)). Here, we suggest that these higher-order beliefs have been created by the generalisation from lower-beliefs particularly because they were efficient and accurate<sup>1</sup>. As such, the higher the number of higher-level beliefs for a cognitive domain, the greater the metacognitive sensitivity in this domain.

We therefore propose that such beliefs are less present for visual perception, and this is perhaps amenable to testing via the use of questionnaires and qualitative reports of how participants think about the tasks they use. Metaperception perhaps relies more on implicit knowledge that has been learnt over time by associative learning (e.g., association between RT and the outcome of a decision). As such, beliefs are more prominent in episodic memory in comparison with visual perception. This is probably strengthened by the fact that the visual perception tasks used in this thesis (and in the literature) have a lower ecological validity compared to the memory task. Whilst learning lists of

---

<sup>1</sup>We also support the idea that pathological generalisation of beliefs exists (see section "Implication for patients").



words or information is relatively common in the real world (e.g., leaning a shopping list, a phone number, information for an exam, etc.), comparing two close visual stimuli (discrimination tasks) or judging whether you saw a stimulus or not (discrimination tasks) rarely occurs in the same overt, repeated manner as when in a laboratory task. Moreover, memory processes are more often questioned in real life and have probably a higher number of metacognitive thoughts about them (e.g., “I am sure I had another item on my shopping list but I cannot remember what it is”) compared to visual perception. We nonetheless believe that metaperception thoughts also occur in daily life, however we suggest that their lower frequency compared to metamemory thoughts and the fact that experimental memory tasks have a higher ecological validity make metamemory beliefs more accurate.

Whilst we have proposed that overall more experience-based metacognition is used in RCJs, it is nonetheless possible that the part of beliefs in RCJs is higher in episodic memory as these beliefs are more prominent. This is also consistent with the fact that the contribution of a putative global resource may differ according to the cognitive domain (see Chapter 4): domains that rely less on metacognitive beliefs are supposed therefore to be more highly correlated.

Although the majority of work regarding beliefs has been dedicated to memory, recent work in visual perception decision making has shown that “global estimates of performance” can be generated from local RCJs and that these estimates are accurate and efficient (Rouault, Dayan, & Fleming, 2019). People are indeed able to detect trial-by-trial variation in local confidence to construct beliefs about the task even in the absence of performance feedback. In return, these beliefs can be used as prior expectations that have been shown to influence metacognitive efficiency even in visual perception (Sherman, Seth, Barrett, & Kanai, 2015). In essence, we suggest that prospective judgements in visual perception are inaccurate (i.e., very low metacognitive sensitivity) as beliefs about these processes are less common than those about memory possibly due to real life training for both memory and metamemory.

Another possibility is that beliefs also differ in terms of stimuli that are used. A large body of work that has focused on prospective judgements (that are mainly influenced by beliefs) has used verbal stimuli. As verbal stimuli are associated with a strong conceptual network, we suggest that beliefs are more prominent for these stimuli compared to non-verbal stimuli. This is likely due to the fact that more information is available in verbal stimuli and trial-by-trial variations might be easier to discriminate. This is

consistent with the fact that we found a very low metacognitive efficiency for a non-verbal episodic memory Experiment 3 in Chapter 6 for prospective judgements despite the fact these judgements are supposed to be mainly underpinned by beliefs.

To disentangle these two likely explanations (more accurate metamemory beliefs and/or more accurate metacognitive beliefs about verbal content), it would be interesting to compare metacognitive efficiency (using for instance RCJs) for two types of visual perception tasks using material with verbalisable content (e.g., object categorisation) and material with less verbalisable content (e.g., visual pattern categorisation). Then, the same distinction can be made for episodic memory tasks (e.g., remembering famous and unknown faces). Comparing these four tasks would allow a quantification of the contribution of the nature of the first-order task and the contribution of the type of material in metacognitive efficiency formation.

### **Using the meta-d framework in memory and perception tasks**

Another explanation for the higher metamemory efficiency compared to metaperception efficiency pertains to the quantification of metacognition. Theoretically, Type 1 performance constraints the Type 2 ROC curve by providing an upper bound as it refers to the maximum available signal to perform the second-order task. Thus, meta- $d'$  is never supposed to be higher than  $d'$ . However, hyper-metacognitive sensitivity (meta- $d' > d'$ ) has been shown to be relatively frequent. [Fleming and Daw \(2017\)](#) suggested that is likely due to “additional valid information from the state world” (p.100). Here, we propose that this information actually pertains to metacognitive beliefs that have been poorly integrated in model of confidence (see Section 2.3.2. “The behavioural dimension”). If metamemory is more information-based as proposed in the last section, it is possible that a framework that quantifies metacognitive sensitivity from the signal available in the first-order task is not the most efficient. The same conclusion can be drawn for prospective judgements as they also are largely based on information-based metacognition.

According to the dual-process theory of recognition memory, whilst the familiarity process is accommodated by classical SDT, recollection is thought as an all-or-none process that cannot be modelled using SDT (e.g., Yonelinas, 1994; Yonelinas, 2002). This is a major difference with visual perception that we can directly model in terms of SDT (e.g., [Macmillan & Creelman, 2004](#)). This is probably strengthened by the fact that visual perception uncertainty has an “objective” value within stimuli characteristics (visual noise such as blur) that is less present in memory as memory decisions are

not related to a stimulus in the environment but to the outputs of a cognitive process. Nonetheless, there is a long tradition of using SDT to model the familiarity process (see [Wixted, 2007](#) for a review). An interesting finding in Chapter 4 goes toward the idea that recollection might be problematic from a SDT point of view. Semantic memory recognition known to be mainly based of familiarity process (according to the fractionation of declarative memory proposed by [Tulving \(1985\)](#); however, see [Kempnich, Urquhart, O'connor, & Moulin, 2017](#) for evidence of a threshold retrieval process in semantic memory) has the highest correlation with the visual perception task which can be due to the fact that first-order performance in these tasks both fit with an SDT model. We suggest that a quantification of metacognitive sensitivity using another framework might allow a better comparison between episodic memory and visual perception. The accumulation of evidence can therefore be a good candidate as it rather focuses on the modelling of RT. Overall, even if the meta-d' framework provides an interesting framework for modelling the link between confidence and first-order accuracy remains an open question ([Van den Berg et al., 2017](#)) and has been always been debated in the field of metacognition (as shown in Chapter 2).

### 7.2.3 Implications for dimensions of metacognition

Chapter 1 defined metacognition according to the three dimensions developed by [Fleming et al. \(2012\)](#). Here we aim to add information to these dimensions according to our proposal of metacognitive judgement formation.

#### The representational dimension

In metacognitive judgement formation, using explicit knowledge and beliefs suggests that people have theories about their own functioning. In this sense, they present to themselves their cognitive activity (akin to the meta-level). On the contrary, the use of implicit knowledge and heuristics do not necessarily need a representational dimension by directly using information stemming from the object-level, such as answer fluency. These heuristics allow a mean of uncertainty estimation. As we have suggested, prospective judgements are more underpinned by information-based metacognition than has likely a representational dimension. Because early model were based on memory functioning, they had a bias on representation dimension because prospective judgements are more pertinent to metacognition in memory tasks, hence their prominence in the model of [Nelson and Narens \(1994\)](#). Moreover, we suggest that their model mainly

propose to explain metacognitive beliefs: the meta-level actually refers to metacognitive knowledge (as originally defined by Flavell (1979)). We even propose that metacognitive experiences (as also defined by Flavell, 1979) also pertain to metacognitive knowledge and information-based metacognition rather than experience-based metacognition. Let us consider the FOK in daily life. The “state” of FOK actually refers to a conflict between a failure to retrieve (which is a first-order process) and the metacognitive belief that we should be able to retrieve this information. This is also the case for other metacognitive experiences such as TOT (e.g., Brown & McNeill, 1966) and déjà-vu experiences (e.g., O’Connor & Moulin, 2010).

### **The behavioural dimension**

On the contrary, metacognitive judgments can be based on information directly stemming from the first-order task which pertains to experience-based metacognition (as the experience of the task). As for the previous section, we suggest that models of confidence have particularly focus on this type of information (e.g., evidence or signal driven by the task) because it is the main source of RCJs formation. As such, models of confidence are thus an extension of first-order models: for instance accumulation models accounting for decision making (mainly in visual perception) have been extended to explain dissociations between metacognition and task performance (see Chapter 2). From a second-order model of confidence point of view, confidence is computed “together with the observed action  $a$  and knowledge of the covariance between [the decision variable] and [the confidence variable]” (p.94; Fleming & Daw, 2017). We have proposed in Chapter 1 that metacognitive behaviours are underpinned by processes involving knowledge either represented explicitly or implicitly. We also have suggested that such implicit knowledge pertains to experience-based metacognition and heuristics such as answer fluency. The motor action, when performed by the participant, gives additional information about the speed of response of the first-order task answer which refers to the correlation between the decision variable and the confidence variable. As suggested above, when this information is disrupted or non-available, participants are less able to rely on the diagnostic answer fluency heuristic to discriminate between correct and incorrect responses. Even if we propose that RCJs are mainly underpinned by such processes, explicit knowledge<sup>2</sup> can also be used. However, models of confidence such as the second-order model of confidence and post-decisional models have “not considered

---

<sup>2</sup>Note that in these two cases, knowledge is used as a broad term, as it can be misleading (i.e., inaccurate) in some situations and pertains more to the notion of beliefs.

the role of learning or prior beliefs about the task structure” (p.106; Fleming & Daw, 2017) that pertain to information-based metacognition. Thus, future work should focus on this issue, modelling the influence of beliefs on metacognitive judgements. Models of confidence regardless of the framework they have been developed (e.g., SDT, accumulation of evidence) do not currently account for these top-down influences on metacognitive judgement formation. Here we propose that even if bottom-up experience-based metacognition is one basis of such judgements (especially for RCJs), top-down beliefs - information based metacognition - is also at play.

In Chapter 1, we also proposed the existence of other metacognitive behaviours (mainly be underpinned by implicit knowledge). These behaviours can be measured by indirect tests and a remaining open question is weather these metacognitive behaviours are also related to metacognitive judgment accuracy. As we proposed that RCJs are mainly based on experience-based metacognition, RCJs accuracy would be likely more related to performance on indirect tests of metacognition compared to prospective judgements. This idea echoes findings in pathologies. For instance, in Alzheimer’s disease patients have deficits on some prospective judgements<sup>3</sup> (e.g., eFOK; Souchay et al., 2002) despite other correct metacognitive behaviours (e.g., time allocation of re-studying item, Moulin, Perfect, & Jones, 2000a. This dissociative pattern have been explained by a preservation of an “implicit awareness” (Mograbi & Morris, 2013) that allow patients to (for further implications for neuropsychology and psychiatry see Section 3.2).

### **The consciousness dimension**

As briefly introduced in Chapter 1, metacognition has a consciousness dimension despite being dissociable from it. Some manifestations of conscious awareness occur without a metacognitive value (e.g., the phenomenology of consciousness) and metacognitive behaviour does not necessarily involve consciousness (e.g., unconscious cognitive control). In that sense, metacognition refers to consciousness in terms of the “access of consciousness”, that is the ability to be aware of conscious states as knowledge. However, metacognitive judgements and particularly RCJs are now extensively used to measure phenomenal consciousness and are seen as subjective measures of consciousness (e.g., Norman & Price, 2015; Rosenthal, 2019; Sandberg et al., 2010; Seth et al., 2008)

---

<sup>3</sup>These deficits are actually a complex pattern resulting in a fractionation of metacognition in this pathology.

with the underlying assumptions that any conscious process should be reportable (in conditions when it is possible (Rosenthal, 2019)).

This thesis did not aim to directly address the concept of phenomenal consciousness, however we want to stress here that there is a tendency to evaluate consciousness in visual perception (perhaps because it is associated with a large phenomenological richness) and we propose that memory is another good candidate for this purpose. As such, FOK, the TOT phenomenon, the *deja-vu* experience, and any other metamemory states where there is a dissociation between first-order processes (e.g., a failure to retrieve a word) and a second-order thought (e.g., being sure of knowing this word), are also associated with a large subjective richness. This would be particularly interesting in the context of the higher-order-thoughts theory of consciousness (Rosenthal, 2000) which proposes that a state is conscious only if one is aware of being in that state. Therefore this ability relates to the individual rather than a in any signal or state from the environment as a notion that is close to our conceptualisation of information-based metacognition.

Some aspects of consciousness research provide interesting windows into metacognitive research. A leading theory of phenomenal consciousness namely the Global Workspace Theory (e.g., Baars, 2005; Dehaene, Sergent, & Changeux, 2003) proposes that consciousness makes available information to different cognitive processes. Such information is available when it is in a “workspace” which is thought to be “global” (i.e., available across cognitive/neural processes). The notion of globalism is interesting here as it proposes as the access of the global workspace occurs after the process has been taken place in domain-specific systems. Shea and Frith (2019) propose integrating metacognition (i.e., low level metacognition from our point of view) into the global workspace. Each state is accompanied by a sense of certainty (how likely this state is correct) that do not imply another level of representation (on the contrary higher-order-thoughts theory of consciousness). These authors suggest that the manipulation of representations is made using metacognitive parameters that weight information. Thus, these parameters stemming from actual states/percepts/outputs would be used regardless of the cognitive domain (as our proposal of experience-based metacognition).

In sum, the representation dimension of metacognition has been mainly proposed in classical metamemory models by explicit metacognitive knowledge and beliefs, whereas models of confidence formation rather focused on evidence driven by the first-order task and implicit metacognitive knowledge. If metacognition has been linked to consciousness, it remains that other forms of metacognitive behaviours do not necessarily

imply conscious awareness. Therefore, the relation between performance on direct and more indirect tests of metacognition should be examined in future research to have a more broad understanding of the diversity of metacognition. Overall, we propose more bridges across theories and fields. The metaperception field and the metamemory field have similarities and differences that are interesting and useful to draw out, as is likely the case for consciousness and metacognitive research<sup>4</sup>.

## 7.3 Other implications

Although the current work mainly aims to contribution to metacognitive research, it also proposes indirect implications for other field of the psychological research. As such, we will now focus on the field on recognition memory and more clinical research.

### 7.3.1 Implication for memory research

As mentioned in Chapter 2, a large amount of work in recognition memory has used confidence as a means of studying memory processes (e.g., DPSD model or the UVSD model; [Yonelinas, 1994](#); [Yonelinas, 2002](#)). Research to date has extensively focused on confidence judgements and ROC curves in order to discriminate between models. Confidence is used to infer trace strength or the involvement of different memory processes and thus it is seen as a first-order decisional process instead of a second order process. As highlighted in Chapter 2, even though metacognition is based on first-order evidence, it can also dissociate from task performance. Thus, post-decisional accumulation models and second-order models of confidence support the involvement of both a representation of sensory evidence during the decision task and additional factors. From this perspective, the use of ROC curves to infer memory trace strength is limited as patterns of results would be explained by both first-order and second-order processes. This is particularly important as we have suggested here that metamemory beliefs are more higher-order beliefs (at least in comparison with those related to visual perception) and that these beliefs play a greater role in the formation of confidence in memory compared to confidence in visual perception.

Moreover, metamemory beliefs have been shown to control the memory activity as they are important in regulating attempts to retrieve information. Participants are more likely to engage in a search if they believe they can recall the information ([Nelson](#)

---

<sup>4</sup>Here we just very briefly considered similarities and differences as it is beyond the scope of this thesis.

& Narens, 1994). As proposed in Chapter 4, a positive feedback loop might ensue in which good metacognitive sensitivity is used to guide memory search which may in turn further increase metacognitive sensitivity. Thus, memory and metamemory are linked in essence and models of memory should consider the role of metacognition as an important component in recognition memory.

Although the present studies do not directly bring additional information for model-comparison in memory, we wanted to highlight (as also suggested in Chapter 2) that a large amount of work related to confidence from of the recognition memory field do not considers confidence as a metacognitive process per se. This means that confidence cannot be (at least completely) used to infer first-order processes as additional information can selectively influence confidence (see section on dissociations between metacognition and first-order performance in Chapter 2). Our proposal of metacognitive judgement formation adds another consideration. Whilst metamemory confidence is mainly based on experience-based metacognition (through the fluency heuristic mainly driven by RT in the first-order task), we propose that metamemory beliefs (information-based metacognition) are more prominent than other metacognitive beliefs (e.g., metaperception beliefs). Thus, these beliefs cannot be directly attributed to memory processes themselves. Research should therefore consider this as an important part of confidence formation in memory and we suggest that recognition memory models should incorporate a metacognitive component when using confidence to infer first-order processes. Thus, it is possible that confidence can be computed from first-order performance for to familiarity (or for an unidimensional trace strength) but according to dual process theories, this might be less the case for recollection. We can even suggest that recollection intrinsically pertains to metacognitive beliefs: if more information is retrieved one can explicitly think (and be sure) that this situation actually occurred. Thus, applying accurate domain-specific metacognitive beliefs in a situation could differentiate between recollection-based recognition and familiarity-based recognition<sup>5</sup>.

### 7.3.2 Implication for neuropsychology and psychiatry

Because metacognitive judgements are self-evaluations of a cognitive activity, they offer a framework by which we can measure disease and symptom awareness in patients. The core idea is that patients that are unaware of their cognitive difficulty should be inaccurate on metacognitive tasks (on metacognitive bias and/or metacognitive sensitivity).

---

<sup>5</sup>Nonetheless, note that recollection and beliefs can be dissociated in the case of recall, e.g., non-believed memory, [Mazzoni, Scoboria, and Harvey \(2010\)](#)



In that sense, it is close to the clinical notion of anosognosia (see Appendix A for an introduction to the key concepts).

A population of that has been particularly studied within both frameworks of metacognition and anosognosia is Alzheimer's disease. In this pathology, it has been proposed that patients exhibit a failure to transfer online awareness of difficulties into metacognitive knowledge that are more generalised (i.e., mnemonic anosognosia, [Agnew & Morris, 1998](#); [Morris & Mograbi, 2013](#)). This failure results in a pattern of findings suggesting a fractionation of metamemory in this pathology ([Souchay, 2007](#)). Patients are able to assess their performance during a cognitive task in the here-and-now both prospectively and retrospectively (see Appendix C for an example in short-term memory) but initial predictions made before any experience with the task are always overestimated ([Ansell & Bucks, 2006](#); [Moulin, Perfect, & Jones, 2000b](#)). According to the view of metacognitive judgement formation we propose here, Alzheimer's disease patients would therefore have an inability to generalise online beliefs into more higher-level beliefs (or to update these beliefs). We have suggested that these beliefs are domain-specific, however from our point of view, the generalisation process occurs regardless of the type of cognitive domain. Nonetheless, we suggest that even if the integration/generalisation process is impaired, such impairment does not necessarily create domain-general metacognitive difficulties. On the contrary, the integration stems from online situation-specific beliefs that can be generated by task performance feedback. If first-order performance is unimpaired, online beliefs and higher-level beliefs remain in accordance. However, as soon as a change in cognitive functioning appears, a mismatch exists between online beliefs and higher-level beliefs. If generalisation processes are inefficient, this results in a metacognitive impairment specific to the impaired cognitive domain.

This is strengthened by studies showing that patients with metamemory impairment have often memory difficulties or a lower task performance than healthy controls<sup>6</sup> (e.g. Alzheimer's disease, [Souchay, 2007](#); schizophrenia, [Souchay et al., 2006](#); patients with frontal lobe lesions, [Schnyer et al., 2004](#); however see patients with autism spectrum disorder, [Wojcik et al., 2013](#)). Similarly, we have shown that patients with Multiple

---

<sup>6</sup>As mentioned in Chapter 2, it is worthy to note that lots of studies in the metamemory field rely on gamma correlation that has been shown to be biased by first-order performance. Therefore, for some work could have found metacognitive sensitivity deficits that do not reveal a genuine impairment due to a confound with first-order performance.

Sclerosis are inaccurate in predicting their performance in a task were they are impaired (i.e., processing speed, see Appendix B). In essence, we suggest that metacognitive impairments in patients relate to faulty belief-updating which can result in domain-specific metacognitive deficit in metacognitive judgements that are mainly underpinned by information-based metacognition (i.e., prospective judgements).

Regarding the domain-generalty of metacognition that we found using RCJs, we suggest that cross-task correlations for metacognitive efficiency are likely due to experience-based metacognition through the fluency heuristic (i.e., RT difference between correct and incorrect responses). As processes involved here are based on implicit knowledge, we propose that they are less likely to be impaired in patients as they are used automatically and do not need a strategic and voluntary use as it is probably the case for explicit knowledge. As such, very few studies have identified metacognitive sensitivity deficits in RCJs in patients (see [Pannu & Kaszniak, 2005](#) in metamemory with neurological population or [Hoven et al., 2019](#) in metadecision with psychiatric populations). Although some studies have found a lower metacognitive sensitivity in schizophrenia (e.g., [Moritz, Woodward, Jelinek, & Klinge, 2008](#)), it appears that when first-order performance is controlled, metacognitive efficiency is the same for controls and patients ([Favre et al., 2020](#)). Nonetheless, [Fleming et al. \(2014\)](#) have highlighted a low metacognitive efficiency specific to a visual perception task in patients with lesion in the anterior PFC despite a preserved metamemory efficiency. This interesting result is not directly consistent with our current proposal. Nonetheless, even if we have suggested that RCJs are mainly domain-general (underpinned by experience-based metacognition), domain-specific processes (underpinned by information-based metacognition) are also at play (as also highlighted by neural networks, [Morales et al., 2018](#); [Vaccaro & Fleming, 2018](#)). This study suggests a specific involvement of the anterior PFC in metacognitive efficiency for visual perception, however we should keep in mind that the neural correlates of metacognition and especially metacognitive sensitivity remains unclear ([Vaccaro & Fleming, 2018](#)). We suggest that further work should be dedicated in the investigation of such high-order representation of specific tasks.

Deficits that have been found in RCJs are mainly on metacognitive bias (see [Rouault, Seow, et al., 2018](#) for a study of psychiatric dimensions in healthy participants; see [Hoven et al., 2019](#) for a review with psychiatric populations) which we have proposed to be domain-general for a variety of first-order tasks and type of judgements (Chapter 4 and Chapter 5). Such deficits have been suggested to be due to domain-general beliefs about self-abilities ([Rouault, Seow, et al., 2018](#)) that could have been created by

chronic dyshomeostasis notably in the case of depression (Stephan et al., 2016). From our perspective, this relates to an over-generalisation of beliefs: extension of a situation-specific perceived failure to other situations and domains .

Overall, few studies have compared metacognitive judgements across domains in pathologies and predictions made here need to be tested in cross-task paradigms. Chapman et al. (2018) recently proposed investigating metacognition for memory and motor tasks (using eFOK and an agency task) in Alzheimer’s disease. Interestingly, a relationship between metamemory accuracy and accuracy in the agency task suggested potential common resources across these two tasks<sup>7</sup>. Future research should investigate how domain-specific and domain-general impairment in metacognition can occur in neurological and psychiatric populations.

Finally, we propose that, in patients, it is the integration of online low-level beliefs to higher-level beliefs that is defective. In some cases, beliefs are not generalised because there are accurate but they are rather over-generalised: situation-specific beliefs becomes too general (e.g., in case of depressive symptoms). In other cases, generalisation processes do not occur at all and a mismatch exists between online beliefs and higher-level beliefs resulting in a domain-specific metacognitive impairment (e.g., in case of Alzheimer’s disease). From a clinical point of view, recent studies have proposed that metacognition can be trained and even that domain-specific training can have a “transfer effect” on another domain (see Carpenter et al., 2019) which is in accordance with our results on domain-generality (Chapter 4). Thus, it opens an interesting window for patients with metacognitive and disease awareness impairment as such awareness has been shown to maximize the efficacy of cognitive rehabilitation techniques (Prigatano, 1999) or to predict conversion to dementia in mild cognitive impairment patients (for a review see Roberts, Clare, & Woods, 2009).

### 7.3.3 Conclusion

This thesis explored the breadth of metacognition asking whether people use a common resource in their metacognitive judgements across different types of first-order tasks and whether this resource is also shared across different metacognitive judgements. We investigated these questions using both correlational studies and experimental manipulations of processes potentially involved in metacognitive judgements formation

---

<sup>7</sup>Note that we do not see judgements of agency as a metacognitive judgement per se therefore we do not see this study as a genuine cross-task comparison (also because this study was carried out for a different purpose).

across domains. This thesis provides new findings for the in-depth understanding of metacognitive processes at play in second-order judgements. In particular, we found that metacognitive efficiency in RCJs can be supported by a domain-general resource which was not the case in prospective FOKs. Our studies investigating the role of fluency however revealed that fluency effects are less unified as previously proposed. Specifically, if perceptual fluency could influence the magnitude of prospective judgement in visual perception and memory, this is not the case for conceptual fluency. Nonetheless, consistent outcomes were found regarding differences across prospective and retrospective judgements suggesting that prospection using non-verbalisable material cannot occur.

Moreover, we proposed that although metacognitive judgements rely on dual processes, the amount of each process differs according to the type of judgement. Prospective judgements rely more on beliefs and higher-order representations whereas retrospective judgements are more based on heuristics driven by the first-order task. In essence, this thesis proposes having a more unified view of metacognitive research than has been mainly studied within field. When related to second-order evaluations, we believe that memory and perception have much more in common than previously thought and the models and methods developed in both field should benefit from each other.

Finally, our results have implications for both research on recognition memory and research on neuropsychology and psychiatry. Recognition memory research has a long tradition of using confidence to infer first-order processes. Here, we propose that models of memory should consider the role of metacognition as an important component in recognition memory, especially the involvement of beliefs in the formation of confidence. Regarding clinical work, metacognition offers an interesting viewpoint on disease awareness. We highlight similarities between models of anosognosia especially in Alzheimer's disease population and our proposal of metacognitive judgement formation. Furthermore, we suggest that identifying processes at play in such evaluation are crucial for patients and the efficacy of rehabilitation techniques, particularly in term of domain-generality.

## **Appendix A**

# **A review of metacognition and self-awareness in Multiple Sclerosis**



## Review

# Metacognition and self-awareness in Multiple Sclerosis

Audrey Mazancieux<sup>a,\*</sup>, Céline Souchay<sup>a</sup>, Olivier Casez<sup>b</sup> and Chris J.A. Moulin<sup>a</sup>

<sup>a</sup> LPNC CNRS 5105, Université Grenoble Alpes, Grenoble, France

<sup>b</sup> Department of Neurology, Université Grenoble Alpes, Grenoble, France

## ARTICLE INFO

## Article history:

Received 2 February 2018

Reviewed 16 May 2018

Revised 3 July 2018

Accepted 9 November 2018

Action editor Robin Morris

Published online 19 November 2018

## Keywords:

Multiple Sclerosis

Metacognition

Self-awareness

Metamemory

Anosognosia

## ABSTRACT

Although a large range of literature on awareness and metacognition focuses on different neurological populations, little attention has been paid to Multiple Sclerosis (MS). This paper gathers literatures related to studies of anosognosia and the theoretical construct of metacognition which both offer a means to operationalize and measure awareness in MS. We focused on both a clinical concern, regarding the relationship between subjective and objective evolution of cognitive performance, and the theoretical issue of metacognitive processes implicated in disease awareness. We identified 26 papers with findings related to awareness of cognitive impairment in MS using questionnaire-based or performance-based methods. We found support for the idea that the relationship between subjective evaluation and neuropsychological evaluation depends on disease duration and is strongly modulated by other variables, such as mood state. We propose that the metacognitive deficit for memory tasks in this population arises from memory impairment. Finally, we discuss methodological issues, variability in MS patients, and the domain specificity of metacognitive impairment.

© 2018 Elsevier Ltd. All rights reserved.

## 1. Introduction

Research into deficit awareness in neuropsychological diseases has shown much development over the last few decades. Understanding awareness is particularly important to alleviate the effect of symptoms, especially cognitive deficits. For instance, in cognitive rehabilitation, in order to maximize the efficacy of cognitive rehabilitation techniques, awareness of deficit is crucial (Prigatano, 1999). Unawareness of deficit has

also been linked to disturbances in understanding of the impact of cognitive disabilities on activities of daily living (McGlynn & Schacter, 1989). A second motivation for researching awareness is to better understand the disease process in neurological disorders: awareness (or lack of it) may be prognostic. For instance, evidence has been found for low awareness as a predictor of conversion to dementia in mild cognitive impairment patients (for a review see Roberts, Clare, & Woods, 2009) independently from others factors (Gerretsen et al., 2017). Moreover, a lack of awareness may contribute to

\* Corresponding author. 1251 Avenue Centrale, St Martin d'Hères, 38040 Grenoble, France.

E-mail address: [audrey.mazancieux@univ-grenoble-alpes.fr](mailto:audrey.mazancieux@univ-grenoble-alpes.fr) (A. Mazancieux).

<https://doi.org/10.1016/j.cortex.2018.11.012>

0010-9452/© 2018 Elsevier Ltd. All rights reserved.

cognitive deficits by altering the efficient allocation of cognitive resources – people who lack awareness will not spontaneously compensate for their cognitive difficulties.

The focus of this review is awareness of deficits and metacognition in Multiple Sclerosis (MS). We aim to define more precisely this concept in this pathology, and critically to determine if MS patients suffer from a disturbance in awareness processes. To do so, we will be concentrating in the classical distinction between metacognitive *knowledge* and metacognitive *experiences* (Flavell, 1979). Metacognitive knowledge refers to a global assessment of cognitive skills and beliefs about functioning whereas metacognitive experiences refer to on-line awareness during task achievement which allows the assessment of the level of expertise of this task. We will also present two parallel literatures (studies of anosognosia in patient groups, and metacognition) which offer a means to operationalize and measure awareness in patient groups. There is considerably variability in the use of terms across the different literatures, so in this introduction we will specify how we have operationalized each type of awareness in terms of measures and the theoretical model, focusing on the distinction between metacognitive knowledge and metacognitive experience.

MS is an autoimmune inflammatory disease affecting the central nervous system. It is characterized by inflammatory lesions which compromise neuronal conduction, resulting in a range of symptoms, including physical and cognitive problems (e.g., McDonald & Compston, 2006). It is now well-documented that cognitive impairment is frequent in MS and its prevalence rate ranges from 43% to 70% (Chiaravalloti & DeLuca, 2008). Rao, Leo, Bernardin, and Unverzagt (1991) suggest that the severity of cognitive impairment is related to the form of MS, disease duration and depends on normative data used to detect cognitive disabilities. Eighty percent of MS patients have the relapsing-remitting (RR) form which refers to new symptoms resulting from inflammation and occurring in isolated attacks or relapses. Fifty percent of the time, RR patients will shift to a progressive form (secondary progressive – SP form) due to neurodegenerative evolution (Thompson et al., 1991). A latest form called primary progressive (PP) refers to a worsening over time without relapse and it is characterized by more cognitive symptoms (Planche, Gibelin, Cregut, Pereira, & Clavelou, 2016). In turn, these factors are related to disease duration and age, since there is a worsening cognitive impairment with time and progressive forms are encountered later in life than the relapsing type.

Because lesions are distributed, there can be multiple cognitive impairments in MS however a typical profile often emerges. The most impaired functions are related to executive functioning which gather planning, flexibility, inhibition, and working memory; broadly speaking, adaptation to new or complex situations (e.g., Rabbitt, 2004). Executive functions partly depend on information processing speed which is a more transversal function referring to the speed of task achievement (for a discussion of speed of processing in MS see Costa, Genova, DeLuca, & Chiaravalloti, 2017). Slowed processing seems to be the earliest and most common impairment in MS and could explain another frequent deficit which is long-term memory (Chiaravalloti & DeLuca, 2008) even if this hypothesis is still under debate.

MS patients are not typically thought of as anosognosic per se, however because this concept refers to the core question of awareness, we briefly define it here. Anosognosia was firstly used for a lack of awareness of deficit in neurological disease, especially hemiplegia (Babinski, 1914), but this term is now being extended to neuropsychological and psychiatric disorders (Prigatano & Schacter, 1991). Several terms co-exist in the literature such as *unawareness of deficit*, *denial of deficit*, *self-awareness*, and *lack of insight*, which all refer to awareness of disabilities, because these terms are mainly used in the context of illness.

Because of the complexity of anosognosia, the effectiveness of the different methods used to assess awareness is often debated (Cocchini & Della Sala, 2010). Broadly speaking, three assessment approaches exist. First, from a medical perspective, the relationship between patients' subjective complaints and neuropsychological evaluation is important in order to inform diagnosis, and to understand the impact of the disease on quality of life and activities of daily living. This approach considers whether subjective complaints (typically measured by patient-completed questionnaires) are related to cognitive abilities as measured on neuropsychological tests. Second, from a neuropsychological perspective, the assessment of self-awareness consists of the comparison of patients' and informants' ratings of abilities of daily living. The level of agreement between the two questionnaires provides information about awareness of deficit. Thirdly, unawareness of deficit can be measured by clinician ratings. As these evaluations of self-awareness are not related to a current cognitive task, we suggest that they measure metacognitive knowledge rather than metacognitive experiences.

These methods assess either the global perceived-abilities of the patients, or focus on specific cognitive domains such as motor abilities, memory or executive functioning. Such a comparison between global or domain specific awareness is motivated by the first theoretical model of awareness, the Dissociation and Interaction and Conscious Experience (DICE; McGlynn & Schacter, 1989) model. This model postulates that a *conscious awareness system* (CAS) allows conscious experiences. Whereas this system can be selectively disconnected from a module leading to impaired awareness for a specific domain, a disconnection between the CAS and the *executive system* is supposed to impair awareness across several domains. From this perspective, different neurological damage leads to different kinds of anosognosia. This idea, which has been reported in neuropsychological studies, was incorporated into a recent more complex model: the *cognitive awareness model* (CAM; Agnew & Morris, 1998; Morris & Hannesdottir, 2004; Morris & Mograbi, 2013; see Fig. 1). This model predicts three types of anosognosia depending on lesion localization: mnemonic anosognosia, executive anosognosia and primary anosognosia.

In parallel to the anosognosia literature, studies from a cognitive psychology viewpoint – mostly concerned with healthy populations – focus on the notion of metacognition. Metacognition as a research topic first derived from studies of child development, and was defined by Flavell (1979) as *cognition about cognition*; knowledge of cognitive abilities and their regulation. In short, the metacognitive research domain focuses on those cognitive processes which allow us to reflect

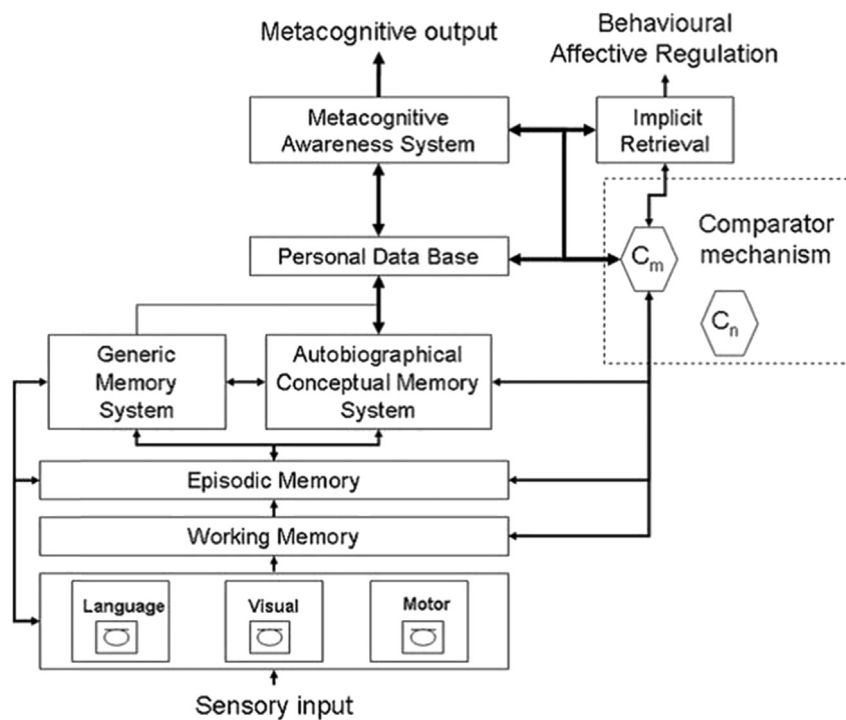


Fig. 1 – The Cognitive Awareness Model (Morris and Mograbi, 2013).

upon and estimate our cognitive capacities. As such, it offers a framework by which we can measure awareness in patient groups (see Souchay, 2007). Here we will just run through a few key concepts. The core idea is that if patients with MS are unaware of the cognitive difficulties this will be detected in their performance on metacognitive tasks.

The metacognitive framework proposed by Nelson and Narens' (1990) defines in a current task (for metacognitive experiences) control and monitoring processes. In a memory task, one can notice in a word-list that is difficult to learn. This self-assessment refers to monitoring processes. In turn, one can decide to allocate more time to study this item. This regulation of cognitive activity and change in strategies refers to control processes. In general, the study of the metacognition of memory processes (metamemory) is much better developed than in other domains but the metacognitive approach is now being extended to other cognitive processes (e.g., visual perception, De Gardelle & Mamassian, 2014; decision making, Yeung & Summerfield, 2012; and motor tasks, Simon & Bjork, 2001).

In metacognition, the main interest is the notion of judgment accuracy which is a measure of a person's awareness of their cognitive abilities. A distinction is made between two classes of measures: global predictions and item-by-item predictions (e.g., Mariné & Huet, 1998). Global predictions are used pre- and post-task, and these two different time points measure different components of metamemory. For instance, pre-task predictions for memory tasks (as an example) ask about the number of items people think they will later recall from a list. As the task has not yet commenced, this measure taps into metacognitive knowledge (Hertzog, 1992), rather than metacognitive experience. A post-task global prediction

is exactly the same prediction, but made after someone has learned the list. This measure draws more directly upon a monitoring process: that is, the extent to which a person updates his or her knowledge about memory abilities according to their experience of the task (Connor, Dunlosky, & Hertzog, 1997). In that sense, these evaluations (except from pre-task global predictions) can be thought of as measuring metacognitive experiences.

Although the metacognitive framework and the study of anosognosia were developed independently, there are similarities between models and constructs. For instance, Toglia and Kirk's model (2000) of awareness divides it into two processes: metacognitive knowledge and online awareness. Their metacognitive knowledge is close to Flavell's concept and online awareness can be thought of as metacognitive experiences (Flavell, 1979). This separation is useful because it is consistent with studies stemming from the anosognosia framework (mainly measuring generalized beliefs – i.e., metacognitive knowledge) and the metacognitive framework (mainly measuring metacognitive experience – i.e., 'online' feelings and evaluations of processes as they are occurring). Recent research focuses on the integration of both constructs. These studies explain results which come from studies on metacognition, especially metamemory according to the most recent model of anosognosia – the CAM model – in the context of dementia (Sunderaraman & Cosentino, 2017). Here, we focused on the distinction between metacognitive knowledge and metacognitive experiences as this distinction allows including both studies referring to anosognosia field and studies from metacognitive framework in the MS literature as has been done in others pathologies (Ernst, Moulin, Souchay, Mograbi, & Morris, 2016).



The focus of this review is not the neuroanatomy of awareness or executive functions, but here it needs to be stressed that the main cognitive symptoms in MS are thought to be due to deficits in executive functioning probably due to a decline in processing speed (Drew, Tippett, Starkey, & Isler, 2008). Neuroanatomically, there is a leading role of the frontal lobe in the neural network related to awareness of deficit. Fleming and Dolan (2012) reviewed the neural basis of metacognition and concluded that it relies on the prefrontal cortex interacting with interior regions such as insula and cingulate cortex. There is considerable overlap between the symptomology of MS, and the concepts of executive functions and these deficits have often been linked to the frontal lobe in this pathology (e.g., Beatty, Goodkin, Beatty, & Monson, 1989; Calabrese, 2006; Chiaravalloti & DeLuca, 2008; Foong et al., 1997). Furthermore, it has been suggested that some facets of executive functions are by definition metacognitive (Fernandez-Duque, Baird, & Posner, 2000), and thus one might expect there to be metacognitive deficits in MS.

Converging on the idea of metacognition and awareness as executive, both the DICE and CAM models include an executive component. Whereas the DICE model supposes a link between the CAS and an executive system, the CAM suggests that frontal lobe damage may cause an executive anosognosia. Some studies have also highlighted links between models of metacognition and models of executive functioning. For instance, Fernandez-Duque et al. (2000) suggest that the meta level and the object level from Nelson and Narens (1994) model have similarities with the executive system and the schemas level from Norman and Shallice's (1986) supervisory attentional system account of executive processes.

Finally, neuropsychological evidence supports a link between executive functioning and metacognition. Several studies in neurodegenerative diseases such as Alzheimer's disease (for a review in memory, see Souchay, 2007) or in traumatic brain injury (e.g., Ciurli et al., 2010) show a positive relationship between metacognitive unawareness and impaired measures of executive function. Likewise, in the healthy aging process, there appears to be simultaneous and correlated declines in metacognition and executive functions (e.g., Souchay & Isingrini, 2004; Souchay, Isingrini, & Espagnet, 2000). Because the main impairment in MS is in executive functioning – or at least in tasks which involve executive functions (e.g., processing speed, memory retrieval, attention) and because these patients have impaired performance on tests of executive functions such as the Stroop task and Wisconsin Sorting Card Test (WCST; e.g., Arnett et al., 1994), phonemic fluency tests (e.g., Henry & Beatty, 2006) and Paced Auditory Serial Addition Test (PASAT; e.g., Rosti, Hämäläinen, Koivisto, & Hokkanen, 2007), we hypothesize that MS patients have also impaired awareness of cognitive functioning. As metacognitive knowledge (through the anosognosia framework) and metacognitive experiences (through the metacognitive framework) have been related to executive functioning, we would expect that MS patients are impaired on both of them. However, we also have to imagine that we will also find individual variability since MS is a highly variable disease.

This systematic review thus aims to elucidate the level of awareness of MS patients according to the distinction between metacognitive knowledge and metacognitive experiences. More specifically, we are focusing on the following questions:

- (1) Regarding metacognitive knowledge, do MS patients make reliable subjective evaluations of their neuropsychological performance?
- (2) Regarding metacognitive experiences, do MS patients have problems in monitoring their cognitive functions?
- (3) What variables are associated with metacognitive knowledge and metacognitive experiences in MS?

A broader review of metamemory in several different neuropsychological populations has already considered MS, citing two articles on the topic. This review (Pannu & Kaszniak, 2005) aimed to understand different outcomes of metamemory impairment focused on metacognitive experiences in different neuropsychological populations such as Parkinson's disease, Alzheimer's disease, MS and traumatic brain injury. Here, we aimed to review studies that evaluate metacognitive experiences for memory as well as other cognitive functions and also studies referring to metacognitive knowledge.

## 2. Method

The flow chart describes the selection process (Fig. 2). We initially identified articles using ScienceDirect and Pubmed databases with the search strings: “multiple sclerosis” and “metacognition”, “metamemory”, “self-awareness”, and “anosognosia”. There was no limit on the year of publication. After screening the abstracts, twenty-five papers were assessed for eligibility. Sixteen more papers were identified as eligible from the citations of these 25 papers. From these 25 papers, 12 were included. From the 16 additional papers, 2 were excluded. Therefore, 26 papers were included in the current review (12 + 14). Our inclusion criteria were the following: (1) one of the measures used in the studies must be a metacognitive measure which is a subjective evaluation of performance (i.e., a validated questionnaire or performance predictions related to a task); (2) these metacognitive measures must concern cognitive abilities or cognitive functioning (e.g., episodic or semantic memory, planning, flexibility, inhibition ...); (3) these measures must be compared to another measure of cognitive functioning such as a neuropsychological evaluation or an experimental measure in the case of metacognitive predictions (see the metacognitive framework above). A summary of the included papers is presented in Table 1.

We organized the discussion of the published research around the distinction of metacognitive knowledge and metacognitive experiences since these refer to the different constructs of awareness as described above (Flavell, 1979; Togliola & Kirk, 2000). Already, a clear imbalance can be observed: most of the papers (22 out of 26) focused on metacognitive knowledge, and the remaining four were ranked as measuring

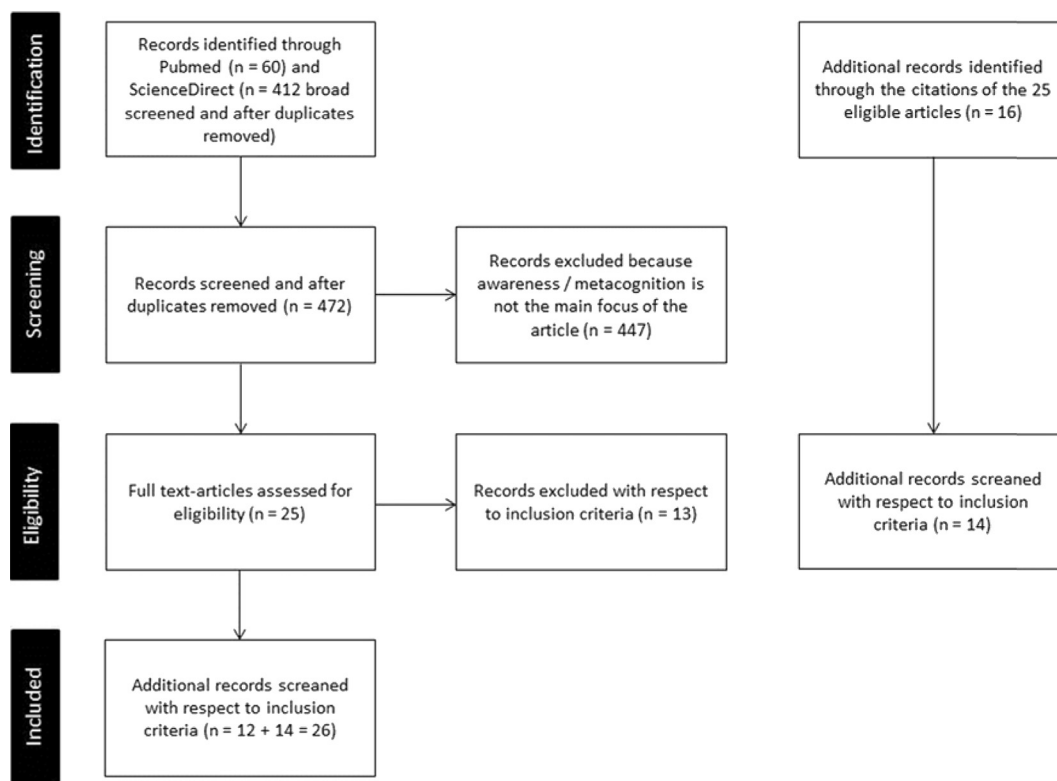


Fig. 2 – Flow chart describing the selection process according to PRISMA 2009.

metacognitive experiences. Table 1 summarizes the methods and main results.

### 3. Results

#### 3.1. Measures of metacognitive knowledge

Metacognitive knowledge was measured by either questionnaires of awareness or experimental measures (i.e., performance predictions). Within the use of questionnaires in order to assess metacognitive knowledge, several methods exist in the literature. These are correlations between neuropsychological evaluation and subjective evaluation (measured by self-rated questionnaires), or patient–informant discrepancies in reported impairments. Some studies compared cohorts of patients who were classified into accurate, underestimating and overestimating groups according to patient–informant discrepancies.

##### 3.1.1. Correlations between neuropsychological and subjective evaluation

Fifteen studies assessing awareness in MS patients by a correlation between subjective evaluations and neuropsychological assessment. These studies focused on executive functioning (verbal fluency, planning, attention, inhibition) and memory evaluation.

The correlations in MS patients are weak for global cognitive functioning (Maor, Olmer, & Mozes, 2001), memory (Randolph, Arnett, & Freske, 2004) and were marginally

significant for processing speed [Stroop ( $r = -.27, p = .016$ ) and SDMT ( $r = -.28, p = .014$ ); Roberg, Bruce, Lovelace, & Lynch, 2012]. Julian, Merluzzi, and Mohr (2007) observed that a neuropsychological index, featuring several tests of executive functioning explained only 8% of the variance of a subjective cognitive index. Additionally, Christodoulou et al. (2005) found no significant correlation between subjective evaluation and neuropsychological testing. Interestingly, however their patients were tested twice and the improvement in neuropsychological test scores was associated with a decrease in subjective complaint. Despite the lack of a correlation, we interpret this as MS patients being aware of their performance changing – a finding which possibly points to intact metacognition – discussed further below.

Other studies in our set show that people with MS are aware of cognitive impairment because they performed worse on neuropsychological testing and also reported more subjective deficits than healthy controls for both global cognitive functioning (Basso et al., 2008; Matotek, Saling, Gates, & Sedal, 2001) and memory (Kujala, Portin, & Ruutiainen, 1996). Hoogervorst et al. (2001) reported a correlation between subjective ratings and neuropsychological evaluations and these correlations were larger for physical disabilities. This suggests that MS patients are aware of their impairments, especially physical difficulties.

Randolph, Arnett, and Higginson (2001) measured self-reported day-to-day memory difficulties – they also assessed these difficulties from an informant's point of view, but did not calculate a discrepancy score as described below. Their results revealed that poorer perceived memory was explained by

**Table 1 – Selected papers dealing with metacognition in MS.**

N°	Référence	Method of assessing awareness	Awareness questionnaire	Objective evaluation	Participants	Results
1	<a href="#">Basso et al. (2008)</a>	Comparison between self-rating and objective testing	FrSBE	TMT, CVLT, WAIS-III letter number sequencing, Digit span, SDMT, COWAT	42 MS + 13 HC	FrSBE accounted for variance across, TMT, Letter Number Sequencing, SDMT and CVLT-II for both patients and controls
2	<a href="#">Beatty and Monson (1991)</a>	Episodic and semantic FOK + global prediction + awareness questionnaire	3 parts questionnaire (friend comparison, everyday memory problems, variable that affect memory)	4 groups of MS according to level in CVLT-II and WCST	45 MS + 22 HC	No difference for awareness questionnaire + Impaired episodic FOK and preserved semantic FOK + impaired global prediction only for the most with low memory and low WCST
3	<a href="#">Bruce et al. (2010)</a>	Comparison between self-rating and objective testing	Prospective and Retrospective Memory Questionnaire	SDMT, WCST, Stroop Test, AVLT, letter-number sequencing	71 MS + 20 HC	Dissociation mediates the relationship between depression, anxiety and self-reported memory complaints.
4	<a href="#">Carone et al. (2005)</a>	Discrepancy between self- and informant-rating + 3 accuracy groups	MSNQ	COWAT, Judgment of Line Orientation test, CVLT-II, BVMT-R, PASAT, SDMT, WCST	122 MS + 37 HC	No difference in objective testing between groups. Overestimation is associated with less depression and more cognitive impairment. Underestimation is associated with more depression.
5	<a href="#">Christodoulou et al. (2005)</a>	Comparison between self-rating and objective testing	PDQ	Brief Repeatable Battery (BRB): SRT, 10/36, SDMT, PASAT, COWAT, TOH	53 MS	No correlation between subjective evaluation and objective performance. Improvement in a second testing is related to lower subjective complains. No link between depression and subjective reports of cognitive impairment.
6	<a href="#">Goverover et al. (2005)</a>	Discrepancy between self- and informant-rating	FrSBE	COWAT, Verbal fluency, BNT, WART-3 reading subtest, WMS, PASAT, WCST	26 MS	Lower level of agreement for the most cognitively impaired patients for executive functions.
7	<a href="#">Goverover et al. (2014)</a>	Discrepancy between self- and informant-rating + global pre- and post-task judgment on experimental task	Functional Behavior Profile	Sorting Test, SDMT, CVLT-II	18 MS + 16 HC	Change in awareness for MS patients between pre- and post-diction (57% of patients were aware before, compared to 81% after) whereas no change was observed for healthy controls (70% against 75%) + lower degree of agreement for patients than controls
8	<a href="#">Hanssen et al. (2014)</a>	Comparison between self-rating and objective testing	BRIEF-A	COWAT, SDMT, Stroop Tests	120 MS	Marginal significance to relation between executive testing and subjective reports. Subjective impairment is related to depression.
9	<a href="#">Hoogervorst et al. (2001)</a>	Comparison between self-rating and objective testing	GNDS	MSFC	290 MS	Correlation between subjective evaluation and objective performance. Correlation for physical disabilities are stronger than those for cognitive abilities.
10	<a href="#">Julian et al. (2007)</a>	Comparison between self-rating and objective testing	Cognitive function subscale of MSQOL-54	SDMT, digit span, Stroop, AVLT, COWAT	58 MS	Objective testing explain 8% of the variance of subjective evaluation. Effective treatment for depression is associated with a stronger relationship between subjective evaluation and objective performance.

(continued on next page)

Table 1 – (continued)

N°	Référence	Method of assessing awareness	Awareness questionnaire	Objective evaluation	Participants	Results
11	<a href="#">Kinsinger et al. (2010)</a>	Comparison between self-rating and objective testing + 3 accuracy groups	PDQ	COWAT, Digit Span, letter-number sequencing, CVLT-II	127 MS	Improvement in accuracy estimation after psychotherapy which was associated with decreased depression and fatigue.
12	<a href="#">Krch et al. (2011)</a>	Comparison between self-rating and objective testing	MFQ	CVLT-II et Memory Assessment Scale	64 MS	Correlation between subjective evaluations and memory performance even after controlling for depression.
13	<a href="#">Kujala et al. (1996)</a>	Comparison between self-rating and objective testing	Self-evaluation of memory and learning problems (from 1 to 5)	WMS,7/24 spatial recall, 10 words learning, Incidental memory of word pairs, Digit Symbol-incidental learning NCSE	45 MS + 35 HC	The most impaired group of patient report more subjective impairment than the less impaired group and controls.
14	<a href="#">Maor et al. (2001)</a>	Comparison between self-rating and objective testing	MSQOL-54		161 MS	Weak correlation between subjective evaluation and objective performance for global cognitive functioning. Subjective impairment is related to depression.
15	<a href="#">Marrie et al. (2005)</a>	Comparison between self-rating and objective testing	PDQ	MSFC	136 MS	Slight decline in memory and processing speed are associated with higher subjective complaints than a major decline. Subjective impairment is related to depression.
16	<a href="#">Matotek et al. (2001)</a>	Comparison between self-rating and objective testing	Subjective difficulties questionnaire	WCST, Verbal intellectual quotient from WAIS-R, WMS	39 MS + 40 HC	Subjective evaluation correlate with verbal fluency and working memory testing. No correlation with anxiety and depression. No link between depression and subjective reports of cognitive impairment
17	<a href="#">Middletown et al. (2006)</a>	Discrepancy between self- and informant-rating + post-task judgment (trail-by-trial and global)	Cognitive Failure Questionnaire	PASAT, SRT, Word List Generation, CVLT-II, TOL	221 MS + 31 HC	No significant correlation perceived/objective abilities for either patients or controls + correlation trial-by-trial and global judgment with test performance
18	<a href="#">Phillips and Stuijbergen (2006)</a>	Comparison between self-rating and objective testing	Multifactorial Memory Questionnaire	ISS	482 MS	Poorer metamemory for the most depressed group.
19	<a href="#">Randolph et al. (2001)</a>	Comparison between self-rating and objective testing	Memory Rating Scale	TOL, CVLT-II, Rivermead Behavioral Memory Test, le 7/24, SDMT, Test of Everyday Attention	79 MS	Poorer perceived memory is explained by slower speed of processing and a lower level of education. No correlation between subjective and objective memory evaluation.
20	<a href="#">Randolph et al. (2004)</a>	Comparison between self-rating and objective testing	MFQ	TOH, letter-number sequencing, SRT	48 MS	Weak correlation between subjective evaluation and objective performance or memory function. Depressive beliefs mediate the relationship between depression, anxiety and self-reported memory complaints.
21	<a href="#">Roberg et al. (2012)</a>	Comparison between self-rating and objective testing	Processing Speed Difficulties Questionnaire	SDMT, Stroop, PASAT, letter-number sequencing, AVLT, COWAT, Conners' Continuous Performance Test II, MSFC	40 MS + 25 HC	Marginal significant correlation between subjective evaluation and objective performance for processing speed.

22	<a href="#">Rosti-Otajarvi et al. (2014)</a>	Comparison between self-rating and objective testing + 3 accuracy groups	MSNQ	Buschke Selective Reminding Test, 10/36, SDMT, PASAT, COWAT	196 MS	RR form: Overestimators are more cognitively impaired and more physically disabled. SP/PP form: Overestimators are more cognitively impaired than have less education.
23	<a href="#">Scarrabelotti and Carroll (1999)</a>	Prospective and retrospective JOL in stem word completion (inclusion and exclusion conditions)	–	–	50 MS + 41 HC	Impaired prospective JOL only for the exclusion condition. No difference for retrospective JOL.
24	<a href="#">Sherman et al. (2008)</a>	Comparison between self-rating and objective testing + 'aware' and 'unaware' groups	Awareness interview	Oral Symbol Digit Test, Brief test attention, judgment of Line orientation, letter-number sequencing, Stroop, COWAT, CVLT-II, WCST	74 MS	Unawareness of global cognitive abilities: 15% of RR patients and 51% of the SP patients. Unawareness is related to lower performance on tests
25	<a href="#">Smith and Arnett (2010)</a>	Discrepancy between self- and informant-rating + 3 accuracy groups	Dysexecutive Questionnaire	COWAT, Animal Naming, Stroop, reading span task	95 MS + 27 HC	No difference between the three groups according to age, disease duration, neuropsychological tests, and depression. Lower level of education for the overestimators.
26	<a href="#">Van der Hiele et al. (2012)</a>	Comparison between self-rating and objective testing + 3 accuracy groups	Dysexecutive Questionnaire	National Adult Reading test, CVLT, Rey's figure, TMT, Stroop, WCST, PASAT	128 MS	Underestimators were slower and showed higher interference on the Stroop Test and performed worse on the WCST + associated with more anxiety and depression.

AVLT: Auditory Verbal Learning Test, BRIEF-A: The Behavior Rating Inventory of Executive Function – Adult version, BNT: Boston Naming Test, BVMT-R: Brief Visual Spatial Memory Test-Revised, COWAT: Controlled Oral Word Association Task, CVLT-II: Californian Verbal Learning Test – version II, FrSBE: Frontal Systems Behavior Scale, ISS: Incapacity Status Scale, GNDS: General Neuropsychological Deficit Scale, MFQ: Memory Function Questionnaire, MSNQ: Multiple Sclerosis Neuropsychological Screening Questionnaire, MSQOL-54: Multiple Sclerosis Quality Of Life Questionnaire, NCSE: Neurobehavioral Cognitive Status Examination, PASAT: Paced Auditory Serial Addition Task, PDQ: Perceive Deficit Questionnaire, SDMT: Symbol Digit Modalities Test, SRT: Selective Reminding Test, TMT: Trail Making Test, TOH: Tower Of Hanoi, WAIS-R: Wechsler Adult Intelligence Scale–Revised, WART-3: Wide Range Achievement Test-3, WMS: Wechsler Memory Scale, WCST: Wisconsin Sorting Card Test. As in [Marrie et al. \(2005\)](#), 'objective' testing refers to neuropsychological evaluation.

slower speed of processing (17% of explained variance) and a lower level of education (15%) rather than the measure of memory impairment. Also, subjective evaluation and neuropsychological memory tests were not correlated with each other (which is a form of impairment of awareness – this will be discussed later). They also found that diminished awareness of memory difficulties was explained by impaired executive functioning, suggesting an association between impaired metamemory and executive functions. Finally, [Marrie, Chelune, Miller, and Cohen \(2005\)](#) suggested an interesting non-linear relationship between the perceived-abilities and the 'objective cognitive impairment' assessed by neuropsychological evaluation including memory and speed of processing. Slight decline in these two functions were associated with higher subjective complaints than a major decline.

Several studies also highlight a relationship between subjective cognitive impairment and the level of depression ([Hanssen, Beiske, Landrø, & Hessen, 2014](#); [Maor et al., 2001](#); [Marrie et al., 2005](#)). [Julian et al. \(2007\)](#) showed that an effective treatment for depression is associated with a stronger relationship between subjective evaluation and neuropsychological evaluation. According to this study, the more depressed MS patients are, the less aware they are. [Phillips and Stuijbergen \(2006\)](#) observed a lower score in a metamemory questionnaire [i.e., self-reported memory ratings about 'contentment, ability, and strategy' (p.429) in this study] for the most depressed group, revealing poorer metamemory in this group.

Two studies have focused on mediation analyses. [Bruce, Bruce, Hancock, and Lynch \(2010\)](#) concluded that dissociation (i.e., the disruption of usually integrated cognitive processes) mediates the relationship between depression, anxiety and self-reported memory complaints in MS whereas [Randolph et al.'s \(2004\)](#) found that depressive beliefs mediated this effect. However, other studies have failed to find a link between depression and subjective reports of cognitive impairments ([Christodoulou et al., 2005](#); [Matotek et al., 2001](#)). [Krch, Sumowski, DeLuca, and Chiaravalloti \(2011\)](#) found a significant correlation between subjective evaluations and memory performance even after controlling for depression.

### 3.1.2. Patient–informant discrepancies

A second method used to evaluate the awareness of cognitive abilities is to consider the level of agreement between patients' and informants' evaluations. The lower the level of agreement, the lower the awareness of patients. Unlike the correlation method described above, this allows the direct comparison of awareness of cognitive functioning without a neuropsychological evaluation. Using such discrepancy measures, [Goverover, Chiaravalloti, and DeLuca \(2005\)](#) found a lower level of agreement for the most cognitively impaired patients (as measured by neuropsychological testing of executive functioning). This result indicates lower awareness of these functions when they are impaired. A lower level of agreement was also related to higher anxiety and depression. Overall, this study concludes that there is an impaired awareness in the most impaired MS patients.

3.1.3. *Under and over-estimation in subgroups of MS patients*  
Another important question in the study of awareness is the assessment of the *magnitude* of the awareness of deficit. In this

way, it is possible to consider whether patients make predictions which over or under estimate their actual function. In the MS literature, six studies classified participants into "accurate", "underestimate" and "overestimate" groups. However, the method used changed from study to study. Some used a discrepancy score between patients' reports and informants' reports (e.g., [Carone, Benedict, Munschauer III, Fishman, & Weinstock-Guttman, 2005](#); [Smith & Arnett, 2010](#)) whereas others measured the accuracy by assessing the discrepancy between subjective evaluation and neuropsychological evaluation ([Kinsinger, Lattie, & Mohr, 2010](#); [Rosti-Otajarvi, Ruutiainen, Huhtala, & Hämäläinen, 2014](#); [Van der Hiele, Spliethoff-Kamminga, Ruimschotel, Middelkoop, & Visser, 2012](#)). For three studies ([Carone et al., 2005](#); [Kinsinger et al., 2010](#); [Smith & Arnett, 2010](#)), the evaluation of magnitude is based on a discrepancy between the t score or z score of the subjective questionnaire and/or the neuropsychological evaluation. Negative scores refer to better subjective cognitive functioning than neuropsychological evaluation. The classification criterion (i.e., cut-off criterion) was fixed by the authors. For instance, [Kinsinger et al. \(2010\)](#) estimated that 'patients were categorized as under-estimators of their cognitive abilities if their discrepancy scores were 1 SD above zero' (p.576). Other studies ([Rosti-Otajarvi et al., 2014](#); [Van der Hiele et al., 2012](#)) classified patients according to neuropsychological performance and subjective evaluation. Thus, accuracy was defined as 'objectively impaired and subjectively impaired' or 'objectively intact and subjectively intact'. Once again, the impairment criterion was fixed by the authors. This separation of patients according to their accuracy allows the comparison of underestimating, overestimating and accurate groups.

The distribution of patients in the five studies according to the three accuracy groups and MS type is summarized in [Table 2](#). We averaged the percentage of groups across studies except for [Rosti-Otajarvi et al. \(2014\)](#) which compared the RR form and SP/PP forms. According to our analysis, fifty-three percent of MS patients are accurate in their subjective reports. Furthermore, MS patients are more likely to underestimate their cognitive abilities than overestimate them (32% compared to 15% for the respective groups). Because these studies mainly included RR patients, the calculated percentages are very close to those related by [Rosti-Otajarvi et al. \(2014\)](#) for the RR group. However, it is important to notice the relatively large differences across studies. For instance, accurate estimators range from 33% ([Kinsinger et al., 2010](#)) to 69% ([Van der Hiele et al., 2012](#)) of the sample. This important variation probably depends on the classification, the definition of accuracy. Due to differences in the criteria used, it is difficult to compare these studies. However, if anything, these studies report more underestimation than overestimation in MS. If we combine the percentages of those who under or over estimate, it appears that approximately half of all patients are inaccurate, but approximately twice as many patients underestimate as overestimate their performance. This collection of studies finds that MS patients, if they lack awareness, tend to think that their cognitive function is more impaired than it actually is.

[Carone et al. \(2005\)](#) showed that overestimation of cognitive abilities in MS is associated with less depression, and more cognitive impairments (memory and executive functioning).

**Table 2 – Distribution of patients according to accuracy groups and MS type.**

Reference	MS type	Underestimate	Accurate	Overestimate	N
Carone et al. (2005)	RR, SP and PP	21%	61%	18%	122
Kinsinger et al. (2010)	RR, SP and PP	65%	33%	2%	124
Smith and Arnett (2010)	RR, SP and PP	27%	49%	24%	95
Van der Hiele et al. (2012)	unknown	16%	69%	15%	128
	Mean	32%	53%	15%	
Rosti-Otajarvi et al. (2014)	RR	27%	53%	20%	138
Rosti-Otajarvi et al. (2014)	PP and SP	7%	64%	28%	58

Conversely, underestimation is only related to increased depression. Although they did not compare directly their three accuracy groups, Kinsinger et al. (2010) observed an improvement in accuracy estimation after psychotherapy. This improvement was associated with decreased depression and fatigue. Smith and Arnett (2010) focused on executive functioning and found no difference between the accurate, underestimating, and overestimating groups according to age, disease duration, neuropsychological tests of executive function, and depression. The authors only observed a lower level of education for the over estimators compared to the accurate/underestimators. Conversely, Van der Hiele et al. (2012) found that underestimators were slower and showed higher interference than the accurate group on the Stroop Test and performed worse on the WCST (but note they did not compare under- and overestimators). In their study, underestimators also showed more anxiety and depression than accurate estimators and overestimators showed more problems in activities of daily living measured by self-reported questionnaire.

One study focused on the comparison between RR form and SP/PP forms of MS. Rosti-Otajarvi et al. (2014) showed that SP/PP patients underestimate significantly less their cognitive abilities compared to RR patients. Overestimators were more cognitively impaired and had more physical disabilities than underestimators for the RR group. Overestimators were also more cognitively impaired than underestimators and had less education than accurate estimators for the progressive group (SP/PP). A second study also compared RR form and SP/PP forms, however only two groups of awareness were created (i.e., aware and unaware) using an awareness interview and the bias/direction of the unawareness was not specified (Sherman, Rapport, & Ryan, 2008). These authors reported that 56% of the MS patients were unaware of their 'thinking abilities', 38% of their attention and 31% of their memory. Across MS forms, 15% of RR and 51% of the SP patients have unawareness of global cognitive abilities. This unawareness was related to lower performance on the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 2000) and several executive function tests such as Controlled Oral Word Association Test (COWAT, Spreen & Strauss, 1991), Stroop and the WCST.

Finally, in order to assess metacognitive knowledge, Beatty and Monson (1991) used a global prediction of performance in a memory (episodic free recall) task. This prediction was then compared to actual performance. The authors compared four patient groups and one control group. The results revealed that only the group with the worse deficit in memory and executive function overestimated their performance. The other four groups were accurate (i.e.,

difference between prediction and performance did not differ from zero).

Overall, across several different measures and in multiple studies, research into metacognitive knowledge in MS points to there being a metacognitive impairment in this disease. There is not a uniform deficit however, and inaccuracy is observed in about 50% of the samples that we could find in the literature. Where these studies have considered the magnitude of judgments, it looks like there is a tendency to underestimate performance, rather than overestimate it. This is interesting, since rather than a failure to adapt to changes in cognition (which would lead to overestimations of performance), this suggests that people with MS may actually be overly concerned with their cognitive changes, and believe themselves to be more impaired than they actually are. We will return to this issue in the discussion.

The comparison of under and over estimates (and the accurate participants) is interesting, but yields equivocal findings. Differences are observed between accuracy groups but the studies do not yield consistent results. They mainly suggest that metacognitive knowledge impairment is associated with depression, fatigue and a worse score on neuropsychological tests. No clear pattern emerges according to over- and underestimation except the fact that overestimation seems to be correlated with less education and seems to be less frequent than underestimation (as above). A not unexpected finding is that underestimation seems to be associated with more depression and anxiety – consistent with the idea that underestimates of performance reflect low self-esteem (e.g., Bandura, 1989).

### 3.2. Measuring metacognitive experiences

Few studies have focused on metacognitive experience or online awareness in MS. As mentioned above, Pannu and Kaszniak (2005) reviewed two studies dealing with metamemory. The first one is a study conducted by Beatty and Monson (1991) in which MS patients were divided into four groups based on performance on the CVLT and the WCST. This study suggested that MS patients perform differently across four metamemory tasks. No difference between patients and healthy controls was found for a semantic feeling-of-knowing (FOK) task, where participants have to predict the future recognition of non-recalled items. However, the mean discrepancy between global predictions and performance in an episodic free recall task was larger in patients than in controls, but only for the patient group with the highest level of cognitive impairments (i.e., the low memory, low WCST group). Accuracy on an episodic FOK task was lower than the

controls for all patient groups except for the “normal memory, normal WCST” group. Pannu & Krasnik's review also mentions Scarrabelotti and Carroll's study (1999) dealing with monitoring. In this study, participants had to make prospective judgments-of-learning (JOLs) about learning words that is they had to say how likely they will be to remember the item later. Then, they performed a stem completion task under an inclusion condition (complete the word with those previously learnt) and an exclusion condition (complete the word with a new one). Finally, they made a retrospective confidence judgment about how correct the completed-word was. Although no difference in accuracy between MS patients and controls was observed for retrospective confidence judgments, MS patients showed a JOL deficit in the exclusion condition. Pannu and Kaszniak (2005) conclude in their review that ‘for tasks that make higher monitoring demands, such as an episodic sentence memory task, or list learning and prediction tasks, MS patients show deficits in comparison to control subjects’ (p.114).

To the best of our knowledge, only two other studies have dealt with online awareness or metacognitive experience in MS. In Goverover, Genova, Griswold, Chiaravalloti, and DeLuca (2014), participants performed an internet-abilities test within which a score was calculated according to the help the participant needed to complete the task. Participants had to respond to questions on a 4-point scale about difficulty, time to perform the task and the ability to do similar task in their home environment. These three questions were gathered into a metacognitive score and were asked pre- and post-task. Participants were classified as “aware” or “unaware” according to their metacognitive score and their performance. Performance and metacognitive scores were divided into two groups based on a median split. The aware group was defined by an agreement between performance and metacognition: low metacognitive prediction and low performance or high metacognitive prediction and high performance. The results revealed a significant change in awareness for MS patients between pre- and post-diction (57% of patients were aware before, compared to 81% after) whereas no change was observed for healthy controls (70% against 75%) suggesting that MS patients need to experience the task to be aware of their performance. However, no interaction effect was reported by the authors and there was only a trend for a difference in percentage of aware participants between controls and MS patients for their predictions. This study suggests that patients may have a deficit in pre-task prediction – which simply measures domain-specific metacognitive knowledge (since it is made before they have experienced the task). On the other hand, after the task is completed, they show normal retrospective monitoring.

Middleton, Denney, Lynch, and Parmenter (2006) also took a retrospective estimation of performance on memory and executive function tasks. They found that neuropsychological evaluation was correlated with estimation for both trial-by-trial estimates and global post-diction. This correlation was significant for MS patients and healthy controls which indicates that patients have an intact relative metacognitive accuracy. However, participants from both patient and control groups consistently underestimated their performance, showing that whereas in general, estimates of performance in

both groups relate to the level of functioning, all participants fail to appreciate how well they have performed the task. In this case, underestimating performance is not specific to the MS patients. In sum, a correlational analysis (which points to those people with worse performance giving lower predictions) and an analysis of prediction magnitude (which points to systematic underestimation) yield different results: predictions are related to performance (in between subject correlations) but differ in magnitude (i.e., they are consistently lower than performance).

Three of these studies also assessed metacognitive knowledge by questionnaires. Beatty and Monson (1991) found no difference in questionnaire responses about awareness of memory abilities between MS patients and healthy controls. Middleton et al. (2006) study revealed no significant correlation between subjective evaluation and neuropsychological evaluation but the same correlation was also not significant for controls. Therefore, it suggests in this case this absence of relationship is not disease-dependent but it also found in normal functioning. However, Goverover et al. (2014) showed a higher discrepancy between self- and informant-rated questionnaires for the patients than for the controls leading to an awareness impairment of global functioning.

In sum, a complex pattern emerges from the results of the research assessing metacognitive knowledge and experiences in MS. These studies point to different profiles of metacognitive impairment in people with MS. They mainly support the idea that patients have a problem with high monitoring-demand tasks and have more of a prospective monitoring impairment than a retrospective one (since it appears that estimation of performance conducted after the task are consistently intact). Beatty and Monson's (1991) results underscore the complexity of the data. They showed different a pattern of impairments according to the patient's level of memory and executive functions. FOK judgments are predominantly impaired in patients (except for the “normal memory, normal WCST” group – who by definition do not have cognitive impairment). However, only the most impaired group overestimated their performance with the global prediction. Additionally, there was no difference in FOK judgments for general knowledge between the groups. These different profiles according to impaired/preserved metacognitive abilities and according to global cognitive impairment will be discussed in the next section.

---

#### 4. Discussion

The present review gathers information about metacognition and the awareness of deficit in the MS population. The results reveal a complex profile of intact and impaired metacognitive knowledge and metacognitive experience in MS. There is considerable heterogeneity in the samples and in the paradigms used, and even in the motivations for carrying out the studies. Here we summarize the key findings and highlight priorities for future research, taking on board the methodological pitfalls in awareness and metacognition research. The discussion is grouped around three key questions.



#### 4.1. Metacognitive knowledge: do MS patients make reliable subjective evaluations of their neuropsychological performance?

As in the results section, we identified two subgroups of studies: those using correlations between self-ratings and neuropsychological testing and those who evaluate under- and over-estimation of the abilities. We will consider these two subgroups and will discuss methodological issues with this respect.

The fifteen studies using correlations between self-rating and neuropsychological testing in order to measure awareness do not provide consistent results for both global and specific functions. They mainly support the idea that MS patients are aware in their subjective evaluations, even if the correlations with neuropsychological evaluations are weak. This relationship is difficult to interpret in terms of metacognition because most of these studies do not include a healthy control group. Without a healthy control group, it is not known how well (in general within any group), task predictions relate to individual differences in performance. That is, even if the correlations are low in the MS group, we might see exactly the same magnitude of correlation in the control group (as shown in the [Middleton et al. 2006](#) study). As such, it is difficult to pinpoint a disease specific effect without considering a comparison of correlations between subjective evaluations and neuropsychological evaluations in healthy groups as well as in MS groups. The only four studies out of fourteen which did have a control group did not statistically compare both groups. This could be explained by the fact that these studies are more focused on the perceived abilities of the patients than an awareness perspective. In any case, in a meta-analysis of 55 studies dealing with the relationship between subjective evaluation and neuropsychological evaluation in healthy participants across several cognitive and physical domains, [Mabe and West \(1982\)](#) found that the correlations between both measures were typically rather low ( $r = .29$ ). Therefore, it is difficult to interpret the observed low correlation between self-evaluation and neuropsychological evaluation in MS as low correlations may also be observed in healthy populations – as stated previously there is a lack of control data regarding correlations between subjective evaluations and neuropsychological evaluations.

However, it is still clinically informative to compare directly self-evaluations and neuropsychological evaluation. Our review reveals that cognitive deficits experienced by MS patients cannot be totally predicted by subjective complaints, but can also be due to emotional factors such as depression ([Hanssen et al., 2014](#); [Julian et al., 2007](#); [Maor et al., 2001](#); [Marrie et al., 2005](#); [Phillips & Stuifbergen, 2006](#)). However, because of the heterogeneity of MS, results differ across studies. As described above, patient-informant discrepancies seem to be a better estimator of awareness because the basis of the metacognitive score directly stems from the same measure and allows classifying patients according to under- or overestimation.

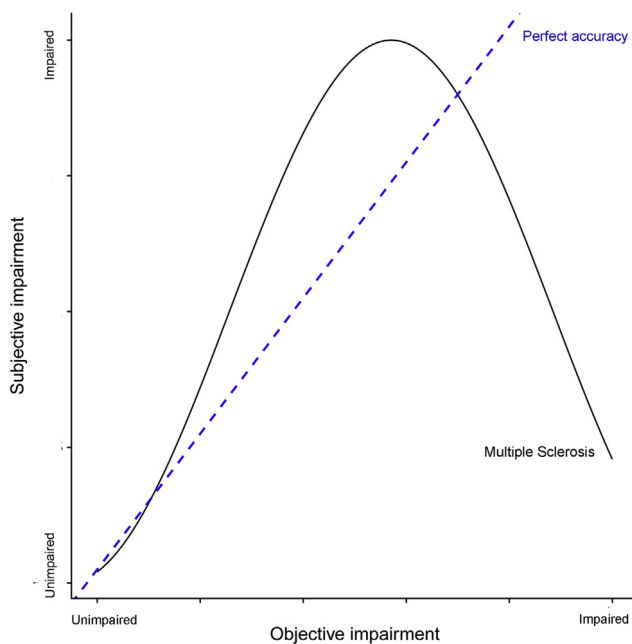
Methodological issues arise when using patient-informant discrepancies, because the measure hinges on how well the informant rating captures real-world performance (or neuropsychological evaluation). Although significant correlations are

found between this evaluation and informant-rated impairment, these correlations are again, moderate (e.g., from  $-.31$  to  $-.47$ ; [Carone et al., 2005](#)). It can also be difficult to interpret these patient-informant discrepancies because, again, there are rarely control comparisons. As with the correlations above, the group differences between controls and patients are often absent. For instance, [Carone et al. \(2005\)](#) had 37 control participants who had ‘informants’, which they compared to their 125 MS patients and their informants. The discrepancies between participants and informants for their patient and control groups were not significant. In a review of measures used in awareness research in Alzheimer’s disease, [Clare, Marková, Verhey, & Kenny. \(2005\)](#) cast some doubt on the use of patient-informant discrepancies – factors such as career burden or depression can influence the ratings, and moreover there are philosophical issues about whether someone else can make accurate estimates of subtle internal changes and subjective difficulties. It might be concluded that apart from some clear differences in subgroups of MS patients, people with MS are only unaware to the same degree that healthy people are unaware if one focuses on subjective complaints.

To evaluate metacognitive knowledge in MS, another subgroup of studies mainly focused on under- or over-estimation of performance. As we predicted, and according to the fact that MS is a heterogeneous disease, these studies suggest that about half of patients have some metacognitive impairment. Within this subgroup, patients more often underestimate their performance than overestimate it.

Overestimation was associated with more cognitive impairment in two studies ([Carone et al., 2005](#); [Rosti-Otajarvi et al., 2014](#)), but [Smith and Arnett \(2010\)](#) found no differences between under and overestimators groups. In line with [Marrie et al. \(2005\)](#), we suggest that the relationship between self-evaluation and neuropsychological evaluation is possibly quadratic. Mildly cognitively impaired patients report a greater degree of subjective impairment than the severely impaired patients. At the beginning of the disease, the diagnosis of MS is integrated into self-knowledge and beliefs (see the notion of *intellectual awareness* from [Toglia & Kirk, 2000](#)) and patients begin to consider the idea that they may have cognitive impairment even if physical disabilities are more often associated with the disease. From the CAM’s perspective, constructs like “*I am a person who suffering from MS*” would be updated in the Personal Database. This leads to an underestimation of performance which is in fact not associated to a real metacognitive impairment, but reflects beliefs, worries, and an updated self-concept. With disease progression, and the increase in cognitive symptoms, MS patients would be more likely to overestimate their performance as is observed in other cognitively impaired populations (e.g., Alzheimer’s disease, [Souchay, 2007](#)). [Fig. 3](#) is a proposal of this relationship between self-evaluation and neuropsychological evaluation based on [Marrie et al.’s](#) sketch.

Thus, patients underestimate their performance whilst they show only a slight decline in their cognitive abilities – in the early stage of the disease. If more severe cognitive deficits develop then it is possible that there is a more genuine metacognitive deficit as the patient may not have the cognitive resources to update or maintain realistic



**Fig. 3 – Proposed non-linear relationship between objective impairment and subjective impairment in MS patients, based on Marrie et al. (2005). Subjective impairment refers to self assessment of functioning, whereas objective impairment refers to performance on standardized neuropsychological tests. The hypothetical ‘perfect accuracy’ line indicates where subjective evaluation is equal to objective performance. The curve shows the observed pattern in MS, with values above the perfect accuracy line being underestimates of performance (because people judge themselves to be more impaired than they actually are – an underestimate of their actual functioning) and values below this line being overestimates (conversely people fail to appreciate their impairment, and overestimate their performance).**

conceptions of task performance. This proposition is not new (cf. Marrie et al., 2005), but it should be noted that this idea is supported by multiple studies in the review. Critically, MS patient overestimators are more cognitively impaired than those who are accurate and those who underestimate (e.g., Carone et al., 2005; Rosti-Otajarvi et al., 2014). Because there is a high fatigue rate in the most cognitively impaired patients, they are rarely included in studies which could explain that we report more underestimation than overestimation in the studies reviewed here. Future research needs to consider this issue in larger, more varied groups and consider disease progression and severity as a variable which is of significance to under and over-estimation. However, it needs to be noted that linear correlations may overlook this critical hypothesis. Therefore, we highlight the importance of measuring under- and overestimation of performance in the assessment of metacognitive knowledge. As related before, there are difficulties with the interpretations of the questionnaires and patient-informant discrepancies. As such, to consider under and overestimation in metacognitive knowledge we suggest that the global prediction method is the more reliable method.

Using such a method, Beatty and Monson's results (1991) only showed an overestimation for the most cognitively impaired group. We suggest that this group contains patients with longer disease duration than the other groups as we suppose that disease progression is associated to more cognitive symptoms. Beatty and Monson's measure of metacognitive knowledge (1991) is thus consistent with the non-linear relationship between self-evaluation and neuropsychological evaluation. However, no underestimation was found for the less impaired patients as with our proposal. We suggest this occurs because questionnaires are more sensitive to variables associated with underestimation in MS (i.e., depression, fatigue, see Section 4.3 for more details), whereas more concrete estimations, such as the number of items that can be recalled from a list are less sensitive to early disease changes. Moreover, when confronted with a word recall test (or similar) in a clinical setting, patients have no suitable reference point. Whereas they may have become worried about their cognitive performance in daily life and adjust their estimates accordingly, they will not necessarily have a concept of an appropriate number of the number of words they should remember from a list. Such an issue would easily be examined in future research by asking participants to give an up-to-date prediction of current functioning, as well as an estimate of functioning before they felt they were affected by the disease.

#### 4.2. Metacognitive experiences: do MS patients have problems in monitoring their cognitive functions?

The above sections point to a deficit in awareness for the patients who are most cognitively impaired. The results however, hinge on measures which can be influenced by other factors such as depression, self-esteem, and disease burden. There is also little consideration of controls' performance or the underlying cognitive processes which are responsible for awareness. To consider the question of awareness more fully it seems apt to focus on metacognitive tasks, although relatively few studies have focused on this in MS.

Our review revealed that patients with MS have problems with prospective metamemory but have intact retrospective metamemory (Goverover et al., 2014; Middleton et al., 2006; Scarrabelotti & Carroll, 1999). Within prospective monitoring, they show heterogeneous profiles according to the object of the metacognitive task: MS patients exhibit impaired episodic FOK but preserved semantic FOK (Beatty & Monson, 1991). As with the sections above, this impairment only seems to arise once the MS sample have low performance on the WSCT or memory tests.

Considerable research exists on the processes underlying FOK judgments, and the putative difference between semantic and episodic FOK is of interest. According to the *noncritical recollection hypothesis of episodic FOK* (Hertzog, Fulton, Sinclair, & Dunlosky, 2014), FOK judgments are partly based on the retrieval of the encoding context; retrieving information or details about the original encoding context. Based on studies finding on a memory retrieval deficit in MS (Calabrese, 2006; Rao, Leo, & St. Aubin-Faubert, 1989), impaired episodic FOK could be explained by a deficit in conscious retrieval of episodic recollection underpinned by frontal lobe dysfunction. This idea is also supported by

studies showing an autobiographical memory impairment in MS (Ernst et al., 2016), which has also been explained by impaired retrieval processes.

However, there is no consensus about the nature of memory disorder in MS (Chiaravalloti & DeLuca, 2008; Defer, Brochet, & Pelletier, 2010). Thus, a second explanation of the memory deficit in MS is the encoding-deficit hypothesis and is often seen as a consequence of processing speed and working memory impairment. From this perspective, impaired episodic FOK could be seen as a contextual encoding deficit and such deficit has been already raised in MS (Thornton, Raz, & Tucker, 2002). In MS, such deficit would be linked to a slower speed of processing or a decrease in efficient strategy used during encoding (Saenz, Bakchine, & Ehrle, 2015).

These two hypotheses (i.e., retrieval and encoding deficit) have also been suggested in order to explain the FOK profile observed in older adults. Likewise, in MS patients, most studies observed an age-related discrepancy between impaired episodic FOK and preserved semantic FOK (for a review see Castel, Middlebrooks, & McGillivray, 2016). This discrepancy has also been explained by two different hypotheses which refer to either a deficit in evaluation of contextual information associated to the retrieval process and executive functioning, or a decrease in encoding strength. With respect to the heterogeneity of the lesions in MS, and because encoding and retrieval both seem to be impaired in MS (Saenz, Bakchine, Jonin, & Ehrle, 2015), both explanations can be considered. In sum, the deficit in episodic FOK in MS could be proposed to be based on the underlying episodic deficit in the disease. When MS participants have impaired episodic memory, they also show a metamemory deficit on these particular tasks. For instance, Beatty and Monson (1991) showed that the group “low WCST, intact memory” exhibited impaired episodic FOK. As these patients also have a lower recall score in this task, we suggest that the memory deficit explains the metamemory impairment in MS patients. This calls into question the level of difficulty of the task, and thus future research needs to carry out multiple FOK episodic memory tasks, where it might be found that more difficult tasks are more sensitive to early changes in metacognitive accuracy in this population. It is therefore possible that a slight decline in memory performance for some particular tasks (those needing more cognitive resources, such as the episodic task in this case) induce a metamemory deficit even if these patients are not impaired on neuropsychological tests.

Regarding the monitoring at encoding measured by JOL, results are less clear. According to Scarrabelotti and Carroll (1999), people with MS have no prospective and retrospective JOL impairment in a word stem completion task. However, although they show intact relative accuracy in the inclusion condition (i.e., completing the sentence with a learned word), MS patients were less accurate than controls in the exclusion condition (i.e., completing the sentence with a new word). Two explanations can be considered. First, the exclusion task is simply more difficult, and as such, as above, this task may be more sensitive to early changes in metacognition. As above, longitudinal studies and those which consider disease severity are a priority for future research.

This impairment can also be explained by the fact that the JOL in the exclusion condition requires more controlled,

effortful processes. Thus, to be aware of an automatic process – to produce a target word in the exclusion condition refers to automatic process – is more difficult for MS patients because, according to Scarrabelotti and Carroll (1999), it requires more ‘controlled (intentional) processes’ (p.1346). Such an interesting hypothesis may be considered in future research, that there is less conscious control and less awareness of automatic processes in memory. This is a proposal (in memory) which resonates with the general tenet of the executive deficit in MS (Beatty et al., 1989; Calabrese, 2006; Chiaravalloti & DeLuca, 2008; Foong et al., 1997). It is also possible that MS patients have specific metamemory impairment in the exclusion condition because having a strong performance in this task mainly depends on recollection which is often impaired in MS and because, as above, the memory deficit seems to explain the metamemory impairment. Prospective monitoring at encoding seems to be intact in patients with MS but impaired under specific conditions which are more demanding in terms of controlled and intentional processes.

Overall, MS patients have an episodic FOK impairment and a JOL impairment which is apparent only under specific (and particularly demanding) conditions. This statement is in line with the executive function deficit in MS, since it has been shown that FOK judgments are correlated with a measure of executive functioning (i.e., performance on the WCST; Souchay, Isingrini, Clarys, Tacconnat, & Eustache, 2004) whereas simpler JOL judgments are not.

MS patients perform normally on retrospective monitoring tasks. This has been found for item-by-item tasks (Scarrabelotti & Carroll, 1999) and global post-task predictions (Goverover et al., 2014; Middleton et al., 2006), both of which are known to refer to monitoring processes (Connor et al., 1997). We suggest that this intact monitoring is due to the fact that retrospective judgments are made after the retrieval process, which acts as a salient cue to make the judgment; people with MS are able to reflect on successfully completed tasks and make accurate assessments of their performance based on experience. Such cues are not present in prospective metacognitive judgments (the FOK and JOL) and these are arguably more demanding in terms of cognitive load.

#### 4.3. What variables are associated with metacognitive knowledge and metacognitive experiences in MS?

A critical variable associated with impaired awareness in this pathology is the type of MS. Although few studies have focused on disease type, the results observed by Shermann et al. (2008) and Rosti-Otajarvi et al. (2014) are interesting. They suggest that patients suffering from SP and PP forms underestimate less their performance compared to the RR form even if no significant difference was found for the accurate group and the overestimators (Rosti-Otajarvi et al., 2014). Because results revealed that patients suffering from the SP form have a greater degree of unawareness than the RR one (51% against 15%), we suggest that progressive forms are associated with more overestimation (Shermann et al., 2008), consistent with the non-linear progression discussed above. Additionally, studies have shown that patients with SP and PP forms exhibit more cognitive impairment (e.g., Wishart & Sharpe, 1997); overestimators are more cognitively impaired

than both the accurate group and the underestimators (Carone et al., 2005; Rosti-Otajarvi et al., 2014). However, we cannot discriminate between a real form-dependent effect and the fact that patients suffering from progressive forms exhibit more cognitive symptoms. In the latter case, this would be consistent with the quadratic relationship between subjective evaluation and neuropsychological performance suggested above. From this perspective, patients with a greater degree of cognitive impairment are more likely to overestimate their performance which seems to be the case for the progressive forms.

A second variable associated with impaired awareness in MS is depression. Studies assessing metacognitive knowledge mainly support the idea that depression is a good predictor of self-rated cognitive impairment (Hanssen et al., 2014; Maor et al., 2001; Marrie et al., 2005). Depression is also related to a lower correlation between subjective evaluation and neuropsychological testing in memory (Phillips & Stuijbergen, 2006). Thus, depression could explain the cases where subjective complaints do not predict tests performance.

From a metacognitive point of view, a decrease in depression is also related to accuracy estimation in MS (Kinsinger et al., 2010). Additionally, underestimation of cognitive abilities is associated with more depression than overestimation (Carone et al., 2005; Van der Hiele et al., 2012). Because we suggest that underestimation occurs in the beginning of the disease – at least for patient with relatively few symptoms – and because depression is related to underestimation, it would be interesting to test the evolution of depression with disease progression. This would be particularly interesting as we suppose that more cognitive impairment is associated with overestimation and that overestimation seems to be linked with less depression. The same results were found with anxiety (Bruce et al., 2010) which was especially associated with underestimation (Van der Hiele et al., 2012). These effects are consistent with studies on metamemory which find an association between poor metamemory and affective disorders (e.g., Cavanaugh & Murphy, 1986). Additionally, this relationship was also suggested in other pathologies such as Alzheimer's disease (Clare, 2004) and Mild Cognitive Impairment (Roberts, Clare, & Wood, 2009).

Finally, Randolph et al. (2004) suggest that depressive beliefs could mediate the relationship between depression and self-rated cognitive impairment. The notion of belief is important in Togliola and Kirk's (2000) construct of metacognitive knowledge and is one of the bases of these judgments. Self-efficacy is especially related to perceived cognitive impairment in MS (Hughes et al., 2015) and should also be considered in studies dealing with metacognitive knowledge. Fatigue is also important to consider because it was also related to perceived cognitive impairment (Jougleux-Vie et al., 2014) and is the most frequent symptom in MS ranging from 53% to 83% of patients (Wood et al., 2013).

Overall, these variables influence measures of metacognitive knowledge which is impaired in half of the MS patients. However, these variables have not yet been measured in studies dealing with metacognitive experiences and should be also considered in future research. This would allow us to determine whether MS patients have a primary deficit in metacognition, or whether it is a consequence of related

problems. Continuing to make reference to the anosognosia framework is critical as it groups together cognitive and noncognitive factors (e.g., depression, anxiety).

## 5. Conclusion

A complex picture emerges, according to whether one considers disease awareness (as operationalized here as metacognitive knowledge) or metacognition per se, and moreover whether we consider metacognition as domain specific or domain general. According to this review, the most impaired half of MS patients exhibit poor subjective evaluation of their disease progression, as measured by questionnaires. Moreover, such patients mostly underestimate their performance. In line with Marrie et al. proposal (2005), we suggest that the relationship between perceived abilities and neuropsychological evaluation is quadratic (see Fig. 3). That is, mildly cognitively impaired patients exhibit a greater degree of subjective impairment – resulting in underestimation – than the severely impaired patients – which results in overestimation. From a CAM (Fig. 1) perspective, the disease concept should be updated in the Personal Database at the beginning of the disease, and any failure to up-date this model will result in over or under-estimation. According to this model, with depression, anxiety and fatigue, such updating would lead patients to underestimate their performance, as we have found in the review here. As the disease progresses, metacognitive impairments may also appear, since cognitive mechanisms which are required to monitor feedback and incorporate it into current goals and stored knowledge, are impaired. This in turn would result in overestimation of performance as has been observed in other pathologies (e.g., Alzheimer's disease; Morris & Mograbi, 2013).

We specifically considered component parts of the metacognitive system. Monitoring processes have almost uniquely been measured in memory tasks. These results support the idea that MS patients exhibit impaired relative accuracy on prospective judgments and especially for episodic FOKs and JOLs on tasks more reliant on recollective processes (see Souchay, 2007 for an explanation of this process in Alzheimer's disease). This suggests that any observed metamemory deficit in MS patients is a consequence of their memory impairment. The memory deficit in MS is slight, and according to Saenz, Bakchine and Ehrle (2015), this impairment seems to be based on both encoding and retrieval deficits. We would add that, in particular, the deficit with controlled recollection processes in MS (see for example Seinelä, Hämäläinen, Koivisto, & Ruutinen, 2002; Scarrabelotti & Carroll, 1999), leads to difficulties in monitoring currently unrecalled information on episodic tasks (see Mograbi & Morris, 2013). That is, the very specific deficits on certain memory tasks are only secondary to the difficulty in generating the requisite information from episodic memory on which to base accurate metacognitive judgments.

In turn, a critical issue is whether to consider metacognition as domain specific or domain general. The studies reviewed in Table 1 point to a bias towards using memory tasks to measure metacognition, and yet conclusions are often drawn about metacognitive function more generally on

the basis of performance in such tasks. Recent evidence suggests that metacognition is not domain general (e.g., Fleming, Ryu, Golfinos, & Blackmon, 2014), but specific for each task and cognitive process, and as such one should be cautious about inferring general deficits from performance on one specific domain. Here, then, we call for a more complete account of metacognition in MS, perhaps focusing more on the deficits which are of most relevance to the disease: executive functions (in terms of cognition) and even fatigue, strength and energy levels (as non-cognitive symptoms of the disease) – in line with Hoogervorst et al. (2001). Moreover, perhaps in the absence of cognitive deficits in MS (as is the case in a large proportion of the population with this disease) it is perhaps inappropriate to interpret erroneous or idealistic predictions of performance as metacognitive impairments. Indeed, we propose that in patients without cognitive difficulties, the search for impaired metacognition is fruitless. In terms of the CAM model (Morris & Mograbi, 2013), discussed above, there is perhaps insufficient studies to adjudicate on the different types of anosognosia, but we can propose based on the review here, that if anything, the deficit would be mnemonic anosognosia, arising from faulty memory and updating.

Finally, the variables associated with impaired metacognitive knowledge (or ‘disease awareness’) in MS should be considered as well in experiments dealing with metacognitive judgments, especially depression and fatigue which are common in this pathology. These variables are of interest for clinicians as they are strongly associated with subjective complaints, but are also known to alter metacognitive accuracy (e.g., Moore & Fresco, 2012).

## Acknowledgments

Audrey Mazancieux is supported by Bourse Ministérielle awarded by the French Government to cover her PhD costs. Chris Moulin is a senior member of the Institut Universitaire de France.

## REFERENCES

- Agnew, S. K., & Morris, R. G. (1998). The heterogeneity of anosognosia for memory impairment in Alzheimer's disease: A review of the literature and a proposed model. *Aging & Mental Health*, 2(1), 7–19.
- Arnett, P. A., Rao, S. M., Bernardin, L., Grafman, J., Yetkin, F. Z., & Lobeck, L. (1994). Relationship between frontal lobe lesions and Wisconsin Card Sorting Test performance in patients with multiple sclerosis. *Neurology*, 44(3 Part 1), 420–425.
- Babinski, M. J. (1914). Contribution à l'étude des troubles mentaux dans l'hémiplégie organique cérébrale. *Revue Neurologique*, 1, 845–848.
- Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology*, 25(5), 729–735.
- Basso, M. R., Shields, I. S., Lowery, N., Ghormley, C., Combs, D., Arnett, P. A., et al. (2008). Self-reported executive dysfunction, neuropsychological impairment, and functional outcomes in multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, 30(8), 920–930.
- Beatty, W. W., Goodkin, D. E., Beatty, P. A., & Monson, N. (1989). Frontal lobe dysfunction and memory impairment in patients with chronic progressive multiple sclerosis. *Brain and Cognition*, 11(1), 73–86.
- Beatty, W. W., & Monson, N. (1991). Metamemory in multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, 13(2), 309–327.
- Bruce, J. M., Bruce, A. S., Hancock, L., & Lynch, S. (2010). Self-reported memory problems in multiple sclerosis: Influence of psychiatric status and normative dissociative experiences. *Archives of Clinical Neuropsychology*, 25(1), 39–48.
- Calabrese, P. (2006). Neuropsychology of multiple sclerosis. *Journal of Neurology*, 253, 10–15.
- Carone, D. A., Benedict, R. H. B., Munschauer, F. E., III, Fishman, I., & Weinstock-Guttman, B. (2005). Interpreting patient/informant discrepancies of reported cognitive symptoms in MS. *Journal of the International Neuropsychological Society*, 11(5), 574–583.
- Castel, A. D., Middlebrooks, C. D., & McGillivray, S. (2016). Monitoring memory in old age: Impaired, spared, and aware. In J. Dunlosky, & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 519–536). Oxford University Press.
- Cavanaugh, J. C., & Murphy, N. Z. (1986). Personality and metamemory correlates of memory performance in younger and older adults. *Educational Gerontology*, 12(4), 385–394.
- Clare, L., Marková, I., Verhey, F., & Kenny, G. (2005). Awareness in dementia: a review of assessment methods and measures. *Aging & Mental Health*, 9(5), 394–413.
- Chiaravalloti, N. D., & DeLuca, J. (2008). Cognitive impairment in multiple sclerosis. *The Lancet Neurology*, 7(12), 1139–1151.
- Christodoulou, C., Melville, P., Scherl, W. F., Morgan, T., MacAllister, W. S., Canfora, D. M., et al. (2005). Perceived cognitive dysfunction and observed neuropsychological performance: Longitudinal relation in persons with multiple sclerosis. *Journal of the International Neuropsychological Society*, 11(5), 614–619.
- Ciurli, P., Bivona, U., Barba, C., Onder, G., Silvestro, D., Azicnuda, E., et al. (2010). Metacognitive unawareness correlates with executive function impairment after severe traumatic brain injury. *Journal of the International Neuropsychological Society*, 16(2), 360–368.
- Clare, L. (2004). Awareness in early-stage Alzheimer's disease: A review of methods and evidence. *British Journal of Clinical Psychology*, 43(2), 177–196.
- Cocchini, G., & Della Sala, S. (2010). Assessing anosognosia for motor and language impairments. In *The study of anosognosia* (pp. 123–144).
- Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging*, 12(1), 50–71.
- Costa, S. L., Genova, H. M., DeLuca, J., & Chiaravalloti, N. D. (2017). Information processing speed in multiple sclerosis: Past, present, and future. *Multiple Sclerosis Journal*, 23(6), 772–789.
- De Gardelle, V., & Mamassian, P. (2014). Does confidence use a common currency across two visual tasks? *Psychological Science*, 25(6), 1286–1288.
- Defer, G., Brochet, B., & Pelletier, J. (2010). *Neuropsychologie de la sclérose en plaques*. Elsevier Masson.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *CVLT-II*. The Psychological Corporation.
- Drew, M., Tippett, L. J., Starkey, N. J., & Isler, R. B. (2008). Executive dysfunction and cognitive impairment in a large community-based sample with Multiple Sclerosis from New Zealand: A descriptive study. *Archives of Clinical Neuropsychology*, 23(1), 1–19.
- Ernst, A., Moulin, C. J., Souchay, C., Mograbi, D. C., & Morris, R. (2016). Anosognosia and metacognition in Alzheimer's

- disease: Insights from experimental psychology. In *The Oxford handbook of metamemory* (Vol. 1, pp. 451–472).
- Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Awareness and metacognition. *Consciousness and Cognition*, 9(2), 324–326.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906–911.
- Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1338–1349.
- Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain: a Journal of Neurology*, 137(10), 2811–2822.
- Foong, J., Rozewicz, L., Quaghebeur, G., Davie, C. A., Kartsounis, L. D., Thompson, A. J., et al. (1997). Executive function in multiple sclerosis. The role of frontal lobe pathology. *Brain: a Journal of Neurology*, 120(1), 15–26.
- Gerretsen, P., Chung, J. K., Shah, P., Plitman, E., Iwata, Y., Caravaggio, F., et al. (2017). Anosognosia is an independent predictor of conversion from mild cognitive impairment to Alzheimer's disease and is associated with reduced brain metabolism. *The Journal of Clinical Psychiatry*, 78(9), 1187–1196.
- Goverover, Y., Chiaravalloti, N., & DeLuca, J. (2005). The relationship between self-awareness of neurobehavioral symptoms, cognitive functioning, and emotional symptoms in multiple sclerosis. *Multiple Sclerosis Journal*, 11(2), 203–212.
- Goverover, Y., Genova, H., Griswold, H., Chiaravalloti, N., & DeLuca, J. (2014). Metacognitive knowledge and online awareness in persons with multiple sclerosis. *NeuroRehabilitation*, 35(2), 315–323.
- Hanssen, K. T., Beiske, A. G., Landrø, N. I., & Hessen, E. (2014). Predictors of executive complaints and executive deficits in multiple sclerosis. *Acta Neurologica Scandinavica*, 129(4), 234–242.
- Henry, J. D., & Beatty, W. W. (2006). Verbal fluency deficits in multiple sclerosis. *Neuropsychologia*, 44(7), 1166–1174.
- Hertzog, C. (1992). Improving memory: The possible roles of metamemory. In *Memory improvement* (pp. 61–78). New York: Springer.
- Hertzog, C., Fulton, E. K., Sinclair, S. M., & Dunlosky, J. (2014). Recalled aspects of original encoding strategies influence episodic feelings of knowing. *Memory & Cognition*, 42(1), 126–140.
- Hoogervorst, E. L. J., Van Winsen, L. M. L., Eikelenboom, M. J., Kalkers, N. F., Uitdehaag, B. M. J., & Polman, C. H. (2001). Comparisons of patient self-report, neurologic examination, and functional impairment in MS. *Neurology*, 56(7), 934–937.
- Hughes, A. J., Beier, M., Hartoonian, N., Turner, A. P., Amtmann, D., & Ehde, D. M. (2015). Self-efficacy as a longitudinal predictor of perceived cognitive impairment in individuals with multiple sclerosis. *Archives of Physical Medicine and Rehabilitation*, 96(5), 913–919.
- Jougleux-Vie, C., Duhin, E., Deken, V., Outteryck, O., Vermersch, P., & Zéphir, H. (2014). Does fatigue complaint reflect memory impairment in multiple sclerosis? *Multiple Sclerosis International*, 2014.
- Julian, L., Merluzzi, N. M., & Mohr, D. C. (2007). The relationship among depression, subjective cognitive impairment, and neuropsychological performance in multiple sclerosis. *Multiple Sclerosis Journal*, 13(1), 81–86.
- Kinsinger, S. W., Lattie, E., & Mohr, D. C. (2010). Relationship between depression, fatigue, subjective cognitive impairment, and objective neuropsychological functioning in patients with multiple sclerosis. *Neuropsychology*, 24(5), 573–580.
- Krch, D., Sumowski, J. F., DeLuca, J., & Chiaravalloti, N. (2011). Subjective memory in multiple sclerosis is associated with initial-trial learning performance. *Journal of the International Neuropsychological Society*, 17(3), 557–561.
- Kujala, P., Portin, R., & Ruutiainen, J. (1996). Memory deficits and early cognitive deterioration in MS. *Acta Neurologica Scandinavica*, 93(5), 329–335.
- Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, 67(3), 280–296.
- Maor, Y., Olmer, L., & Mozes, B. (2001). The relation between objective and subjective impairment in cognitive function among multiple sclerosis patients—the role of depression. *Multiple Sclerosis Journal*, 7(2), 131–135.
- Mariné, C., & Huet, N. (1998). Techniques d'évaluation de la métacognition. I Les mesures indépendantes de l'exécution de tâches. II Les mesures dépendantes de l'exécution de tâches. *L'année psychologique*, 98(4), 711–742.
- Marrie, R. A., Chelune, G. J., Miller, D. M., & Cohen, J. A. (2005). Subjective cognitive complaints relate to mild impairment of cognition in multiple sclerosis. *Multiple Sclerosis Journal*, 11(1), 69–75.
- Matotek, K., Saling, M. M., Gates, P., & Sedal, L. (2001). Subjective complaints, verbal fluency, and working memory in mild multiple sclerosis. *Applied Neuropsychology*, 8(4), 204–210.
- McDonald, I., & Compston, A. (2006). The symptoms and signs of multiple sclerosis. *McAlpine's Multiple Sclerosis*, 4, 321–322.
- McGlynn, S. M., & Schacter, D. L. (1989). Unawareness of deficits in neuropsychological syndromes. *Journal of Clinical and Experimental Neuropsychology*, 11(2), 143–205.
- Middleton, L. S., Denney, D. R., Lynch, S. G., & Parmenter, B. (2006). The relationship between perceived and objective cognitive functioning in multiple sclerosis. *Archives of Clinical Neuropsychology*, 21(5), 487–494.
- Mograbi, D. C., & Morris, R. G. (2013). Implicit awareness in anosognosia: Clinical observations, experimental evidence, and theoretical implications. *Cognitive Neuroscience*, 4(3–4), 181–197.
- Morris, R. G., & Mograbi, D. C. (2013). Anosognosia, autobiographical memory and self knowledge in Alzheimer's disease. *Cortex*, 49(6), 1553–1565.
- Moore, M. T., & Fresco, D. M. (2012). Depressive realism: A meta-analytic review. *Clinical Psychology Review*, 32(6), 496–509.
- Morris, R. G., & Hannesdottir, K. (2004). Loss of 'awareness' in Alzheimer's disease. In R. G. Morris, & J. Becker (Eds.), *Cognitive neuropsychology of Alzheimer's disease* (Vol. 2, pp. 375–398). Oxford University Press.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In *The psychology of learning and motivation*. Academic Press, 125–173.
- Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. In *Metacognition: Knowing about knowing* (pp. 1–25).
- Norman, D. A., & Shallice, T. (1986). Attention to action. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation*. Boston, MA: Springer.
- Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychology Review*, 15(3), 105–130.
- Phillips, L. J., & Stuifbergen, A. K. (2006). Predicting continued employment in persons with multiple sclerosis. *Journal of Rehabilitation*, 72(1), 35–43.
- Planche, V., Gibelin, M., Cregut, D., Pereira, B., & Clavelou, P. (2016). Cognitive impairment in a population-based study of patients with multiple sclerosis: Differences between late relapsing–remitting, secondary progressive and primary progressive multiple sclerosis. *European Journal of Neurology*, 23(2), 282–289.
- Prigatano, G. P. (1999). *Principles of neuropsychological rehabilitation*. Oxford University Press.

- Prigatano, G. P., & Schacter, D. L. (Eds.). (1991). *Awareness of deficit after brain injury: Clinical and theoretical issues*. Oxford University Press.
- Rabbitt, P. (Ed.). (2004). *Methodology of frontal and executive function*. Psychology Press.
- Randolph, J. J., Arnett, P. A., & Freske, P. (2004). Metamemory in multiple sclerosis: Exploring affective and executive contributors. *Archives of Clinical Neuropsychology*, 19(2), 259–279.
- Randolph, J. J., Arnett, P. A., & Higginson, C. I. (2001). Metamemory and tested cognitive functioning in multiple sclerosis. *The Clinical Neuropsychologist*, 15(3), 357–368.
- Rao, S. M., Leo, G. J., & Aubin-Faubert, P. S. (1989). On the nature of memory disturbance in multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, 11(5), 699–712.
- Rao, S. M., Leo, G. J., Bernardin, L., & Unverzagt, F. (1991). Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. *Neurology*, 41(5), 685–691.
- Roberg, B. L., Bruce, J. M., Lovelace, C. T., & Lynch, S. (2012). How patients with multiple sclerosis perceive cognitive slowing. *The Clinical Neuropsychologist*, 26(8), 1278–1295.
- Roberts, J. L., Clare, L., & Woods, R. T. (2009). Subjective memory complaints and awareness of memory functioning in mild cognitive impairment: A systematic review. *Dementia and Geriatric Cognitive Disorders*, 28(2), 95–109.
- Rosti-Otajarvi, E., Ruutiainen, J., Huhtala, H., & Hämäläinen, P. (2014). Relationship between subjective and objective cognitive performance in multiple sclerosis. *Acta Neurologica Scandinavica*, 130(5), 319–327.
- Rosti, E., Hämäläinen, P., Koivisto, K., & Hokkanen, L. (2007). PASAT in detecting cognitive impairment in relapsing-remitting MS. *Applied Neuropsychology*, 14(2), 101–112.
- Saenz, A., Bakchine, S., Jonin, P. Y., & Ehrlé, N. (2015). Atteinte de la memoire episodique verbale dans la sclerose en plaques: Revue critique des processus cognitifs concernes et de leur exploration. *Revue Neurologique*, 171(8), 624–645.
- Scarrabelotti, M., & Carroll, M. (1999). Memory dissociation and metamemory in multiple sclerosis. *Neuropsychologia*, 37(12), 1335–1350.
- Seinela, A., Hämäläinen, P., Koivisto, M., & Ruutiainen, J. (2002). Conscious and unconscious uses of memory in multiple sclerosis. *Journal of the Neurological Sciences*, 198(1), 79–85.
- Sherman, T. E., Rapport, L. J., & Ryan, K. A. (2008). Awareness of deficit in multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, 30(3), 301–311.
- Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(4), 907–912.
- Smith, M. M., & Arnett, P. A. (2010). Awareness of executive functioning deficits in multiple sclerosis: Self versus informant ratings of impairment. *Journal of Clinical and Experimental Neuropsychology*, 32(7), 780–787.
- Souchay, C., Isingrini, M., & Espagnet, L. (2000). Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology*, 14(2), 299–309.
- Souchay, C. (2007). Metamemory in Alzheimer's disease. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 43(7), 987–1003.
- Souchay, C., & Isingrini, M. (2004). Age related differences in metacognitive control: Role of executive functioning. *Brain and Cognition*, 56(1), 89–99.
- Souchay, C., Isingrini, M., Clarys, D., Taconnat, L., & Eustache, F. (2004). Executive functioning and judgment-of-learning versus feeling-of-knowing in older adults. *Experimental Aging Research*, 30(1), 47–62.
- Spren, O., & Strauss, E. (1991). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.
- Sunderaraman, P., & Cosentino, S. (2017). Integrating the constructs of anosognosia and metacognition: A review of recent findings in dementia. *Current Neurology and Neuroscience Reports*, 17(3), 17–27.
- Thompson, A. J., Kermode, A. G., Wicks, D., MacManus, D. G., Kendall, B. E., Kingsley, D. P. E., et al. (1991). Major differences in the dynamics of primary and secondary progressive multiple sclerosis. *Annals of Neurology*, 29(1), 53–62.
- Thornton, A. E., Raz, N., & Tucker, K. A. (2002). Memory in multiple sclerosis: Contextual encoding deficits. *Journal of the International Neuropsychological Society*, 8(3), 395–409.
- Toglia, J., & Kirk, U. (2000). Understanding awareness deficits following brain injury. *NeuroRehabilitation*, 15(1), 57–70.
- Van der Hiele, K., Spliethoff-Kamminga, N. G. A., Ruimschotel, R. P., Middelkoop, H. A. M., & Visser, L. H. (2012). The relationship between self-reported executive performance and psychological characteristics in multiple sclerosis. *European Journal of Neurology*, 19(4), 562–569.
- Wishart, H., & Sharpe, D. (1997). Neuropsychological aspects of multiple sclerosis: A quantitative review. *Journal of Clinical and Experimental Neuropsychology*, 19(6), 810–824.
- Wood, B., Van Der Mei, I. A. F., Ponsonby, A. L., Pittas, F., Quinn, S., Dwyer, T., et al. (2013). Prevalence and concurrence of anxiety, depression and fatigue over time in multiple sclerosis. *Multiple Sclerosis Journal*, 19(2), 217–224.
- Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1310–1321.

## **Appendix B**

# **A multidimensional assessment of metacognition across domains in Multiple Sclerosis**



# **A multidimensional assessment of metacognition across domains in Multiple Sclerosis**

Audrey Mazancieux

LPNC CNRS 5105, Université Grenoble Alpes

Chris J.A. Moulin

LPNC CNRS 5105, Université Grenoble Alpes

and

Institut Universitaire de France

Olivier Casez

Department of Neurology, Université Grenoble Alpes

Céline Souchay

LPNC CNRS 5105, Université Grenoble Alpes

## **Corresponding Author**

Audrey Mazancieux

1251 avenue Centrale, St Martin d'Hères

38040 Grenoble FRANCE

audrey.mazancieux@univ-grenoble-alpes.fr

## **Abstract**

**Objective:** In neurological diseases, metacognitive judgements have been widely used in order to assess the degree of disease awareness. However, as yet little research of this type has focused on Multiple Sclerosis (MS).

**Method:** We here focused on an investigation of item-by-item metacognitive predictions (using Feeling-Of-Knowing judgements) in episodic and semantic memory and global metacognitive predictions in standard neuropsychological tests pertinent to MS (processing speed and verbal fluency). Twenty-seven relapsing-remitting MS (RR-MS) patients and 27 comparison participants took part.

**Results:** We found that RR-MS patients were as accurate as the group of comparison participants on our episodic and semantic item-by-item judgements. However, for the global predictions we found that the MS group initially overestimated their performance ( $d_s = .64$ ), but only on a task on which performance was also impaired ( $d_s = .89$ ; processing speed). We suggest that MS patients, under certain conditions, show inaccurate metacognitive knowledge. However, postdictions and item-by-item predictions indicate that on-line metacognitive processes are no different from participants without MS.

**Conclusion:** We conclude that there is no monitoring deficit in RR-MS and as such these patients should benefit from adaptive strategies and symptom education.

**KEYWORDS:** Multiple Sclerosis, metacognition, self-awareness, metamemory, anosognosia

In the context of pathology, awareness is critical for patient care. Being aware of cognitive or physical impairments is crucial for both the efficacy of cognitive rehabilitation programmes (Prigatano, 1999) and the understanding of the impact of cognitive disabilities on activities of daily living (McGlynn & Schacter, 1989). The focus of this paper is Multiple Sclerosis (MS). Whilst a considerable number of studies have examined the question of disease awareness in MS (for a review see Mazancieux, Souchay, Casez & Moulin, 2019), most research has considered self-report and questionnaire measures. In this article we invoke the metacognition framework to consider disease awareness in MS. Metacognition broadly refers to the knowledge of, the monitoring of (self-evaluation) and the control of (strategy implementation) cognitive activity (Nelson & Narens, 1990). It allows the evaluation of awareness in asking patients to make metacognitive judgments. These judgments refer to a self-assessment of performance on a particular cognitive task.

Although metacognition has been widely evaluated in different neurological and psychiatric diseases (e.g., Pannu & Kaszniak, 2005), there are surprisingly few studies focusing on the evaluation of metacognition in MS despite the high incidence of this pathology. MS is an autoimmune inflammatory disease characterized by lesions which can appear across the whole central nervous system. These lesions produce a neural and neuronal demyelination which compromises the conduction of information (Trapp & Nave, 2008). In addition to physical disabilities, cognitive impairment is also frequent in MS with prevalence rates ranging from 43% to 70% (Chiaravalloti & DeLuca, 2008). Although cognitive symptoms vary in MS, a common profile emerges where the majority of these symptoms are related to an executive functioning deficit as a potential consequence of processing speed impairments (Drew, Tippett, Starkey, & Isler, 2008). As has been shown in other pathologies (see Souchay, 2007 for a review in Alzheimer's disease), in traumatic brain injury (Ciurli et al., 2010) or in healthy aging (e.g., Souchay & Isingrini, 2004), impaired performance of

executive function tests are associated with metacognitive difficulties. This leads to the expectation that MS patients might exhibit metacognitive impairment. Perhaps more importantly, from a clinical viewpoint, assessing the level of symptom awareness is crucial in order to help patients to use pertinent strategies when dealing with their symptoms. Apart from executive function related symptoms (including planning, flexibility, inhibition, working memory, Rabbitt, 2004) and processing speed, impairments in long-term memory in verbal and visual modalities has been reported (e.g., Calabrese, 2006; Ruet, 2015).

The majority of studies investigating awareness in MS have focused on a comparison of self-evaluations of cognitive functioning (mainly by questionnaires) with more objective neuropsychological evaluations (e.g., Maor, Olmer, & Mozes, 2001; Randolph, Arnett, & Freske, 2004; Roberg, Bruce, Lovelace, & Lynch, 2012). In a recent review of the scant literature on metacognition in MS (Mazancieux et al., 2019), we suggested a non-linear relationship between the subjective evaluation of cognitive impairment and a more objective evaluation (i.e. neuropsychological assessment). Patients with a slight decline in their cognitive abilities tend to underestimate their abilities whereas patients with more cognitive impairment tend to overestimate them. This failure in self-evaluation is also associated with emotional disturbances and fatigue which are prevalent in MS patients (Kesselring & Klement, 2001). For instance, it has been shown that depression is associated with metacognitive inaccuracy in MS (Kinsinger Lattie, & Mohr, 2010) and a lower correlation between subjective evaluation and neuropsychological memory tests (Phillips & Stuijbergen, 2006). Beliefs about cognitive functioning (referred here to metacognitive knowledge (Flavell, 1979), and which are easily operationalised in questionnaire studies) have been the most evaluated metacognitive construct in MS. However, since methodological issues arise from the comparison between subjective evaluation via questionnaire and neuropsychological

evaluation, a more reliable way to measure metacognitive knowledge is the use of global predictions, as used in the current paper.

In the global prediction paradigm, participants are asked to predict their performance in a cognitive task. For instance on an episodic memory task, they predict the number of items they think they will be able to recall from a list. By comparing the prediction to the actual performance (i.e., the number of recalled items), it is possible to have an idea of the awareness of the cognitive function. When the prediction is made before the task, it allows an estimation of or metacognitive knowledge (generalised beliefs about the task which might include lay understandings of aging or the disease process; Hertzog, 1992). Measured after the task, 'postdictions' evaluate metacognitive experience, especially monitoring processes referring to the update of self-evaluation derived from on-line monitoring of the ongoing task (Connor, Dunlosky & Hertzog, 1997).

A more fine-grained analysis of monitoring is achieved by asking participants to make item-by-item judgements. In MS, Beatty and Monson (1991) asked patients and non-MS participants to perform item-by-item Feeling-Of-Knowing (FOK) judgements, where participants have to predict their future ability to recognize an item that they have failed to recall. Two versions of the task exist. In the episodic FOK (eFOK) task, participants first learn paired-words and then have to recall the target from a presented cue. If they are unable to recall the target word, they report their likelihood of recognizing it in a list of words. This judgement is the FOK. In the semantic FOK (sFOK) task, participants answer general knowledge questions. As in the eFOK task, they have to say if they think they will be able to recognize the answer if they are not able to recall it. From these FOKs, it is possible to examine metacognitive biases (the over- and underestimation of performance) and metacognitive sensitivity (the discrimination between correct and incorrect recognition). In the only study in MS, patients exhibited poor metacognitive sensitivity in the eFOK task

(Beatty & Monson, 1991). However, this study is inconclusive for several reasons. First, the authors did not distinguish between different forms of MS. Second, alternative - more reliable - measures of metacognitive sensitivity have subsequently been developed. Third, the examination of metacognition was limited to memory tasks. The current study aimed to address these shortcomings.

The present study aims to further examine metacognitive functioning in people with RR-MS, the most common form of MS (80% of patients, Rao et al., 1991). The present study proposes a general overview of metacognition in MS with relapsing-remitting MS (RR-MS) patients, since this is the most common form (80% of patients, Rao et al., 1991). First, we decided to measure eFOK and sFOK, a common strategy for exploring metacognition in cognitive impaired groups (e.g., Alzheimer's disease, Souchay, 2007; patients with focal frontal lesions, Schnyer et al., 2004; Korsakoff's syndrome, Shimamura & Squire, 1986; autism spectrum disorders, Wojcik, Moulin, & Souchay, 2013). A typical profile is of impaired eFOK accuracy when patients exhibit episodic memory impairment. On the contrary, sFOK accuracy is preserved in these studies. There is an overwhelming bias for measuring metacognition through memory tasks in MS (Mazancieux et al., 2019). As such, even though eFOK and sFOK tasks are robust and often used as measures of metacognition, memory function may not be the most pertinent task on which to test the metacognition of people with MS.

We assume that focusing on more relevant functions would allow a more complete picture of awareness in this pathology. From a clinical point of view, we assume that measuring awareness of a cognitive activity is especially relevant when there is a specific impairment in this cognitive activity. Therefore, we also adopted a procedure where participants can make metacognitive judgements about standard neuropsychological tasks where MS patients are often impaired: the Symbol Digit Modalities Test (SDMT) and the

conceptual verbal fluency task (Planche, Gibelin, Cregut, Pereira, & Clavelou, 2016; Ruet, 2015).

The SDMT is a processing speed task where first an association of symbols with digits is provided. In the test phase, only the symbols are presented, and participants have to say the digit associated with each symbol as rapidly as possible. As slowing is the main cognitive impairment in MS, patients often exhibit a deficit in this task. In the conceptual fluency task, participants have to generate as many words as possible in a given time from a semantic category (e.g., animals). Similarly, MS patients often show significant impairments (slowing) on this task where self-initiated processes and strategic search in memory are involved. In order to assess awareness of these cognitive abilities, we added metacognitive judgements to these two tasks focusing on global predictions to measure both metacognitive knowledge and metacognitive monitoring.

Our rationale was to have a protocol that mixed very commonly used monitoring tasks (FOKs) and global predictions on tasks that are pertinent for MS. In particular, we proposed metacognitive judgements on neuropsychological tests that are particularly used in this population. The SDMT is one of the most used tests in MS (e.g., Planche et al., 2015; Walker, et al., 2016; O'Brien, et al., 2007; Basso, et al., 2008; Ruet, 2015). Regarding verbal fluency, several studies have found that it is also a good predictor of RR-MS severity (Prakash, Snook, Lewis, Motl, & Kramer, 2008) and a selective impairment of semantic fluency in RR-MS has been shown (despite a preserved phonemic fluency, Santiago, Guardia, Casado, Carmona & Arbizu, 2007). Thus from a neuropsychological viewpoint, these are tasks where we may expect to find deficits, and as such examining metacognitive awareness in these tasks would be of critical interest, even though these are less typically studied in a metacognitive context.

In sum, there is very little existing research into metacognition in this population, and existing works focus mainly on memory function with varying disease types. This study aims

to explore more precisely metacognitive processes in MS that are likely to be impaired due to the neuropsychological profile with executive deficits in this population.

## **Method**

### **Participants**

Twenty-seven patients (21 female, 6 male;  $M_{age} = 39.48$ ,  $SD_{age} = 9.93$ ) were included in the study. The diagnosis of MS was established by a neurologist who also informed the patients about the study. Inclusion criteria were to have no recent exacerbation of MS symptoms, and no other neurological disease. Participants were excluded if they had a form of MS other than RR-MS, or a history of alcohol or drug abuse. Twenty-seven non-MS volunteer participants (21 female, 6 male;  $M_{age} = 39.03$ ,  $SD_{age} = 10.80$ ) also took part in the study as a comparison group. People in the comparison group voluntarily chose to participate to the study without being paid for their participation. Information about the study was given in the hospital where patients were tested and in Grenoble Alpes University. This advertisement targeted the general public, but no patient family member was recruited to the control group. Only people with no history of neurological disease, psychiatric disease, or alcohol or drug abuse were included in the control group. Patients and comparison group participants were matched one-by-one in terms of gender, age ( $\pm 5$  years) and years of education ( $\pm 3$  years). Demographic and clinical data are summarised in Table 1.

Participants were tested either in the Laboratoire de Psychologie et Neurocognition (LPNC) or in the Centre Hospitalier Universitaire in Grenoble. The study was approved by the Ethics Committee for Non-Interventional Research of Grenoble. All data included in this manuscript was obtained in compliance with the Helsinki Declaration.

INSERT TABLE 1 HERE



## **Material and procedure**

All participants were tested individually in one 60 to 75 minute session. The whole procedure included two metacognition tasks: global predictions and the eFOK and sFOK tasks, as well as the completion of two questionnaires. Global prediction and FOK task order was randomly assigned for each participant.

**Global prediction.** Participants performed two neuropsychological tasks: a verbal fluency task and the SDMT (oral version). The standard tasks were slightly modified in order to assess and compare metacognitive awareness across tasks. Participants had 45s (instead of 120s) to give as many numbers as possible for the SDMT task. Two versions were created in order to have two trials (see Figure 1). For the verbal fluency task, participants again had two trials and had 45s (instead of 120s) to give words either from the category ‘animals’ or ‘fruits and vegetables’. After the task was explained to the participants, they were asked to predict the score they would achieve. For the fluency task, participants were asked ‘how many words from the category do you think you will generate in 45s?’ For the SDMT task, participants were asked ‘how many numbers do you think you will read in 45s?’ These predictions were made once before the task was performed (prediction) and for a second time after completion (postdiction). For the postdiction, participants were asked to estimate their prior performance on the same basis (number of items achieved). There were two trials per task, which enables the examination of the ability to integrate feedback from having completed the task into the predictions for a second trial. Therefore, for each task, participants performed an initial prediction of performance, then conducted the task, and following the task, made a postdiction. Then, they had to make a second prediction, complete a different version of the task, and make a second postdiction. Trial order (version 1 and 2 for the SDMT and animal category or fruit and vegetable category for the verbal fluency task) was randomly assigned for each participant.

INSERT FIGURE 1 HERE

**FOK tasks.** The material used for the eFOK and sFOK tasks was similar to those used by Souchay, Moulin, Clarys, Taconnat and Isingrini (2007). These materials allow some control of difficulty between the episodic memory task and the semantic memory task since the same target word is used in both tasks. Each target has a definition used in the sFOK task and an associative cue used in the eFOK task. All the targets were divided into two lists so that each participant would not have the same target word in both tasks. Half of the participants had the first list for the episodic task and the second list for the semantic task, with the other half having the reverse pattern.

The eFOK task included three stages: encoding, cued recall, and recognition. Participants firstly attempted to learn 40 paired-words with the first word written in uppercase and the second written in lowercase. Each word pair was presented for 5 seconds. During the recall stage, only the cue (i.e., the word written in lowercase) was presented and the participant was asked to retrieve the associated target word (i.e., the uppercase word) with 15 s to do so. After this time passed, they had to give a FOK judgment, reporting whether they thought they would recognize the correct target amongst a 5-word list. As in Souchay and colleagues (2007), the FOK decisions were in a 'yes' or 'no' format. No feedback about the correctness of the recall was given to the participants, and FOK judgments were made for all items. After the recall stage for all cues had been completed, participants performed a five-alternative forced choice recognition task. The 40 cues were presented again and the participants had to find the correct associated target with the presented cue. There was no time limit for this stage.

The sFOK task included recall and recognition phases. First of all, participants attempted recall for 40 general-information questions. As in the eFOK task, they had 15 s to respond and then made an FOK judgment in the same manner as the eFOK procedure. After this, they performed a recognition task, where participants were again presented the 40 general-information questions with five-alternative responses. The two tasks were constructed using E-prime software and were presented to the participants on a 15.6 inch computer screen. Half of the participants started with eFOKs and half with sFOKs.

**Emotional and fatigue assessment.** Both patients and the comparison group completed two questionnaires at the end of the testing session. The first one was the Beck Depression Inventory II (BDI-II; Beck, Steer, & Brown, 1996) and the second was the Fatigue Impact Scale (FIS; Fisk, Ritvo, Ross, Haase, Marrie, & Schleich, 1994). For participants who reported being too tired by the experimental procedure, questionnaires were sent by e-mail and were completed within one week.

### **Statistical analyses**

Analyses were conducted using R software. Data and analysis scripts are available on Open Science Framework (<https://osf.io/fyshb/files/>). The main interest in metacognition is the accuracy of the judgments; that is the comparison between the judgment and the performance. Regarding global predictions, we first focused on the magnitude of predictions as simply the number of items participants predict. Then, we calculated accuracy scores in terms of the relation between predicted and actual performance. This score is non-directional meaning that it allows an estimate of how precise are participants without being influenced by metacognitive bias (underestimation or overestimation of performance; see Moulin et al., 2002). These two measures capture different aspects of metacognition: someone can consistently overestimate their performance, but yet be relatively accurate with a small discrepancy between their prediction and the score. Because we expect differences in terms of

task performance, and because these differences might influence accuracy scores, the prediction was transposed into a percentage of performance. More precisely, each prediction was expressed in a proportion of performance for each trial using the following formula:  $\text{Prediction} * 100 / \text{Performance}$  (e.g., a participant with task performance of 30 and prediction of 10 would have predicted 1/3 of their performance having therefore a percentage of performance of 33%. A participant with task performance of 20 and prediction of 40 would have predicted 150% of his or her performance). To control for bias, the non-directional difference between this score and performance (that refers to 100% in this context) was computed. Therefore, an accuracy score of 0 suggests that the participant has a perfect accuracy, and an accuracy score of 10 refers to a deviation of 10% from performance. Due to recording issues, one patient did not have prediction and postdiction scores for the fluency task.

To avoid effects of potential outliers which might be found in patients who have by definition a non-normal behaviour, we used linear mixed-effect models computed using ‘lmerTest’ and ‘lme4’ packages (Kuznetsova, Brockhoff, & Christensen, 2017). Therefore, we estimated for each model an intercept per participant as a random effect. These effects are not the main focus of this paper, therefore we only reported fixed-effects. As there is no consensus regarding the calculation of effect size for mixed-effects models especially when several variables are included in the model (Westfall, Kenny, & Judd, 2014), we decided to calculate Cohen’s  $d$  from the  $t$  value as is done for regular  $t$ -tests (Lakens, 2013). When the effect included the between-subject group comparison the  $ds$  value was calculated, and we used the  $dz$  formula in cases where the effect included only within-subject variables (Lakens, 2013).

For the FOK tasks, we focused on both metacognitive bias and metacognitive sensitivity. Metacognitive sensitivity was estimated by two different approaches. First we

calculated the Type 2  $d'$  (Higham, 2007; Nelson, 1984) as follows: Type 2  $d' = z(H2) - z(FA2)$  where  $z$  is the inverse of the cumulative normal distribution function. Here H2 refers to Type 2 hits which are the proportion of reported 'yes' FOKs for correct responses and FA2 refers to Type 2 false alarms which are the proportion of reported 'yes' FOKs for incorrect responses. When H2 and the FA2 rates were equal to either 1 or 0, we used standard corrections (Green & Swets, 1966), using  $1/(2NC)$  instead of a rate of 0 and  $1-1/(2NI)$  instead of a rate of 1 (where NC is the number of correct responses and NI the number of incorrect responses). However, because Type 2  $d'$  is influenced by metacognitive bias (see Fleming and Lau, 2014), we also computed mixed-effects logistic regressions between task performance (correct and incorrect responses) and FOK (yes and no). The difference (i.e., the slope) between yes and no FOK allows the estimation of the capacity to judge future recognition according to task performance. Therefore, the larger the difference, the higher the discrimination between correct and incorrect responses in the recognition task. Moreover, this mixed-effect model effect allows the estimation of an intercept and a slope for FOK per participant as a random effect controlling for cross-participants variability. Finally, we calculated the percentage of correct answers for the 'yes' FOKs for each participant in order to estimate bias in metamemory judgments. Other analyses were standard  $t$ -tests.

## Results

### Global predictions

**Task performance.** Analyses of task performance for the SDMT task showed a main effect of group,  $t(52) = 3.27, p = .002, d_s = 0.89$ , with patients having a lower score. There was neither an effect of trial,  $t(52) = 1.21, p = .232$ , nor an interaction between the two factors,  $t(52) = -1.21, p = .232$ . Regarding the fluency task, we found no main effect of group,  $t(52) = 0.94, p = .352$ , no effect of trial,  $t(52) = 1.59, p = .119$ , and no interaction,  $t(52) = 1.02, p = .313$  (see Table 2).

**Magnitude of predictions.** Magnitudes of raw predictions were compared according to group, trial, and judgement type (prediction vs postdiction) for each task (Figure 2 and 3). For the SDMT task, the analyses revealed a main effect of group,  $t(52) = 2.04, p = .047, d_s = 0.56$ , with patients overall predicting completing fewer items compared to the comparison group. There was a main effect of trial,  $t(156) = 3.29, p = .001, d_z = 0.45$ , with a higher prediction (i.e. more items) for the first trial compared to the second trial. We also found a significant interaction between group and trial,  $t(156) = -2.33, p = .021, d_s = 0.63$ . Irrespective of judgement type, MS patients have lower predictions compared to the comparison group for the second trial,  $t(64.15) = 2.67, p = .010, d_z = 0.73$ , but not for the first trial,  $t(64.15) = 1.19, p = .240$ . Finally, the analyses revealed an interaction between trial and judgement type,  $t(156) = 3.25, p = .001, d_s = 0.44$ . Irrespective of groups, participants have a trend for lower postdictions compared to predictions in the first trial,  $t(156) = 1.91, p = .059$ , and have the opposite pattern of results in the second trial,  $t(156) = -2.68, p = .008, d_z = 0.37$ . Regarding the fluency task, we found a main effect of trial,  $t(145.09) = 3.80, p < .001, d_z = 0.52$ , with a larger prediction for the first trial compared to the second trial. No other effect was significant.

INSERT TABLE 2 HERE

**Metacognitive accuracy.** We calculated accuracy scores as outlined above which were compared according to group, trial, and judgement type (prediction vs postdiction) for each task (Figure 2 and 3). For the SDMT task, the analyses revealed a main effect of judgement type,  $t(156) = 3.90, p < .001, d_z = 0.53$ , predictions being less accurate than postdictions. No other main effects or interactions were significant, but we found a trend for a three-way-interaction,  $t(156) = 1.92, p = .056, d_s = 0.52$ . Therefore, we compared the

interaction between group and trial for prediction on the one hand and postdiction on the second hand. Although we found no effect for postdiction,  $t(156) = -0.15, p = .884$ , predictions showed a significant interaction between group and trial,  $t(156) = 2.58, p = .011, d_s = 0.70$ . Critically, patients were less accurate at predicting their performance than the comparison group for the first trial,  $t(156) = 2.37, p = .019, d_s = 0.64$ , which was not the case for the second trial,  $t(156) = -0.79, p = .433$ . Regarding the fluency task<sup>1</sup>, the analyses revealed only a main effect of judgement type,  $t(155.66) = 2.45, p = .015, d_z = 0.33$ , predictions being less accurate than postdictions.

INSERT FIGURE 2 HERE

INSERT FIGURE 3 HERE

### **FOK tasks**

**Recall and recognition.** The percentage of correct recall and correct recognition were calculated for each task and each participant. No difference between MS patients and the comparison group was found for recall either in the episodic memory task or the semantic memory task. The same result was found for the recognition performance (see Table 3).

INSERT TABLE 3 HERE

**Metacognitive sensitivity.** A Type 2  $d'$  was calculated for each participant and each task (Figure 4). For the episodic memory task, four participants had a performance rate of 1 so they were excluded from the following analysis. Overall, participants had a Type 2  $d'$

---

<sup>1</sup> For this analysis, we excluded one prediction of a comparison participant in the second trial which was extremely inaccurate (deviation of 328%). Running the same analysis but leaving in this participant did not change the pattern of significant results.

significantly different from 0 for both the sFOK task,  $t(52) = 4.15, p < .001$ , and the episodic FOK task,  $t(52) = 5.35, p < .001$ . There were no differences between metacognitive sensitivity between MS patients and non-MS participants for both the sFOK task,  $t(52) = -1.31, p = .195$ , and the eFOK task,  $t(52) = -1.78, p = .082$ .

INSERT FIGURE 4 HERE

Moreover, we fitted two mixed effects logistic regressions on sensitivity, with FOKs and group as fixed effects (Figure 5). We estimated an intercept and a slope for FOKs by participants as random effects. For the episodic memory task, the model showed a significant relationship between task accuracy and FOKs (estimate = 0.48,  $Z = 2.85, p = .004$ ) revealing that participants were able to predict correctly their memory performance. This relationship was not different according to group (estimate = -0.51,  $Z = -1.66, p = .097$ ), MS patients being as accurate as comparison group participants. For the semantic memory task, the model only showed a trend between task performance and FOKs (estimate = 0.33,  $Z = 1.83, p = .067$ ). This relationship was not different according to group (estimate = -0.45,  $Z = -1.32, p = .188$ ), MS patients being as accurate as non-MS participants.

INSERT FIGURE 5 HERE

**Metacognitive bias.** Metacognitive bias was estimated by calculating the percentage correct responses in the recognition task for the ‘yes’ FOKs for each participants and each task (Figure 6). For the episodic memory task, the analysis showed no effect of group,  $t(52) = -0.56, p = .578$ , as well as for the semantic memory task,  $t(52) = 0.51, p = .614$ . MS patients and the comparison group have therefore the same tendency to report ‘yes’ FOK for correct responses in the recognition task.



INSERT FIGURE 6 HERE

### **Relationship between metacognition and others variables**

We compared scores to the FIS and the BDI between MS patients and comparison group participants. MS patients had a higher score than the comparison group on the BDI ( $M_{patients} = 13.63$ ,  $SD_{patients} = 9.63$ ;  $M_{comparison} = 6.70$ ,  $SD_{comparison} = 5.02$ ),  $t(52) = 3.32$ ,  $p = .002$ ,  $d_s = 0.90$ , and on the FIS ( $M_{patients} = 75.33$ ,  $SD_{patients} = 27.57$ ;  $M_{comparison} = 54.22$ ,  $SD_{comparison} = 33.22$ ),  $t(52) = 2.54$ ,  $p = .014$ ,  $d_s = 0.69$ . To investigate the relationship between metacognitive sensitivity and emotional and fatigue variables, we performed correlational analyses with patients. No type 2  $d'$  values correlated with the BDI scale, the FIS scale, or the EDSS. Recall did not correlate with any of these individual difference variables either. Finally, as the first prediction for the SDMT was impaired in patients, we explored the relationship between this score and depression, fatigue, and disease duration. No correlation reached significance.

### **Discussion**

The current study proposes a multidimensional assessment of metacognition in RR-MS patients. We used global predictions and item-by-item predictions to measure both metacognitive knowledge and metacognitive monitoring. The novelty of this experiment was to measure metacognition on tasks which are relevant in MS (the SDMT and the verbal fluency) as well as typical metacognitive tasks (FOKs). Patients showed only significant impairment for the SDMT task which is consistent with the fact that processing speed is one of the main cognitive impairments in MS (Planche et al., 2015).

Regarding global predictions, predictions before the tasks were less accurate than postdictions in both groups and for the two tasks therefore replicating previous results in memory (e.g. Moulin, Perfect, & Jones, 2000). For the fluency task, there was no group

difference in terms of performance: MS patients predicted their performance at the same magnitude as our comparison group, and were therefore as accurate. For the SDMT task, MS patients had a lower task performance. However, patients overall predicted the same number of items as our comparison group therefore being less accurate. This was the case for the prediction of the first trial only. Thus, MS patients were able to have accurate predictions when having experienced the task (i.e., for postdictions and second-trial predictions). Across all tasks, MS patients can update their self-evaluation even though they have dysfunctional beliefs at first.

Patients' metacognitive knowledge was inaccurate, as gauged by the initial global predictions, for tasks before completing the task. As proposed in Mazancieux et al. (2019), such predictions are more associated with mood variables (depression, anxiety, etc.), fatigue, and self-esteem than with executive functions (that are more involved in monitoring processes). Although our sample of patients were more depressed and reported more fatigue than comparison group participants, these variables were not correlated with the accuracy on the first prediction for the SDMT task. As we have previously suggested (Mazancieux, et al. 2019), depression and fatigue in MS could lead to an underestimation of performance. However, in the present study, patients on average predict the same number of items as comparison group participants. Therefore, we looked at the signed difference between this initial prediction and performance. The number of overestimators (16 patients) and underestimators (11 patients) was almost the same, however there were no difference between these groups in terms of depression,  $t(25) = -0.04, p = .971$ , fatigue,  $t(25) = 1.37, p = .181$ , or disease duration,  $t(25) = -0.20, p = .844$ . There was thus no systematic under or overestimation in the MS group, and no relation to other measures.

The processes involved in under- and overestimation are not the same. The underestimation of performance might be associated with low self-beliefs and concern about

function. On the contrary, overestimation of performance can occur when patients have more cognitive impairment and therefore do not have enough cognitive resources perform accurate predictions and to update their metacognitive knowledge when faced with changes in their function. In previous studies in MS, overestimation was also associated with more cognitive impairments (Carone et al., 2005; Rosti-Otajärvi et al., 2014; but see Smith & Arnett, 2010). Moreover, these impairments were more related to tasks measuring executive functioning, however we did not measure these abilities, so we do not know if our two subsets of patients differ with this respect. In conclusion, in our sample of MS patients we have a pattern which is consistent not with over or underestimation but a lack of accuracy in estimating an upcoming task for which they have not experienced. They are less accurate than the comparison group in this regard only on a task where they are impaired (SDMT). If anything, future research could consider beliefs prior to conducting tasks, but in all other regards we did not find deficits in MS patients in metacognitive awareness per se with global measures, even when there is a significant deficit in performance. Once they have had the opportunity to experience a task, people with MS make an appropriate evaluation of their performance. Metacognition and disease awareness are complex multidimensional constructs, and it is clear that mood and knowledge impinge on people's evaluations. We proposed a multi-dimensional consideration of metacognition in a previous review (Mazancieux et al., 2019) but less is known about how these factors relate in MS than in other pathologies such as Alzheimer's disease (Mograbi & Morris, 2014).

Likewise, regarding FOKs, MS patients have the same metacognitive sensitivity as the comparison group participants which does not reproduce previous findings (Beatty & Monson, 1991). The main difference between our work and the previous study is that we exclusively focused on RR-MS. Primary progressive (PP-MS) and secondary progressive (SP-MS) are the forms of MS with the most cognitive impairment (e.g., Planche et al., 2015).

In an awareness interview, Sherman et al. (2008) showed that 51.5% of SP-MS patients have an unawareness of deficit compared to only 14.7% for RR-MS patients. Similarly, in Beatty and Monson's (1991), groups with impairment in episodic memory monitoring included at least half of PP-MS and SP-MS. It is therefore very likely that their patients are both more impaired and heterogenous than our sample (note that they had a lower score than controls on a verbal fluency test which was not the case in the present study). It remains a priority to consider disease type, severity and duration to produce a full picture of metacognitive function in MS.

The present study suggests that RR-MS patients with slight cognitive impairments can adequately update their evaluations, therefore showing intact metacognitive monitoring. In our sample of MS patients, cognitive impairment results in lower performance on our version of the SDMT task only. If there is any evidence of metacognitive impairment, it is in inaccurate self-knowledge on a task where the MS group showed impairment, namely processing speed. On this initial prediction, consistent with the large variability in MS, half of the patients overestimated their performance whereas the other half underestimated it. The difference between under and over estimation was not captured by depression, fatigue, or disease duration measures in this study, perhaps due to our sample size. On a clinical note, it suggests that these patients are likely able to have adaptive strategies in daily living activities and will benefit from cognitive rehabilitation techniques more efficiently (Prigatano, 1999). A priority is now to verify this pattern in relatively homogenous groups of MS patients as used here but with more pronounced cognitive impairment to observe whether monitoring dysfunction occurs with more cognitive impairment (such as executive functions) rather than a MS-trait. It will also be of clinical and theoretical relevance to take the metacognitive approach into domains which are perhaps more sensitive to the cognitive changes in MS, such as autobiographical memory (e.g. Ernst et al., 2013).

## **Acknowledgement**

The authors declare that there is no conflict of interest. No specific financial support was provided for this research.

## References

- Beatty, W. W., & Monson, N. (1991). Metamemory in multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, *13*(2), 309-327.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio*, *78*(2), 490-498.
- Carone, D. A., Benedict, R. H. B., Munschauer III, F. E., Fishman, I., & Weinstock-Guttman, B. (2005). Interpreting patient/informant discrepancies of reported cognitive symptoms in MS. *Journal of the International Neuropsychological Society*, *11*(5), 574-583.
- Chiaravalloti, N. D., & DeLuca, J. (2008). Cognitive impairment in multiple sclerosis. *The Lancet Neurology*, *7*(12), 1139-1151.
- Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and aging*, *12*(1), 50-71.
- Ernst, A., Blanc, F., Voltzenlogel, V., De Seze, J., Chauvin, B., & Manning, L. (2013). Autobiographical memory in multiple sclerosis patients: assessment and cognitive facilitation. *Neuropsychological rehabilitation*, *23*(2), 161-181.
- Fisk, J. D., Ritvo, P. G., Ross, L., Haase, D. A., Marrie, T. J., & Schlech, W. F. (1994). Measuring the Functional Impact of Fatigue: Initial Validation of the Fatigue Impact Scale. *Clinical Infectious Diseases*, *18*, 79-83.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American psychologist*, *34*(10), 906-911.
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in human neuroscience*, *8*, 443.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York: Wiley.

- Hertzog, C. (1992). Improving memory: The possible roles of metamemory. In *Memory Improvement* (pp. 61-78). Springer New York.
- Higham, P. A. (2011). Accuracy discrimination and type-2 signal detection theory: Clarifications, extensions, and an analysis of bias. In *Constructions of Remembering and Metacognition* (pp. 109-127). Palgrave Macmillan, London.
- Kesselring, J., & Klement, U. (2001). Cognitive and affective disturbances in multiple sclerosis. *Journal of neurology*, 248(3), 180-183.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82(13).
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, 4, 863.
- Maor, Y., Olmer, L., & Mozes, B. (2001). The relation between objective and subjective impairment in cognitive function among multiple sclerosis patients-the role of depression. *Multiple Sclerosis Journal*, 7(2), 131-135.
- Mazancieux, A., Souchay, C., Casez, O., Moulin, C. J. A. (2019). Metacognition and self-awareness in Multiple Sclerosis. *Cortex*, 111, 238-255.
- McGlynn, S. M., & Schacter, D. L. (1989). Unawareness of deficits in neuropsychological syndromes. *Journal of Clinical and Experimental Neuropsychology*, 11(2), 143-205.
- Moulin, C. J., Perfect, T. J., & Jones, R. W. (2000). Global predictions of memory in Alzheimer's disease: Evidence for preserved metamemory monitoring. *Aging, Neuropsychology, and Cognition*, 7(4), 230-244.
- Nelson, T. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95(1), 109-133.
- Nelson, T. O., & Narens, L. (1990). The psychology of learning and motivation. *Metamemory: A theoretical framework and new findings*.

- Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychology review*, *15*(3), 105-130.
- Planche, V., Gibelin, M., Cregut, D., Pereira, B., & Clavelou, P. (2016). Cognitive impairment in a population-based study of patients with multiple sclerosis: differences between late relapsing– remitting, secondary progressive and primary progressive multiple sclerosis. *European journal of neurology*, *23*(2), 282-289.
- Prigatano, G. P. (1999). *Principles of neuropsychological rehabilitation*. Oxford University Press.
- Rabbitt, P. (Ed.). (2004). *Methodology of frontal and executive function*. Psychology Press
- Randolph, J. J., Arnett, P. A., & Freske, P. (2004). Metamemory in multiple sclerosis: exploring affective and executive contributors. *Archives of Clinical Neuropsychology*, *19*(2), 259-279.
- Rao, S. M., Leo, G. J., Bernardin, L., & Unverzagt, F. (1991). Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. *Neurology*, *41*(5), 685-691.
- Roberg, B. L., Bruce, J. M., Lovelace, C. T., & Lynch, S. (2012). How patients with multiple sclerosis perceive cognitive slowing. *The Clinical neuropsychologist*, *26*(8), 1278-1295.
- Rosti-Otajärvi, E., Ruutiainen, J., Huhtala, H., & Hämäläinen, P. (2014). Relationship between subjective and objective cognitive performance in multiple sclerosis. *Acta Neurologica Scandinavica*, *130*(5), 319-327.
- Ruet, A. (2015). Cognitive Impairment in Multiple Sclerosis. In *Neuropsychiatric Symptoms of Inflammatory Demyelinating Diseases* (pp. 227-247). Springer, Cham.
- Schnyer, D. M., Verfaellie, M., Alexander, M. P., LaFleche, G., Nicholls, L., & Kaszniak, A. W. (2004). A role for right medial prefrontal cortex in accurate feeling-of-knowing



- judgments: Evidence from patients with lesions to frontal cortex. *Neuropsychologia*, 42(7), 957-966.
- Shimamura, A. P., & Squire, L. R. (1986). Memory and metamemory: a study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(3), 452-460.
- Smith, M. M., & Arnett, P. A. (2010). Awareness of executive functioning deficits in multiple sclerosis: Self versus informant ratings of impairment. *Journal of clinical and experimental neuropsychology*, 32(7), 780-787.
- Souchay, C. (2007). Metamemory in Alzheimer's disease. *Cortex*, 43(7), 987-1003.
- Souchay, C., Moulin, C. J., Clarys, D., Taconnat, L., & Isingrini, M. (2007). Diminished episodic memory awareness in older adults: Evidence from feeling-of-knowing and recollection. *Consciousness and cognition*, 16(4), 769-784.
- Trapp, B. D., & Nave, K. A. (2008). Multiple sclerosis: an immune or neurodegenerative disorder? *Annual review of neuroscience*, 31, 247-269.
- Wojcik, D. Z., Moulin, C. J., & Souchay, C. (2013). Metamemory in children with autism: Exploring “feeling-of-knowing” in episodic and semantic memory. *Neuropsychology*, 27(1), 19-27.

## Tables

**Table 1.** Means and standard deviation for demographic and clinical data for the MS patients and the comparison group. EDSS: the Expanded Disability Status Scale.

	MS patients	Comparison Group	Cohen's <i>d</i>
N =	27	27	
Age in years	39.48 (9.93)	39.03 (10.80)	0.04
Education in years	14.04 (2.08)	14.56 (2.03)	0.25
EDSS	2.56 (1.93)	n.a	
Disease duration in years	6.96 (3.23)	n.a	

**Table 2.** Means and standard deviations for predictions, performance, and postdictions (in number of items) according to groups, trials, and tasks.

	<i>SDMT task</i>		<i>Fluency task</i>	
	MS patients	Comparison Group	MS patients	Comparison Group
<i>Trial 1</i>				
Prediction	30.07 (12.81)	30.48 (12.16)	21.70 (8.50)	22.15 (6.67)
Performance	28.78 (7.76)	33.74 (6.62)	21.04 (5.04)	23.04 (3.69)
Postdiction	25.41 (9.00)	30.89 (10.74)	20.65 (7.29)	22.56 (6.64)
<i>Trial 2</i>				
Prediction	21.81 (6.20)	28.41 (11.33)	18.22 (6.25)	19.96 (6.03)
Performance	27.41 (6.39)	33.74 (5.80)	20.59 (5.83)	21.00 (6.97)
Postdiction	24.78 (6.47)	31.44 (9.08)	18.88 (7.09)	20.44 (7.96)

**Table 3.** Means and standard deviations for proportion of correct recall and recognition according to group and memory task.

	MS patients	Comparison Group	t(52) value	p value
<i>Episodic memory</i>				
Recall	0.37 (0.19)	0.38 (0.20)	0.28	0.784
Recognition	0.85 (0.11)	0.86 (0.10)	0.23	0.816
<i>Semantic memory</i>				
Recall	0.43 (0.17)	0.45 (0.23)	0.29	0.775
Recognition	0.75 (0.11)	0.77 (0.14)	-0.56	0.581

## Figures

### Trial 1

1	2	3	4	5	6	7	8	9
↔	‡	=		≠	□	∅	∈	∃

---

‡	∃	‡	∃		∃		∃	↔	∈	∃	=	↔	∅	‡	=	□		∈	=	↔	∅	∈	‡	≠
	∅	↔	∅	≠	∈		↔	≠	‡	□	∃	∃	≠	□	∅	□	‡	∃		∈	∅	‡	∈	□
∈	□	‡	∈	‡	∃		∅		∈	□	∅	=	↔	□	‡	↔	∈	∅		=	↔	□	‡	∃

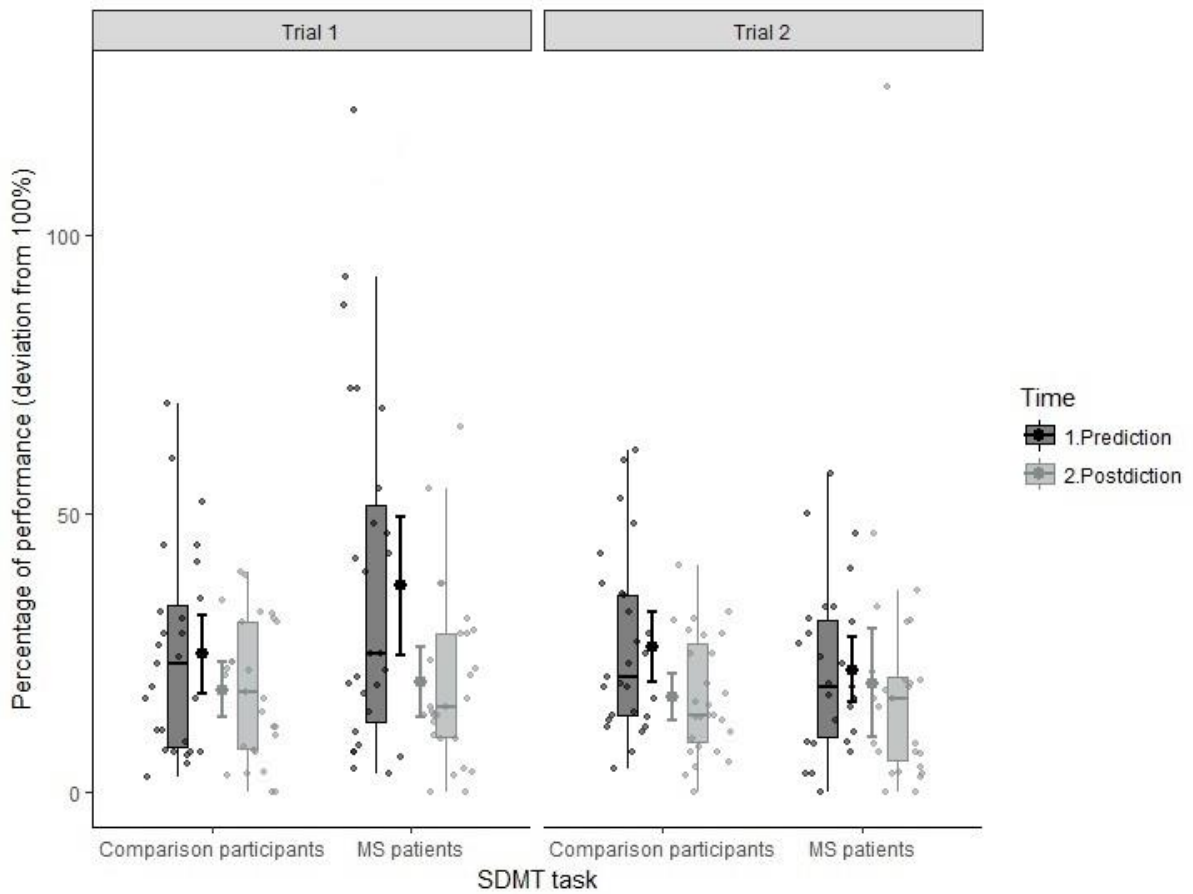
### Trial 2

1	2	3	4	5	6	7	8	9
	×	=	∧	⊂	⊃	⊥	⊃	↗

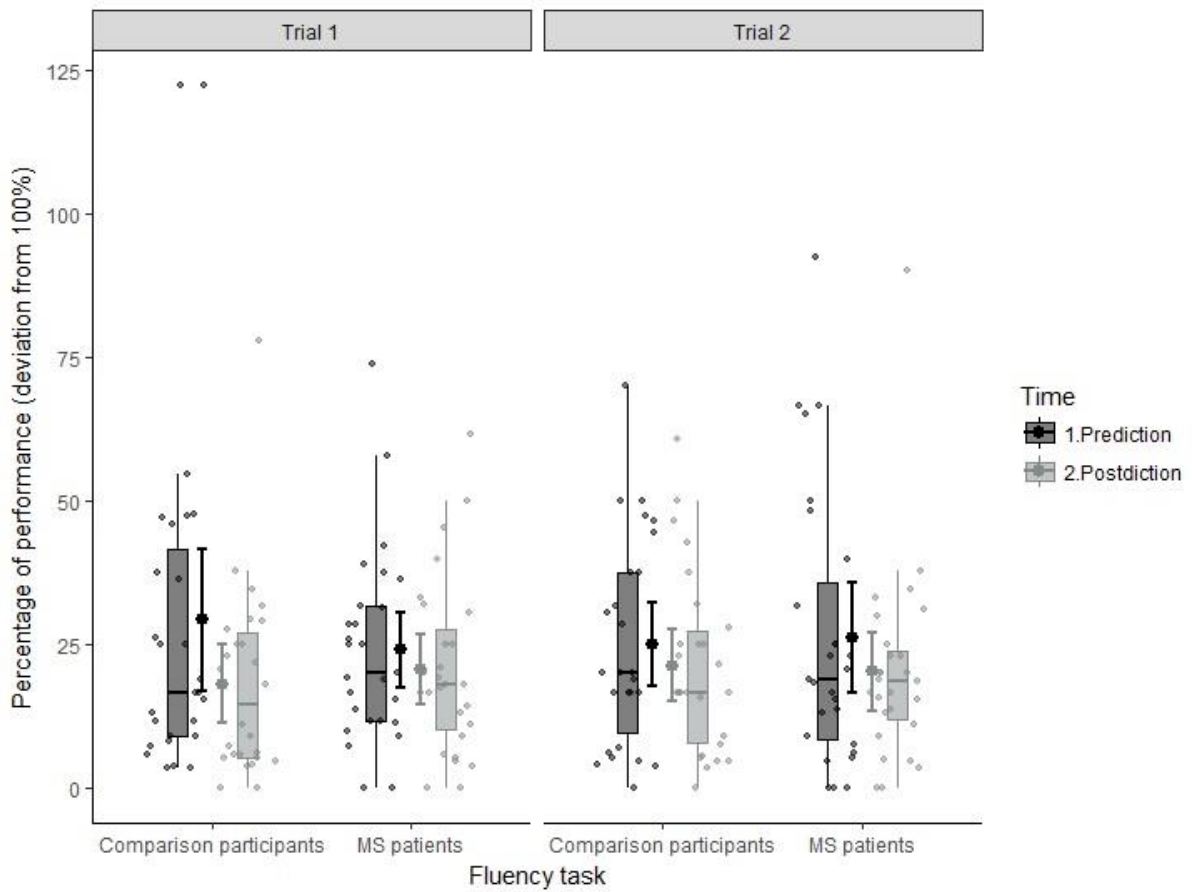
---

×	↗	×	↗	∧	↗	∧	↗		⊃	↗	=		⊥	×	=	⊃	∧	⊃	=		⊥	⊃	×	⊂
∧	⊥		⊥	⊂	⊃	∧		⊂	×	⊃	↗	↗	⊂	⊃	⊥	⊃	×	↗	∧	⊃	⊥	×	⊃	⊂
⊃	⊂	×	⊃	×	↗	∧	⊥	∧	⊃	⊂	⊥	=		⊂	×		⊃	⊥	∧	=		⊂	×	↗

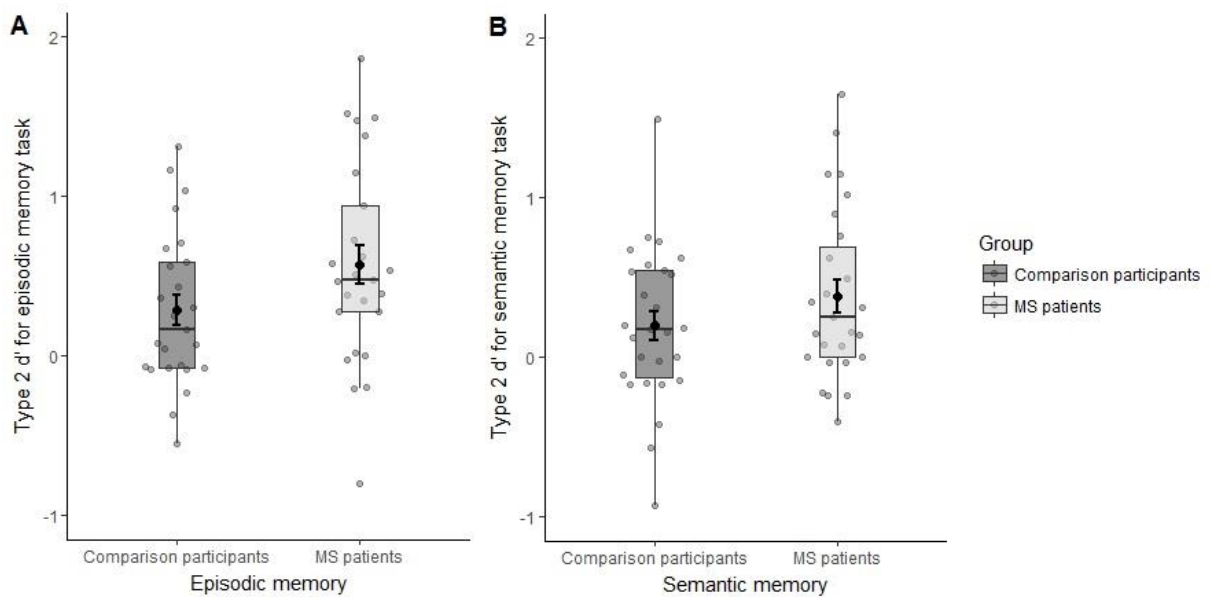
**Figure 1.** The two trials of our version of the SMDT tasks. Participants have to read aloud digits that correspond to the presented symbols as rapidly as possible. They have 45 seconds to read as many digits as they can.



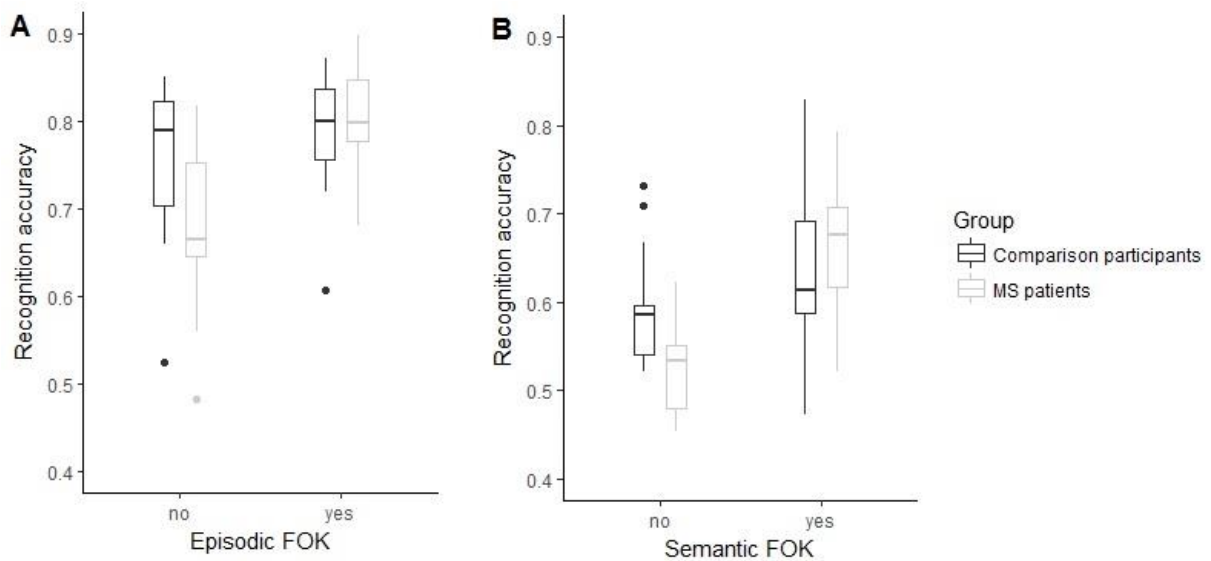
**Figure 2.** Means and standard errors for predictions and postdiction metacognitive accuracy scores according to groups and trials for the SDMT task.



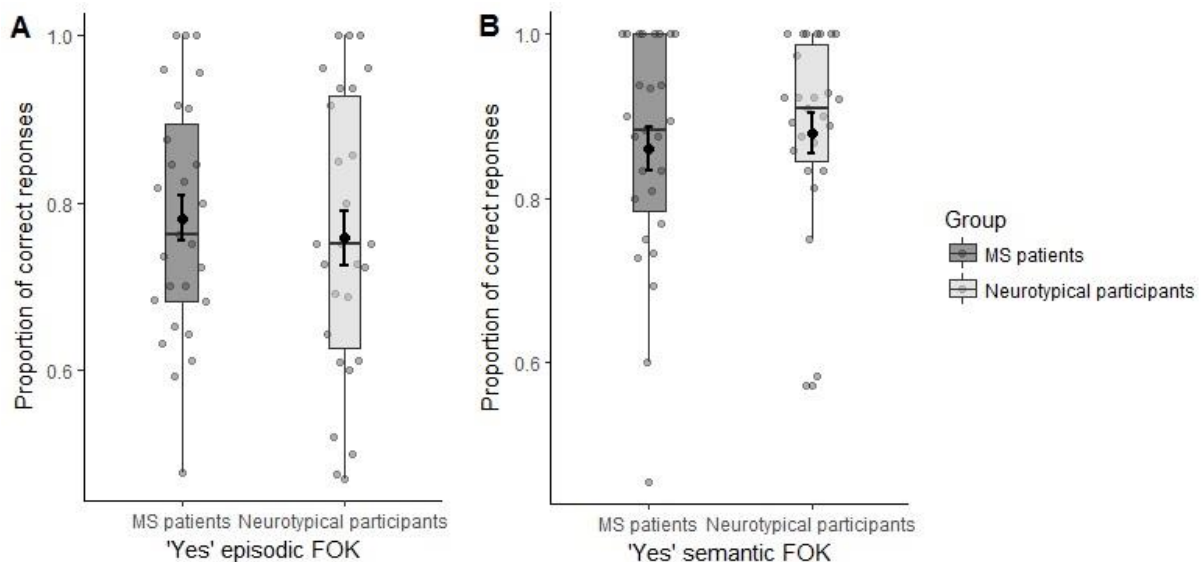
**Figure 3.** Means and standard errors for predictions and postdiction metacognitive accuracy scores according to groups and trials for the fluency task.



**Figure 4.** Individual values, means, and standard errors for metacognitive sensitivity measured by Type 2  $d'$  according to group for the episodic memory task (A) and semantic memory task (B).



**Figure 5.** Boxplots for the mixed logistic regressions between task accuracy in the recognition tasks and confidence in MS patients and comparison group participants for the episodic memory task (A) and semantic memory task (B).



**Figure 6.** Individual values, means, and standard errors for metacognitive bias (proportion of correct responses for 'yes' FOK) according to group for the episodic and semantic memory tasks.

## **Appendix C**

**In the here and now: Short term  
memory predictions are preserved in  
Alzheimer's disease**





## Note

## In the here and now: Short term memory predictions are preserved in Alzheimer's disease



Julie M. Bertrand <sup>a,b,1</sup>, Audrey Mazancieux <sup>a,1</sup>, Chris. J.A. Moulin <sup>a</sup>,  
Yannick Béjot <sup>b</sup>, Olivier Rouaud <sup>c</sup> and Céline Souchay <sup>a,\*</sup>

<sup>a</sup> Laboratoire de Psychologie et Neurocognition, LPNC CNRS 5105, Université Grenoble Alpes, France

<sup>b</sup> EA4184, Medical School of Dijon, Université de Bourgogne, France

<sup>c</sup> Centre Leenaards de la mémoire, Lausanne, Switzerland

## ARTICLE INFO

## Article history:

Received 13 November 2018

Reviewed 1 March 2019

Revised 27 March 2019

Accepted 28 March 2019

Action editor Robin Morris

Published online 2 May 2019

## Keywords:

Metacognition

Short-term memory

Alzheimer's disease

Global predictions

## ABSTRACT

According to neuropsychological models of anosognosia, there is a failure to transfer online awareness of dysfunction into a more generalised long term belief about memory function in Alzheimer's disease. This failure results in specific metamemory deficits for global predictions: patients overestimate their performance before the task but are able to monitor their memory performance after having experienced the task. However, after a delay, they are still not able to make accurate predictions. As previous work has mainly focused on long-term memory, the present study investigates this issue in short-term and working memory. Using both global and item-by-item metacognitive judgements in a digit span task, we showed that Alzheimer's disease patients are as accurate as older adults in monitoring their performance despite impaired memory. When they have the opportunity to test themselves, or when they have already performed the task, patients are able to use feedback to adjust their metacognitive judgements. Overall, these results show that even for a relatively complex task, patients with Alzheimer's disease are aware of their difficulties in the here-and-now.

© 2019 Elsevier Ltd. All rights reserved.

Metamemory, defined as the ability to monitor, regulate and predict one's memory performance (Flavell, Miller & Miller, 2002) has often been explored in Alzheimer's disease (e.g., Souchay, 2007; Sunderaraman & Cosentino, 2017). The literature has focused on patients' ability to predict their memory performance on an upcoming test. Such predictions can occur at an item level (i.e., how people expect to perform for each specific item) or a global level (i.e.,

participants' expectations for their overall test performance). At the item level, metacognitive tasks with episodic and semantic materials have showed a diversity of spared and impaired performance (e.g., Bäckman & Lipinska, 1993; Souchay, Isingrini, & Gil, 2002), supporting the idea that the metamemory impairment in Alzheimer's disease is a consequence of the memory deficit (for a review see Ernst, Moulin, Souchay, Mograbi, & Morris, 2016). Critically,

\* Corresponding author. Laboratoire de Psychologie et Neurocognition, LPNC, Grenoble Alpes University, France.

E-mail address: [celine.souchay@me.com](mailto:celine.souchay@me.com) (C. Souchay).

<sup>1</sup> Joint first authors.

<https://doi.org/10.1016/j.cortex.2019.03.027>

0010-9452/© 2019 Elsevier Ltd. All rights reserved.

experimental metamemory tasks are known to be related to clinical classifications of awareness (e.g., Consentino, et al., 2016).

Fewer studies have explored metamemory in Alzheimer's patients using global judgments but all have shown that patients overestimate their performance initially (Ansell & Bucks, 2006; Barrett, Eslinger, Ballentine, & Heilman, 2005; Moulin, Perfect, & Jones, 2000a; 2000b). These studies also compared predictions made before (prospective judgements) and after the task (retrospective judgements), finding that Alzheimer's patients revise their predictions to accurate levels after having experienced the task (Ansell & Bucks, 2006; Moulin, Perfect, & Jones, 2000a, 2000b; Stewart, McGeown, Shanks, & Venneri, 2010). Thus patients are poorly calibrated but they show intact awareness of their memory performance after having experienced the task. Interestingly, after a delay, people with Alzheimer's disease continue to overestimate their function even after having made accurate evaluations during a task (Silva, Pinho, Macedo, Souchay, & Moulin, 2017): there is a failure to transfer the on-line awareness of dysfunction into a more generalised long-term belief about memory function. Several models have been proposed to explain awareness deficits in Alzheimer's disease such as the Levels of Awareness framework (Clare, Marková, Roth, & Morris, 2011) or the Cognitive Awareness Model (CAM, Morris & Mograbi, 2013). Such models converge on a failure to consolidate their knowledge regarding their memory abilities over a long period; *mnemonic anosognosia*.

To date, the bias in such research has overwhelmingly been towards examining the ability to monitor long term memory. Clearly, however, the global judgements literature reviewed above points to an ability to make accurate judgements in the short term which are not maintained in the long term. Here we sought to directly assess the ability to monitor short term memory, hitherto unexamined in Alzheimer's disease. Short-term memory underpins many activities of daily living, and evaluation this domain is critical. Our experimental design is based on Flavell's original design (Flavell, Friedrichs, & Hoyt, 1970; see also; Murphy, Sanders, Gabrieheski, & Schmitt, 1981) and adapted by Bertrand, Moulin, and Souchay (2017) in a recent study. A novelty of our protocol is that it allows the measure of both global predictions and item-by-item judgements. As it has been suggested by previous work in episodic memory, we hypothesised that Alzheimer's disease patients would be impaired on initial global predictions (made before the task). However, according to the idea that on-line metacognitive processes are intact, they should be preserved for item-by-item judgements and global post-diction because these are based on access to short term representations of task performance. Following standard practice in neuropsychological assessment, we examined both forward and backward span, although we made no specific predictions about differences between the two tasks, although the fact that backwards span is more demanding than forward span may be of interest (although this difference between the two tasks is far from clear-cut, e.g., Hester, Kinsella, & Ong, 2004).

## 1. Method

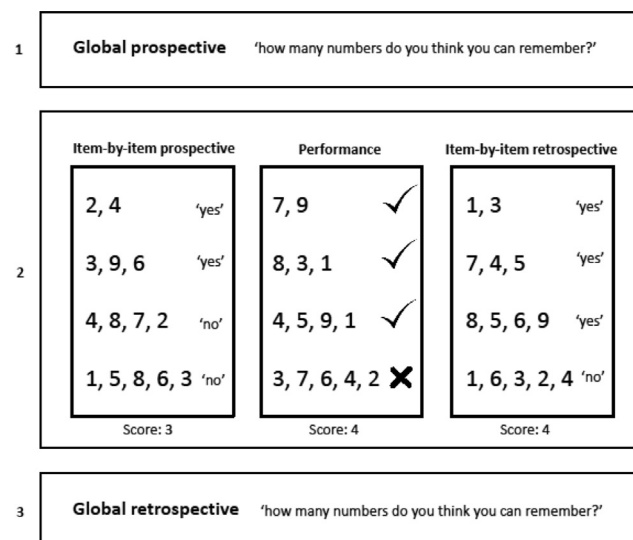
### 1.1. Participants

Twenty-three older adults ( $M_{age} = 73.09$ ,  $SD_{age} = 6.04$ ; 17 females) and eighteen Alzheimer's patients ( $M_{age} = 76.44$ ,  $SD_{age} = 5.89$ ; 6 females) participated in the study. The healthy older adults were recruited from in the local community. Participants were defined as cognitively healthy if they had a mini-mental state exam (MMSE; Folstein, Folstein, & McHugh, 1975) score of 28–30.

All patients were recruited from the Memory Clinic at the Dijon university hospital. Diagnosis was determined by a group of neurologists at the memory clinic. Patients had a MMSE score ranging from 14 to 28 ( $M = 21.67$ ,  $SD = 4.38$ ). Participants were excluded if they had a history of clinical stroke, traumatic brain injury, alcohol or drug abuse or medical/psychiatric condition. The study was approved by the Institutional Review Board of the Dijon Hospital.

### 1.2. Materials and procedure

For the two tasks (i.e., forward digit span and backward digit span), there were three phases (for a summary of the entire procedure see Fig. 1). The first phase was a global prospective judgement task, where participants had to report how many digits they thought they would be able to remember (from 0 to 9). The second phase was an online task where participants gave item-by-item metacognitive judgements for the short-term memory task (either forward digit span or backward digit span). Here, there were two types of judgements: the prospective judgements (made before a trial) and the retrospective judgements (made after the trial). Item-by-item judgements were made after being presented a set of digits of a certain length: they were based on the participant's recent experience of the to-be-



**Fig. 1 – Summary of the procedure. The first phase is a global prospective judgement. The second includes prospective judgements, the actual task (either forward span or backward span) and retrospective judgements. The third phase is a global retrospective judgement.**

remembered digits. Participants were presented with first one series of digits and then asked whether they would be able to recall the number by giving a Yes/No answer. The number of digits increased sequentially (to a maximum of 9), until the participants said 'No'. For example, if after the presentation of 4 items a participant decided to say 'No', the item-by-item judgement stopped, with the participants therefore predicting having a span of 3. Note that in this paradigm, performance and predictions are not taken for the same trial. Rather, there is blocked presentation, such that a first set of digit spans are used to make the item-by-item prospective judgements, then there is the block of digit spans where recall is measured in the standard fashion, and finally a third block where the retrospective item-by-item predictions are made (see Fig. 1). For the span tasks where recall was measured, the digit forward and backward span tasks from the Wechsler Adult Intelligence Scale-IV (WAIS-IV, Wechsler, 2011) were used. For the item-by-item judgements, lists of numbers matched in length with the span task were created and these differed for the two judgements. Lists were counterbalanced across judgements and participants. The third phase was a global retrospective judgement task. As in the first phase, participants had to say how many digits they thought they would be able to remember.

### 1.3. Data analyses

Analyses were conducted in R and included linear regressions (yielding  $t$  statistics which can be interpreted exactly as factorial ANOVAs) and Pearson's correlations. We use a standard analysis protocol, starting by examining the mean recall (span performance) and the mean prediction values (prediction magnitude). Then, we focus on metacognitive accuracy in a standard fashion, considering the non-directional discrepancy between prediction and performance (e.g., Moulin et al., 2000a). This procedure allows the estimation of metacognitive accuracy independently from the bias (underestimation or overestimation of performance). A score of zero reflects perfect accuracy and the higher the score, the bigger the discrepancy between the metacognitive judgement and the performance (see Table 1). [Following publication we will make the dataset and analysis script available on-line. Data and script are part of the submission.]

## 2. Results

### 2.1. Span performance

The mean span performance for each group is found in Table 1. We conducted linear regressions with group as a between-

subjects factor and task as a within-subjects factor. As expected, we found a main effect of group, revealing that Alzheimer's patients have a lower performance than older adult controls,  $t(39) = 3.28, p = .002, d = 1.03$ . We also found a main effect of task. Performance on the forward span was higher than performance on the backward span,  $t(39) = 6.20, p < .001, d_z = .97$ , i.e., spans were significantly longer for forwards rather backwards recall. There was no interaction between the two factors,  $t(39) = 1.36, p = .180, d = .43$ .

### 2.2. Magnitude of metacognitive judgements

**Forward span.** The mean values of predictions are given in Table 1. We conducted linear regressions with, group, prediction time (prospective vs retrospective), and judgement type (global vs item-by-item) as factors. Analyses revealed a significant effect of group,  $t(39) = 2.61, p = .013, d = .82$ , and a non-significant trend of judgement type,  $t(39) = 1.82, p = .077, d_z = .28$ . Patients made lower judgements overall than controls therefore predict having a lower span, which is appropriate given the differences in performance reported above. The analysis showed neither a significant effect of prediction time,  $t(39) = 1.12, p = .269, d_z = .17$ , and we found no significant interactions [all  $t(39) < 1.43$ , and all  $p > .05$ ].

**Backward span.** The mean values of predictions are given in Table 1. As for forward span task, we conducted linear regressions with, group, prediction time (prospective vs retrospective), and judgement type (global vs item-by-item) as factors. Analyses revealed only that patients have a non-significant trend for lower judgements than older adults,  $t(39) = 1.85, p = .072, d = .58$ . The analysis again showed neither a significant effect of prediction time,  $t(39) = 1.33, p = .192, d_z = .21$ , nor a significant effect of judgement type,  $t(39) = 1.30, p = .203, d_z = .20$ . There were no significant interactions [all  $t(39) < 1.29$ , and all  $p > .05$ ]. Whereas we found significant differences in magnitude of predictions for forward spans, no such pattern was observed for the backwards span.

### 2.3. Metacognitive accuracy

**Forward span.** We conducted linear regressions with, group, prediction time (prospective vs retrospective), and judgement type (global vs item-by-item) as factors. These analyses showed neither a significant effect of group,  $t(39) = 1.09, p = .284, d = .34$ , nor a significant effect of prediction time,  $t(39) = 1.76, p = .086, d_z = .27$ . We did however, find a main effect of judgement type,  $t(39) = 2.18, p = .036, d_z = .34$ . Global judgements were less accurate than item-by-item judgements. There was also a significant interaction between judgement

**Table 1 – Mean and standard errors for global judgements, item-by-item judgements, and performance for AD patients and older adults.**

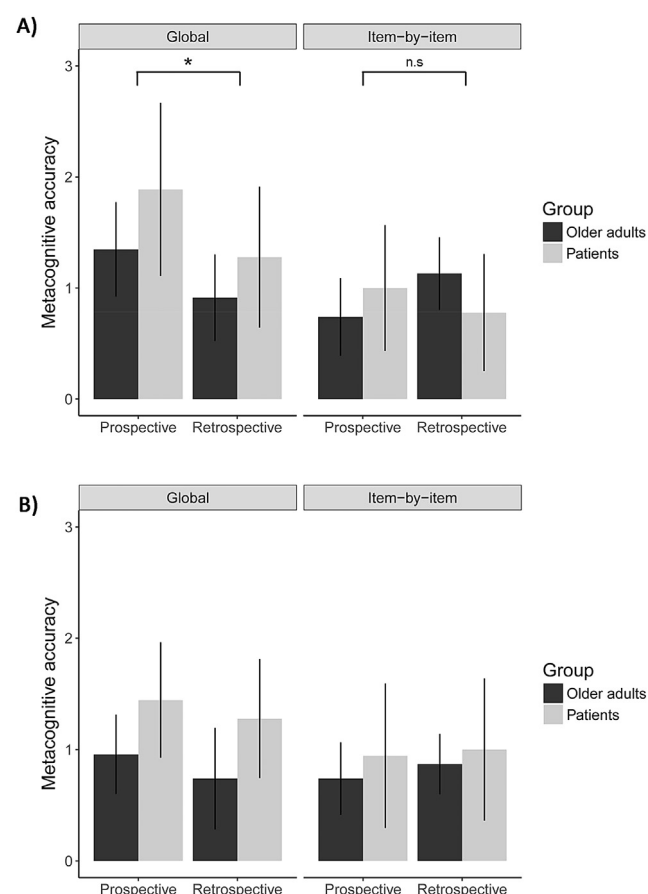
	Forward span		Backward span	
	Older adults	AD patients	Older adults	AD patients
Global prospective	5.17 (1.03)	4.78 (2.02)	4.35 (.78)	3.89 (1.81)
Item-by-item prospective	5.96 (1.26)	4.67 (1.88)	4.48 (.95)	3.83 (2.04)
Performance	6.00 (1.09)	4.67 (1.14)	4.61 (1.31)	3.78 (.94)
Item-by-item retrospective	6.26 (1.60)	4.89 (2.03)	4.96 (1.11)	4.00 (1.50)
Global retrospective	5.61 (1.26)	4.61 (2.00)	4.39 (.84)	3.83 (1.62)

type and judgement time,  $t(39) = 2.17, p = .036, d_z = .34$ . Retrospective judgements are more accurate for global predictions,  $t(39) = 2.41, p = .021, d_z = .38$ , which is not the case for item-by-item judgements,  $t(39) = -.56, p = .582, d_z = -.08$ . No other interaction was significant [all  $t(39) < 1.59$ , and all  $p > .05$ ]. As there was no group difference, these results show that patients are as accurate as controls at predicting their short-term memory performance (Fig. 2A).

**Backward span.** As for forward span task, we conducted linear regressions with, group, prediction time (prospective vs retrospective), and judgement type (global vs item-by-item) as factors. The analysis showed neither a significant effect of group,  $t(39) = 1.47, p = .149, d = .46$ , nor a significant effect of judgement time,  $t(39) = .51, p = .614, d_z = .08$ , nor a significant effect of judgement type,  $t(39) = 1.45, p = .156, d_z = .23$ . No interaction was significant [all  $t(39) < 1.23$ , and all  $p > .05$ ]. For the backward span task, these results show that patients are as accurate as controls at predicting their performance (Fig. 2B).

#### 2.4. Correlational analyses

In order to examine the accuracy at the group level, we analyzed the correlations between the metacognitive



**Fig. 2 – (A) Mean and confidence intervals for metacognitive accuracy according to judgement type, judgement time, and group for the forward span task. (B) Mean and confidence intervals for metacognitive accuracy according to judgement type, judgement time, and group for the backward span task.**

judgements and digit span tasks. In these analyses, individuals' predictions are correlated with individuals' performance, such that as a group, we can see if those people with lower predictions actually have a worse performance (see Connor, Dunlosky & Hertzog, 1997). Table 2 shows that this relationship is always positive except for the global prospective judgements in both tasks where correlations are not significant. The same analysis can be carried out for each group individually, to compare these correlations for AD patients and older adults. Tables 3 and 4 show that this relationship is overall positive between judgement and performance (although not always significant) except for the global prospective judgements in the forward span where correlations coefficients are near 0. Moreover, there was no difference in the magnitude of correlation across AD patients and older adults (all  $z$  value  $< |1.96|$ ).

### 3. Discussion

Previous studies of metacognition in Alzheimer's disease have focused on long term memory. Here we investigated the awareness of short term memory and working memory. We replicated the documented deficits in both digit span forward and digit span backward in Alzheimer's disease (e.g., Morris & Baddeley, 1988). In addition, we showed that people with Alzheimer's disease are as accurate as controls at assessing this function, despite the deficit in performance. To consider the importance of this finding for our understanding of metacognition and anosognosia in Alzheimer's disease, we must consider how participants are able to make accurate judgements on these tasks.

Regarding item-by-item predictions, we propose that when given the digits to memorise, even in the 'dry-run' prediction phase, participants test themselves. People with Alzheimer's disease run through the digits presented, as do controls, and have access to whether or not they will be able to complete the task. Because there are no dual demands of performance and prediction, participants are able to directly report this information: in the here-and-now they can accurately gauge whether they can retain (or retain and manipulate, in the case of digits backward) the information. For retrospective judgements, where there is a preservation in Alzheimer's disease in long term memory (e.g., Gallo, Cramer, Wong, & Bennett, 2012; Moulin, James, Perfect, & Jones, 2003), patients are able to use correctly the feedback arising from this self-test to make accurate predictions.

Turning to global predictions, we found that for the forward span the first prediction is less accurate than the retrospective one and item-by-item judgements (for both patients and older adults). This effect is also typically observed in long term memory tasks for Alzheimer patients (e.g., Silva et al., 2017). Moreover, correlational analyses at the group level bring additional evidence to this. It has been shown in both Alzheimer's disease (Silva et al., 2017) and with older adults (Connor, Dunlosky, & Hertzog, 1997) very low correlations between initial global predictions and performance. This was not the case for later retrospective predictions. Thus, when they can experience the task, both older adults and Alzheimer's disease patients update their knowledge about the

task and make accurate judgements. Although results are clear for the forward span task (i.e., significant difference and very low correlation,  $r = .06$ ), this is less the case for the backward digit span. We did not find a significant difference in terms of accuracy for this task, and correlational analyses showed no significant relationship between global prospective judgement and performance. However, this low correlation was not different from the other three (Fisher's  $z$ , all  $p > .05$ ). The backwards digit span data do not therefore follow exactly the pattern of overestimation and initial inaccuracy found in the Alzheimer's group on previous long term memory tasks and here in our own experiment. Critically, we find no evidence for any group differences in accuracy or magnitude of predictions on this task either. We might hypothesise in general that people anticipate the backwards digit span is a difficult task.

It is important to discuss the large variability for accuracy in patients. It is possible that more patients overestimate their performance (see [supplementary results](#)). We counted the number of participants who overestimated and indeed found that patients overestimate more than controls for the prospective global prediction. For the forward span, 33% of patients overestimate their performance compared to 17% for controls. The same is observed for backward span, 56% of patients overestimate their performance compared to 26% for controls.

Overall, patients and controls have the same judgement accuracy distributions (see [supplementary results](#)). This result has implications for metacognition and anosognosia more generally in Alzheimer's disease. Despite having a deficit in short-term and working memory, the patients with Alzheimer's are nonetheless able to reliably report their difficulties with this task: the magnitude of judgements is different from controls (although a trend for the backward

**Table 2 – Bonferroni corrected correlations between metacognitive judgements and performance for both forward digit span and backward digit span. As there are 4 correlations per tasks the significance threshold is equal to  $.05/4 = .013$ .**

	Forward span	Backward span
Global prospective	$r = .06, p = .727$	$r = .30, p = .057$
Item-by-item prospective	$r = .66, p < .001$	$r = .45, p = .003$
Item-by-item retrospective	$r = .72, p < .001$	$r = .49, p = .001$
Global retrospective	$r = .47, p = .002$	$r = .43, p = .005$
Significant correlations are in bold.		

**Table 3 – Bonferroni corrected correlations between metacognitive judgements and performance for the forward digit span. As there are 4 correlations for each group the significance threshold is equal to  $.05/4 = .013$ .**

	Forward span		
	AD patients	Older adults	z value
Global prospective	$r = -.03, p = .914$	$r = .04, p = .854$	-.21
Item-by-item prospective	$r = .60, p = .009$	$r = .56, p = .005$	.18
Item-by-item retrospective	$r = .47, p = .049$	$r = .20, p = .362$	.93
Global retrospective	$r = .77, p < .001$	$r = .55, p = .007$	1.21
Significant correlations are in bold.			

**Table 4 – Bonferroni corrected correlations between metacognitive judgements and performance for the backward digit span. As there are 4 correlations for each group the significance threshold is equal to  $.05/4 = .013$ .**

	Backward span		
	AD patients	Older adults	z value
Global prospective	$r = .26, p = .297$	$r = .37, p = .087$	-.37
Item-by-item prospective	$r = .28, p = .256$	$r = .60, p = .003$	-1.22
Item-by-item retrospective	$r = .56, p = .015$	$r = .36, p = .098$	.77
Global retrospective	$r = .17, p = .508$	$r = .65, p < .001$	-1.82
Significant correlations are in bold.			

span). This is in direct contrast with tasks which require memory retrieval. On (long term) episodic memory feeling of knowing tasks (e.g., [Souchay et al., 2002](#)), patients are unable to reliably gauge whether a previously studied word is available or not when tested by recognition. This is proposed to be due to the impoverished information available to the person with Alzheimer's disease: they cannot evaluate their memory accurately, because they cannot retrieve from memory enough pertinent information on which to base their judgement. In contrast, even for a relatively complex task, such as reversing and repeating a series of digits as tested here, in the here-and-now patients with Alzheimer's disease are aware of their difficulties.

Taken together, these results support the idea of a preservation of online monitoring in Alzheimer's disease. When they can test themselves or when they have already performed the task, both older adults and patients are able to use the performance feedback to adjust their metacognitive judgements. Naturally, this has major clinical implications. On-line, whilst struggling with a task, a patient with Alzheimer's disease will be aware of their difficulties, even if when asked later they are not aware of how difficult the task was, or indeed, when encountering the same task again, they will not beforehand know how difficult they will find it. It would be important to replicate the likely deficit for patients in initial global prediction and to add a second trial after a delay. If this impairment is also found for a second trial, this would be in line with the failure to transfer information from online evaluations into long-term representations ([Morris & Mograbi, 2013](#)). Interestingly, [Stewart et al. \(2010\)](#) show that whilst global judgements may be accurate for long term memory tasks, the accuracy that is acquired is forgotten as soon as one

hour later. We add another task to the literature for which people with Alzheimer's can accurately gauge their performance. The impact of this work is that people with Alzheimer's are able to reflect upon their performance in a task which is critical for daily function: short term memory.

Anosognosia, however, is likely to remain a multifaceted construct, with varying causes and manifestations. Whilst it is clear memory mechanisms are pertinent to tasks which involve memory, different domains should be compared (Chapman et al., 2018) and the involvement of other process such as executive function (Scherling, Wilkins, Zakrezewski, et al., 2016) perspective taking (Serino & Riva, 2017), and emotion [need to be examined in detail (Bertrand et al., 2016)].

---

### Conflict of interest

The authors have no conflict of interest to report.

---

### CRedit authorship contribution statement

**Julie M. Bertrand:** Conceptualization, Data curation, Investigation, Methodology, Project administration, Writing - original draft. **Audrey Mazancieux:** Data curation, Formal analysis, Software, Visualization, Writing - original draft, Writing - review & editing. **Chris J.A. Moulin:** Conceptualization, Writing - original draft, Writing - review & editing. **Yannick Béjot:** Resources, Investigation, Methodology, Supervision. **Olivier Rouaud:** Resources, Funding acquisition, Investigation, Methodology, Supervision. **Céline Souchay:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing - original draft.

---

### Acknowledgments

This work was supported by a research studentship given to Julie Bertrand, provided by the Region of Bourgogne (FABER), "Conscience et Mémoire dans la Maladie d'Alzheimer" awarded to Céline Souchay and Chris Moulin. Chris Moulin is supported by the Institut Universitaire de France.

---

### Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cortex.2019.03.027>.

---

### REFERENCES

- Ansell, E. L., & Bucks, R. S. (2006). Mnemonic anosognosia in Alzheimer's disease: A test of Agnew and Morris (1998). *Neuropsychologia*, 44(7), 1095–1102.
- Bäckman, L., & Lipinska, B. (1993). Monitoring of general knowledge: Evidence for preservation in early Alzheimer's disease. *Neuropsychologia*, 31(4), 335–345.
- Barrett, A. M., Eslinger, P. J., Ballentine, N. H., & Heilman, K. M. (2005). Unawareness of cognitive deficit (cognitive anosognosia) in probable AD and control subjects. *Neurology*, 64(4), 693–699.
- Bertrand, E., Dourado, M. C., Laks, J., Morris, R. G., Landeira-Fernandez, J., & Mograbi, D. C. (2016). Mood-congruent recollection and anosognosia in Alzheimer's disease. *Cortex*, 84, 55–62.
- Bertrand, J. M., Moulin, C. J. A., & Souchay, C. (2017). Short-term memory predictions across the lifespan: Monitoring span before and after conducting a task. *Memory*, 25(5), 607–618.
- Chapman, S., Colvin, L. E., Vuorre, M., Cocchini, G., Metcalfe, J., Huey, E. D., et al. (2018). Cross domain self-monitoring in anosognosia for memory loss in Alzheimer's disease. *Cortex*, 101, 221–233.
- Clare, L., Marková, I. S., Roth, I., & Morris, R. G. (2011). Awareness in Alzheimer's disease and associated dementias: Theoretical framework and clinical implications. *Aging & Mental Health*, 15(8), 936–944.
- Cosentino, S., Zhu, C., Bertrand, E., Metcalfe, J., Janicki, S., & Cines, S. (2016). Examination of the metacognitive errors that contribute to anosognosia in Alzheimer's disease. *Cortex*, 84, 101–110.
- Ernst, A., Moulin, C. J., Souchay, C., Mograbi, D. C., & Morris, R. (2016). Anosognosia and metacognition in Alzheimer's disease: Insights from experimental psychology. *The Oxford Handbook of Metamemory*, 1, 456–479.
- Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology*, 1(4), 324–340.
- Flavell, D. M., Jamshidi, Y., Hawe, E., Pineda Torra, I., Taskinen, M. R., Frick, M. H., et al. (2002). Peroxisome proliferator-activated receptor  $\alpha$  gene variants influence progression of coronary atherosclerosis and risk of coronary artery disease. *Circulation*, 105(12), 1440–1445.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189–198.
- Gallo, D. A., Cramer, S. J., Wong, J. T., & Bennett, D. A. (2012). Alzheimer's disease can spare local metacognition despite global anosognosia: Revisiting the confidence–accuracy relationship in episodic memory. *Neuropsychologia*, 50(9), 2356–2364.
- Hester, R. L., Kinsella, G. J., & Ong, B. E. N. (2004). Effect of age on forward and backward span tasks. *Journal of the International Neuropsychological Society*, 10(4), 475–481.
- Morris, R. G., & Baddeley, A. D. (1988). Primary and working memory functioning in Alzheimer-type dementia. *Journal of Clinical and Experimental Neuropsychology*, 10(2), 279–296.
- Morris, R. G., & Mograbi, D. C. (2013). Anosognosia, autobiographical memory and self knowledge in Alzheimer's disease. *Cortex*, 49(6), 1553–1565.
- Moulin, C. J., James, N., Perfect, T. J., & Jones, R. W. (2003). Knowing what you cannot recognise: Further evidence for intact metacognition in Alzheimer's disease. *Aging, Neuropsychology, and Cognition*, 10(1), 74–82.
- Moulin, C. J., Perfect, T. J., & Jones, R. W. (2000a). Global predictions of memory in Alzheimer's disease: Evidence for preserved metamemory monitoring. *Aging, Neuropsychology, and Cognition*, 7(4), 230–244.
- Moulin, C. J., Perfect, T. J., & Jones, R. W. (2000b). Evidence for intact memory monitoring in Alzheimer's disease: Metamemory sensitivity at encoding. *Neuropsychologia*, 38(9), 1242–1250.
- Murphy, M., Sanders, R., Gabrielseski, A., & Schmitt, F. (1981). Metamemory in the aged. *Journal of Gerontology*, 36(2), 185–193.

- Scherling, C. S., Wilkins, S. E., Zakrezewski, J., Kramer, J. H., Miller, B. L., Weiner, M. W., et al. (2016). Decreased self-appraisal accuracy on cognitive tests of executive functioning is a predictor of decline in mild cognitive impairment. *Frontiers in Aging Neuroscience*, 8, 120.
- Serino, S., & Riva, G. (2017). The proactive self in space: How egocentric and allocentric spatial impairments contribute to anosognosia in Alzheimer's Disease. *Journal of Alzheimer's Disease*, 55(3), 881–892.
- Silva, A. R., Pinho, M. S., Macedo, L., Souchay, C., & Moulin, C. (2017). Mnemonic anosognosia in Alzheimer's disease is caused by a failure to transfer online evaluations of performance: Evidence from memory training programs. *Journal of Clinical and Experimental Neuropsychology*, 39(5), 419–433.
- Souchay, C. (2007). Metamemory in Alzheimer's disease. *Cortex*, 43(7), 987–1003.
- Souchay, C., Isingrini, M., & Gil, R. (2002). Alzheimer's disease and feeling-of-knowing in episodic memory. *Neuropsychologia*, 40(13), 2386–2396.
- Stewart, G., McGeown, W. J., Shanks, M. F., & Venneri, A. (2010). Anosognosia for memory impairment in Alzheimer's disease. *Acta Neuropsychiatrica*, 22(4), 180–187.
- Sunderaraman, P., & Cosentino, S. (2017). Integrating the constructs of anosognosia and metacognition: A review of recent findings in dementia. *Current Neurology and Neuroscience Reports*, 17(3), 17–27.
- Wechsler, D. (2011). WASI-II: Wechsler abbreviated scale of intelligence. PsychCorp.

## References

- Ackerman, R., & Zalmanov, H. (2012). The persistence of the fluency–confidence association in problem solving. *Psychonomic bulletin & review*, *19*(6), 1187–1192.
- Agnew, S. K., & Morris, R. (1998). The heterogeneity of anosognosia for memory impairment in alzheimer’s disease: A review of the literature and a proposed model. *Aging & Mental Health*, *2*(1), 7–19.
- Ais, J., Zylberberg, A., Barttfeld, P., & Sigman, M. (2016). Individual consistency in the accuracy and distribution of confidence judgments. *Cognition*, *146*, 377–386. doi: 10.1016/j.cognition.2015.10.006
- Alban, M. W., & Kelley, C. M. (2013). Embodiment meets metamemory: Weight as a cue for metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1628–1634.
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome open research*, *4*.
- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and social psychology review*, *13*(3), 219–235.
- Ansell, E. L., & Bucks, R. S. (2006). Mnemonic anosognosia in Alzheimer’s disease: A test of Agnew and Morris (1998). *Neuropsychologia*, *44*(7), 1095–1102. doi: 10.1016/j.neuropsychologia.2005.10.019
- Arango-Muñoz, S. (2011). Two levels of metacognition. *Philosophia*, *39*(1), 71–82.
- Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of experimental psychology*, *81*(1), 126–131.
- Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in brain research*, *150*, 45–53.
- Bach, D. R., & Dolan, R. J. (2012). Knowing how much you don’t know: a neural organization of uncertainty estimates. *Nature reviews neuroscience*, *13*(8), 572–586.
- Bacon, E., Danion, J.-M., Kauffmann-Muller, F., & Bruant, A. (2001). Consciousness in schizophrenia: a metacognitive approach to semantic memory. *Consciousness and*



- cognition*, 10(4), 473–484.
- Bacon, E., Danion, J.-M., Kauffmann-Muller, F., Schelstraete, M.-A., Bruant, A., Sellal, F., & Grange, D. (1998). Confidence level and feeling of knowing for episodic and semantic memory: an investigation of lorazepam effects on metamemory. *Psychopharmacology*, 138(3-4), 318–325.
- Baird, B., Cieslak, M., Smallwood, J., Grafton, S. T., & Schooler, J. W. (2015). Regional White Matter Variation Associated with Domain-specific Metacognitive Accuracy. *Journal of Cognitive Neuroscience*, 27(3), 440–452. doi: 10.1162/jocn\_a\_00741
- Baird, B., Smallwood, J., Gorgolewski, K. J., & Margulies, D. S. (2013). Medial and Lateral Networks in Anterior Prefrontal Cortex Support Metacognitive Ability for Memory and Perception. *Journal of Neuroscience*, 33(42), 16657–16665. doi: 10.1523/JNEUROSCI.0786-13.2013
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2), 191–215.
- Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & psychophysics*, 55(4), 412–428.
- Baranski, J. V., & Petrusic, W. M. (1995). On the calibration of knowledge and perception. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 49(3), 397.
- Barrett, A. B., Dienes, Z., & Seth, A. K. (2013). Measures of metacognition on signal-detection theoretic models. *Psychological methods*, 18(4), 535–552.
- Barthelme, S., & Mamassian, P. (2010). Flexible mechanisms underlie the evaluation of visual confidence. *Proceedings of the National Academy of Sciences*, 107(48), 20834–20839. doi: 10.1073/pnas.1007704107
- Beatty, W. W., & Monson, N. (1991). Metamemory in multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, 13(2), 309–327.
- Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? a meta-analysis. *Psychological bulletin*, 137(2), 211–241.
- Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & cognition*, 31(2), 297–305.
- Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In *Implicit memory and metacognition* (pp. 321–350). Psychology Press.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, 127(1), 55–68.

- Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In *Handbook of metamemory and memory* (pp. 73–94). Taylor and Francis.
- Beran, M. J., Brandl, J. L., Perner, J., & Proust, J. (2012). On the nature, evolution, development, and epistemology of metacognition: introductory thoughts. In *The foundations of metacognition* (pp. 289–301). Oxford University Press.
- Besken, M., & Mulligan, N. W. (2014). Perceptual fluency, auditory generation, and metamemory: Analyzing the perceptual fluency hypothesis in the auditory modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 429–440. doi: 10.1037/a0034407
- Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in cognitive sciences*, 15(12), 567–575.
- Boduroglu, A., Pehlivanoglu, D., Tekcan, A. İ., & Kapucu, A. (2015). Effects of self-referencing on feeling-of-knowing accuracy and recollective experience. *Memory*, 23(5), 736–747.
- Borkowski, J. G., Carr, M., Rellinger, E., Pressley, M., et al. (1990). Self-regulated cognition: Interdependence of metacognition, attributions, and self-esteem. *Dimensions of thinking and cognitive instruction*, 1, 53–92.
- Bornstein, R. F., & D’agostino, P. R. (1992). Stimulus recognition and the mere exposure effect. *Journal of personality and social psychology*, 63(4), 545–552.
- Bosco, A., Lancioni, G. E., Belardinelli, M. O., Singh, N. N., O’Reilly, M. F., & Sigafos, J. (2009). Learning as a possible sign of non-reflective consciousness in persons with a diagnosis of vegetative state and pervasive motor impairment. *Cognitive processing*, 10(4), 355.
- Brown, R., & McNeill, D. (1966). The “tip of the tongue” phenomenon. *Journal of verbal learning and verbal behavior*, 5(4), 325–337.
- Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General*, 148(1), 51–64.
- Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and brain sciences*, 32(2), 121–138.
- Castel, A. D., Middlebrooks, C. D., & McGillivray, S. (2016). Monitoring memory in old age: Impaired, spared, and aware. In *The oxford handbook of metamemory* (pp. 519–535). Oxford library of psychology.
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among amazon mechanical turk workers: Consequences and solutions for behavioral researchers. *Behavior*

- research methods*, 46(1), 112–130.
- Chapman, S., Colvin, L. E., Vuorre, M., Cocchini, G., Metcalfe, J., Huey, E. D., & Cosentino, S. (2018). Cross domain self-monitoring in anosognosia for memory loss in alzheimer's disease. *Cortex*, 101, 221–233.
- Charles, L., Van Opstal, F., Marti, S., & Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage*, 73, 80–94.
- Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 40(4), 343–367.
- Chua, E. F., Schacter, D. L., & Sperling, R. A. (2009). Neural correlates of metamemory: a comparison of feeling-of-knowing and retrospective confidence judgments. *Journal of Cognitive Neuroscience*, 21(9), 1751–1765.
- Clarys, D., Isingrini, M., & Gana, K. (2002). Mediators of age-related differences in recollective experience in recognition memory. *Acta psychologica*, 109(3), 315–329.
- Davis, K., Christodoulou, J., Seider, S., & Gardner, H. E. (2011). The theory of multiple intelligences. Davis, K., Christodoulou, J., Seider, S., & Gardner, H. (2011). *The theory of multiple intelligences*. In R.J Sternberg & S.B Kaufman (Eds.), *Cambridge Handbook of Intelligence*, 485–503.
- de Gardelle, V., Le Corre, F., & Mamassian, P. (2016). Confidence as a Common Currency between Vision and Audition. *Plos One*, 11(1), e0147901. doi: 10.1371/journal.pone.0147901
- de Gardelle, V., & Mamassian, P. (2014). Does confidence use a common currency across two visual tasks? *Psychological science*, 25(6), 1286–1288.
- Dehaene, S., Sergent, C., & Changeux, J.-P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences*, 100(14), 8520–8525.
- De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature neuroscience*, 16(1), 105–110.
- de Sousa, R. (2008). Epistemic feelings. In *Epistemology and emotions* (pp. 139–161). London: Routledge.
- Dienes, Z. (2007). Subjective measures of unconscious knowledge. *Progress in brain research*, 168, 49–269.
- Dienes, Z., Altmann, G., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning*,

- Memory, and Cognition*, 21(5), 1322–1338.
- Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and brain sciences*, 22(5), 735–808.
- Dixon, R. A., & Hultsch, D. F. (1983). Structure and development of metamemory in adulthood. *Journal of Gerontology*, 38(6), 682–688.
- Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory & Cognition*, 25(2), 129–139.
- Dokic, J. (2014). Feeling the past: A two-tiered account of episodic memory. *Review of Philosophy and Psychology*, 5(3), 413–426.
- Dougherty, M. R., Robey, A. M., & Buttaccio, D. (2018). Do metacognitive judgments alter memory performance beyond the benefits of retrieval practice? A comment on and replication attempt of Dougherty, Scheck, Nelson, and Narens (2005). *Memory & Cognition*, 46(4), 558–565.
- Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Memory & cognition*, 33(6), 1096–1115.
- Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and aging*, 15(3), 462–474.
- Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic–extrinsic distinction of judgments of learning (jols): Effects of relatedness and serial position on jols. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(5), 1180–1191.
- Dunlosky, J., Mueller, M. L., & Tauber, S. K. (2014). The contribution of processing fluency (and beliefs) to people’s judgments of learning. In *Remembering* (pp. 70–88). Psychology Press.
- Dunlosky, J., Mueller, M. L., & Thiede, K. W. (2018). Methodology for investigating human metamemory. In *The oxford handbook of metamemory* (pp. 491–518). Oxford University Press.
- Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (jols) to the effects of various study activities depend on when the jols occur? *Journal of Memory and Language*, 33(4), 545–565.
- Eakin, D. K., Hertzog, C., & Harris, W. (2014). Age invariance in semantic and episodic metamemory: Both younger and older adults provide accurate feeling-of-knowing for names of faces. *Aging, Neuropsychology, and Cognition*, 21(1), 27–51.
- Elman, J. A., Klostermann, E. C., Marian, D. E., Verstaen, A., & Shimamura, A. P. (2012).

- Neural correlates of metacognitive monitoring during episodic and semantic retrieval. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(3), 599–609.
- Ernst, A., Moulin, C., Souchay, C., Mograbi, D., & Morris, R. (2016). Anosognosia and metacognition in alzheimer's disease: Insights from experimental psychology. In *The oxford handbook of metamemory* (pp. 451–472). Oxford University Press.
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.*, *59*, 255–278.
- Faivre, N., Filevich, E., Solovey, G., Kühn, S., & Blanke, O. (2018). Behavioral, Modeling, and Electrophysiological Evidence for Supramodality in Human Metacognition. *The Journal of Neuroscience*, *38*(2), 263–277.
- Faivre, N., Vuillaume, L., Bernasconi, F., Salomon, R., Blanke, O., & Cleeremans, A. (2020). Sensorimotor conflicts alter metacognitive and action monitoring. *Cortex*, *124*, 224–234.
- Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 238–244. doi: 10.1037/0278-7393.33.1.238
- Fitzgerald, L. M., Arvaneh, M., & Dockree, P. M. (2017). Domain-specific and domain-general processes underlying metacognitive judgments. *Consciousness and Cognition*, *49*, 264–277. doi: 10.1016/j.concog.2017.01.011
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, *34*(10), 906.
- Fleming, S. M. (2017). HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, *2017*(1). doi: 10.1093/nc/nix007
- Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 91–114. doi: 10.1037/rev0000045
- Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012). Metacognition: computation, biology and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1280–1286. doi: 10.1098/rstb.2012.0021
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*. doi: 10.3389/fnhum.2014.00443
- Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-specific disruption of perceptual confidence. *Psychological science*, *26*(1), 89–98.
- Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2016). Metacognition about

- the past and future: quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neuroscience of Consciousness*, 2016(1), niw018. doi: 10.1093/nc/niw018
- Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*, 137(10), 2811–2822. doi: 10.1093/brain/awu221
- Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, 329(5998), 1541–1543.
- Fodor, J. A. (1983). *The modularity of mind*. MIT press.
- Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(5), 680–693.
- Frensch, P. A., & Buchner, A. (1999). Domain-general versus domain-specificity in cognition. In *The nature of cognition* (pp. 137–172). MIT Press.
- Frith, C. D. (2019). The neural basis of consciousness. *Psychological medicine*, 1–13.
- Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review*, 10(4), 843–876.
- Gelman, A., & Rubin, D. B. (1992). A single series from the gibbs sampler provides a false sense of security. *Bayesian statistics*, 4, 625–631.
- Geurten, M., Meulemans, T., & Lemaire, P. (2018). From domain-specific to domain-general? the developmental path of metacognition for strategy selection. *Cognitive Development*, 48, 62–81.
- Gherman, S., & Philiastides, M. G. (2015). Neural representations of confidence emerge from the process of decision formation during perceptual choices. *Neuroimage*, 106, 134–143.
- Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain sciences*, 16(1), 1–14.
- Goupil, L., & Kouider, S. (2016). Behavioral and neural indices of metacognitive sensitivity in preverbal infants. *Current Biology*, 26(22), 3038–3045.
- Green, D. M., Swets, J. A., et al. (1966). *Signal detection theory and psychophysics* (Vol. 1). Wiley New York.
- Greifeneder, R., Bless, H., & Pham, M. T. (2011). When Do People Rely on Affective and Cognitive Feelings in Judgment? A Review. *Personality and Social Psychology*

- Review*, 15(2), 107–141. doi: 10.1177/1088868310367640
- Grimaldi, P., Lau, H., & Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. *Neuroscience & Biobehavioral Reviews*, 55, 88–97. doi: 10.1016/j.neubiorev.2015.04.006
- Ha, M., Haury, D. L., & Nehm, R. H. (2012). Feeling of certainty: Uncovering a missing link between knowledge and acceptance of evolution. *Journal of Research in Science Teaching*, 49(1), 95–121.
- Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of educational psychology*, 56(4), 208–216.
- Harvey, N. (1997). Confidence in judgment. *Trends in cognitive sciences*, 1(2), 78–82.
- Hertzog, C., & Dixon, R. A. (1994). Metacognitive development in adulthood and old age. In *Metacognition: Knowing about knowing* (pp. 227–251).
- Hertzog, C., Dixon, R. A., & Hulstsch, D. F. (1990). Chapter four metamemory in adulthood: Differentiating knowledge, belief, and behavior. In *Advances in psychology* (Vol. 71, pp. 161–212). Elsevier.
- Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(1), 22–34.
- Hertzog, C., Dunlosky, J., & Sinclair, S. M. (2010). Episodic feeling-of-knowing resolution derives from the quality of original encoding. *Memory & cognition*, 38(6), 771–784.
- Hertzog, C., Fulton, E. K., Sinclair, S. M., & Dunlosky, J. (2014). Recalled aspects of original encoding strategies influence episodic feelings of knowing. *Memory & cognition*, 42(1), 126–140.
- Hicks, J., Marsh, R., & Ritschel, L. (2002). The role of recollection and partial information in source monitoring. *Journal of experimental psychology. Learning, memory, and cognition*, 28(3), 503–508.
- Higham, P. A. (2007). No special k! a signal detection framework for the strategic regulation of memory accuracy. *Journal of Experimental Psychology: General*, 136(1), 1–22.
- Higham, P. A., Perfect, T. J., & Bruno, D. (2009). Investigating strength and frequency effects in recognition memory using type-2 signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), 57–80.

- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological review*, 95(4), 528–551.
- Hoven, M., Lebreton, M., Engelmann, J. B., Denys, D., Luigjes, J., & van Holst, R. J. (2019). Abnormalities of confidence in psychiatry: an overview and future perspectives. *Translational psychiatry*, 9(1), 1–18.
- Hu, X., Li, T., Zheng, J., Su, N., Liu, Z., & Luo, L. (2015). How much do metamemory beliefs contribute to the font-size effect in judgments of learning? *PloS one*, 10(11), e0142351.
- Isingrini, M., Sacher, M., Perrotin, A., Taconnat, L., Souchay, C., Stoehr, H., & Bouazzaoui, B. (2016). Episodic feeling-of-knowing relies on noncriterial recollection and familiarity: Evidence using an online remember-know procedure. *Consciousness and Cognition*, 41, 31–40.
- Janes, J. L., Rivers, M. L., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both? *Psychonomic bulletin & review*, 25(6), 2356–2364.
- Jang, Y., Wallsten, T. S., & Huber, D. E. (2012). A stochastic detection and retrieval model for the study of metacognition. *Psychological Review*, 119(1), 186–200.
- Jia, X., Li, P., Li, X., Zhang, Y., Cao, W., Cao, L., & Li, W. (2016, January). The Effect of Word Frequency on Judgments of Learning: Contributions of Beliefs and Processing Fluency. *Frontiers in Psychology*, 6.
- Juslin, P. (1993). An explanation of the hard-easy effect in studies of realism of confidence in one's general knowledge. *European Journal of Cognitive Psychology*, 5(1), 55–71.
- Kanai, R., Walsh, V., & Tseng, C.-h. (2010). Subjective discriminability of invisibility: a framework for distinguishing perceptual and attentional failures of awareness. *Consciousness and cognition*, 19(4), 1045–1057.
- Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition*, 28(1), 92–107.
- Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, 32(1), 1–24.
- Kelley, C. M., & Rhodes, M. G. (2002). Making sense and nonsense of experience: Attributions in memory and judgment. In *Psychology of learning and motivation* (Vol. 41, pp. 293–320). Elsevier.



- Kempnich, M., Urquhart, J. A., O'connor, A. R., & Moulin, C. J. (2017). Evidence for the contribution of a threshold retrieval process to semantic memory. *The Quarterly Journal of Experimental Psychology*, *70*(10), 2026–2047.
- Kentridge, R., & Heywood, C. (2000). Metacognition and awareness. *Consciousness and Cognition*, *9*, 308–312.
- Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1322–1337.
- Keren, G. (1991). Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologica*, *77*(3), 217–273.
- Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, *84*(6), 1329–1342.
- Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, *324*(5928), 759–764.
- King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology*, 329–343.
- Klatzky, R. L. (1984). Armchair theorists have more fun. *Behavioral and Brain Sciences*, *7*(2), 244–244.
- Koriat, A. (1993). How do we know that we know? the accessibility model of the feeling of knowing. *Psychological review*, *100*(4), 609–639.
- Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, *124*(3), 311–333.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of experimental psychology: General*, *126*(4), 349.
- Koriat, A. (2007). Metacognition and consciousness. In *The cambridge handbook of consciousness* (pp. 289–325). Cambridge University Press.
- Koriat, A. (2008). Subjective confidence in one's answers: the consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 945–1001.
- Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1133–1145.
- Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (n.d.). Predicting One's Own Forgetting: The

- Role of Experience-Based and Theory-Based Processes. *Journal of Experimental Psychology: General*, 133(4), 643–656. doi: 10.1037/0096-3445.133.4.643
- Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In *Dual-process theories in social psychology* (pp. 483–502). Guilford Press.
- Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Learning, Memory*, 27(1), 34–53.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human learning and memory*, 6(2), 107–118.
- Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of memory and Language*, 52(4), 478–492.
- Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science*, 18(1), 64–71.
- Kruschke, J. (2014). *Doing bayesian data analysis: A tutorial with r, jags, and stan*. Academic Press.
- Kruskal, W. H., & Goodman, L. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, 49(268), 732–764.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82(13).
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4.
- Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences*, 103(49), 18763–18768.
- Lee, A. L., Ruby, E., Giles, N., & Lau, H. (2018). Cross-domain association in metacognitive efficiency depends on first-order task types. *Frontiers in psychology*, 9, 2464.
- Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational behavior and human performance*, 20(2), 159–183.
- Logan, J. M., Castel, A. D., Haber, S., & Viehman, E. J. (2012). Metacognition and the spacing effect: the role of repetition, feedback, and instruction on judgments of learning for massed and spaced rehearsal. *Metacognition and Learning*, 7(3), 175–195.
- Luna, K., Martín-Luengo, B., & Albuquerque, P. B. (2018, July). Do delayed judgements

- of learning reduce metamemory illusions? A meta-analysis. *Quarterly Journal of Experimental Psychology*, 71(7), 1626–1636.
- Luna, K., Nogueira, M., & Albuquerque, P. B. (2019). Words in larger font are perceived as more important: explaining the belief that font size affects memory. *Memory*, 27(4), 555–560.
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.
- Mamassian, P. (2016). Visual Confidence. *Annual Review of Vision Science*, 2(1), 459–481. doi: 10.1146/annurev-vision-111815-114630
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological review*, 87(3), 252–271.
- Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, 21(1), 422–430. doi: 10.1016/j.concog.2011.09.021
- Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of type 1 and type 2 data: meta-d, response-specific meta-d, and the unequal variance sdt model. In *The cognitive neuroscience of metacognition* (pp. 25–66). Springer.
- Maniscalco, B., & Lau, H. (2015). Manipulation of working memory contents selectively impairs metacognitive sensitivity in a concurrent visual discrimination task. *Neuroscience of Consciousness*, 2015(1). doi: 10.1093/nc/niv002
- Maniscalco, B., Peters, M. A., & Lau, H. (2016). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Attention, Perception, & Psychophysics*, 78(3), 923–937.
- Masson, M. E., & Rotello, C. M. (2009). Sources of bias in the goodman–kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2), 509.
- Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (jols): An analytic or nonanalytic basis for jols? *Memory & Cognition*, 29(2), 222–233.
- Mazancieux, A., Fleming, S., Souchay, C., & Moulin, C. (2018). Retrospective confidence judgments across tasks: domain-general processes underlying metacognitive accuracy.
- Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of*

- Experimental Psychology: Learning, Memory, and Cognition*, 21(5), 1263–1274.
- Mazzoni, G., Scoboria, A., & Harvey, L. (2010). Nonbelieved memories. *Psychological Science*, 21(9), 1334–1340.
- McCabe, D. P., & Soderstrom, N. C. (2011). Recollection-based prospective metamemory judgments are more accurate than those based on confidence: Judgments of remembering and knowing (JORKS). *Journal of Experimental Psychology: General*, 140(4), 605–621. doi: 10.1037/a0024014
- McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013). Anatomical Coupling between Distinct Metacognitive Systems for Memory and Visual Perception. *Journal of Neuroscience*, 33(5), 1897–1906. doi: 10.1523/JNEUROSCI.1890-12.2013
- Metcalfe, J., Schwartz, B., & Joaquim, S. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(4), 851–861.
- Metcalfe, J., & Schwartz, B. L. (2016). The ghost in the machine: Self-reflective consciousness and the neuroscience of metacognition. In *Oxford handbook of metamemory* (pp. 407–424). Oxford University Press.
- Metcalfe, J., & Son, L. K. (2012). Anoetic, noetic, and auto-noetic metacognition. In *The foundations of metacognition* (pp. 289–301). Oxford University Press.
- Meuwese, J. D., van Loon, A. M., Lamme, V. A., & Fahrenfort, J. J. (2014). The subjective experience of object recognition: comparing metacognition for object detection and object categorization. *Attention, Perception, & Psychophysics*, 76(4), 1057–1068.
- Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as bayesian probability: From neural origins to behavior. *Neuron*, 88(1), 78–92.
- Miller, G. (2000). How to keep our metatheories adaptive: Beyond cosmides, tooby, and lakatos. *Psychological Inquiry*, 11(1), 42–46.
- Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, 145(2), 200–219.
- Mograb, D. C., Brown, R. G., Salas, C., & Morris, R. G. (2012). Emotional reactivity and awareness of task performance in alzheimer's disease. *Neuropsychologia*, 50(8), 2075–2084.
- Mograb, D. C., & Morris, R. G. (2013). Implicit awareness in anosognosia: Clinical observations, experimental evidence, and theoretical implications. *Cognitive*

- Neuroscience*, 4(3-4), 181–197. doi: 10.1080/17588928.2013.833899
- Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex. *The Journal of Neuroscience*, 38(14), 3534–3546. doi: 10.1523/JNEUROSCI.2360-17.2018
- Moritz, S., Woodward, T., Jelinek, L., & Klinge, R. (2008). Memory and metamemory in schizophrenia: a liberal acceptance account of psychosis. *Psychological medicine*, 38(6), 825–832.
- Moro, V., Pernigo, S., Zapparoli, P., Cordioli, Z., & Aglioti, S. M. (2011). Phenomenology and neural correlates of implicit and emergent motor awareness in patients with anosognosia for hemiplegia. *Behavioural Brain Research*, 225(1), 259–269.
- Morris, R. G., & Mograbi, D. C. (2013). Anosognosia, autobiographical memory and self knowledge in Alzheimer's disease. *Cortex*, 49(6), 1553–1565. doi: 10.1016/j.cortex.2012.09.006
- Morson, S. M., Moulin, C. J., & Souchay, C. (2015). Selective deficits in episodic feeling of knowing in ageing: A novel use of the general knowledge task. *Acta Psychologica*, 157, 85–92. doi: 10.1016/j.actpsy.2015.02.014
- Moulin, Perfect, T., & Jones, R. (2000a). The effects of repetition on allocation of study time and judgements of learning in alzheimer's disease. *Neuropsychologia*, 38(6), 748–756.
- Moulin, Perfect, T. J., & Jones, R. W. (2000b). Global predictions of memory in alzheimer's disease: Evidence for preserved metamemory monitoring. *Aging, Neuropsychology, and Cognition*, 7(4), 230–244.
- Moulin, & Souchay, C. (2014). Epistemic feelings and memory. In *The sage handbook of applied memory* (pp. 520–539). SAGE Publications Ltd.
- Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people's metamemory judgments: What are the contributions of processing fluency and beliefs about memory? *The Quarterly Journal of Experimental Psychology*, 69(4), 781–799.
- Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, 70, 1–12.
- Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic bulletin & review*, 20(2), 378–384.

- Murdock, B. B. (1983). A distributed memory model for serial-order information. *Psychological Review*, *90*(4), 316–338.
- Narens, L., Jameson, K. A., & Lee, V. (1994). Subthreshold priming and memory monitoring. In *Metacognition: Knowing about knowing* (pp. 71–92). MIT Press Cambridge, MA.
- Navajas, J., Bahrami, B., & Latham, P. E. (2016). Post-decisional accounts of biases in confidence. *Current Opinion in Behavioral Sciences*, *11*, 55–60. doi: 10.1016/j.cobeha.2016.05.005
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological bulletin*, *95*(1), 109.
- Nelson, T. O. (1988). Predictive accuracy of the feeling of knowing across different criterion tasks and across different subject populations and individuals. In *Practical aspects of memory: Current research and issues, vol. 1. memory in everyday life* (pp. 190–196). John Wiley & Sons.
- Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. In *Psychology of learning and motivation* (Vol. 26, pp. 125–173). Elsevier.
- Nelson, T. O. (1996). Consciousness and metacognition. *American psychologist*, *51*(2), 102.
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (jols) are extremely accurate at predicting subsequent recall: The "delayed-jol effect". *Psychological Science*, *2*(4), 267–271.
- Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect.". *Journal of experimental psychology: Learning, Memory, and Cognition*, *14*(4), 676–686.
- Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. In *Metacognition: Knowing about knowing* (pp. 1–25). MIT Press.
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new french lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516–524.
- Norman, E., & Price, M. C. (2015). Measuring consciousness with confidence ratings. *Behavioural methods in consciousness research*, 159–180.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606.
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in cognitive sciences*, *12*(6), 237–241.
- Overgaard, M., & Sandberg, K. (2012). Kinds of access: different methods for report

- reveal different kinds of metacognitive access. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1287–1296.
- O'Connor, A. R., & Moulin, C. J. (2010). Recognition without identification, erroneous familiarity, and déjà vu. *Current Psychiatry Reports*, 12(3), 165–173.
- Palmer, E. C., David, A. S., & Fleming, S. M. (2014). Effects of age on metacognitive efficiency. *Consciousness and Cognition*, 28, 151–160.
- Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory Experiments in Neurological Populations: A Review. *Neuropsychology Review*, 15(3), 105–130.
- Pappas, B. A., Sunderland, T., Weingartner, H. M., Vitiello, B., Martinson, H., & Putnam, K. (1992). Alzheimer's disease and feeling-of-knowing for knowledge and episodic memory. *Journal of Gerontology*, 47(3), 159–164.
- Pereira, M., Faivre, N., Iturrate, I., Wirthlin, M., Serafini, L., Martin, S., ... Millan, J. d. R. (2018). Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *BioRxiv*, 496877.
- Perfect, T., & Hollins, T. (1996). Predictive feeling of knowing judgements and postdictive confidence judgements in eyewitness memory and general knowledge. *Applied Cognitive Psychology*, 10(5), 371–382.
- Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science*, 9(3), 319–332.
- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological review*, 117(3), 864–901.
- Plummer, M. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, pp. 10–18).
- Poppel, E., Held, R., & Frost, D. (1973). Residual function after brain wounds involving the central visual pathways in man. In *Brain and behaviour: Critical concepts in psychology* (pp. 322–327). Taylor & Francis.
- Prigatano, G. P. (1999). *Principles of neuropsychological rehabilitation*. Oxford University Press.
- Proust, J. (2003). Does metacognition necessarily involve metarepresentation? *Behavioral and brain sciences*, 26(3), 352–352.
- Proust, J. (2007). Metacognition and metarepresentation: is a self-directed theory of mind a precondition for metacognition? *Synthese*, 159(2), 271–295.
- Rabbitt, P. (1966). Errors and error correction in choice-response tasks. *Journal of*

- experimental psychology*, 71(2), 264.
- Rabbitt, P., & Vyas, S. (1981). Processing a display even after you make a response to it. how perceptual errors can be corrected. *The Quarterly Journal of Experimental Psychology*, 33(3), 223–239.
- Rahnev, D. A., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S. H. (2011). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of neurophysiology*, 107(6), 1556–1563.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, 20(4), 873–922.
- Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological review*, 116(1), 59–83.
- Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological review*, 120(3), 697–719.
- Rausch, M., & Zehetleitner, M. (2017). Should metacognition be measured by logistic regression? *Consciousness and cognition*, 49, 291–312.
- Reder, L. M. (1987). Strategy selection in question answering. *Cognitive psychology*, 19(1), 90–138.
- Reder, L. M. (1988). Strategic control of retrieval strategies. In *Psychology of learning and motivation* (Vol. 22, pp. 227–259). Elsevier.
- Reggev, N., Zuckerman, M., & Maril, A. (2011). Are all judgments created equal? *Neuropsychologia*, 49(5), 1332–1342.
- Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. (2019). From knowing to remembering: The semantic–episodic distinction. *Trends in cognitive sciences*, 23(12), 1041–1057.
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: evidence for metacognitive illusions. *Journal of experimental psychology: General*, 137(4), 615–625.
- Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, 16(3), 550–554.
- Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual review of psychology*, 39(1), 475–543.
- Roberts, J., Clare, L., & Woods, R. (2009). Subjective memory complaints and awareness of memory functioning in mild cognitive impairment: a systematic



- review. *Dementia and geriatric cognitive disorders*, 28(2), 95–109.
- Rosenthal, D. (2000). Consciousness and metacognition. In *Metarepresentation: Proceedings of the tenth vancouver cognitive science conference* (pp. 265–295).
- Rosenthal, D. (2019). Consciousness and confidence. *Neuropsychologia*, 128, 255–265.
- Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature communications*, 10(1), 1141.
- Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human metacognition across domains: Insights from individual differences and neuroimaging. *Personality neuroscience*, 1.
- Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance. *Biological Psychiatry*, 84(6), 443–451. doi: 10.1016/j.biopsych.2017.12.017
- Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive neuroscience*, 1(3), 165–175.
- Sadeghi, S., Ekhtiari, H., Bahrami, B., & Ahmadabadi, M. N. (2017). Metacognitive Deficiency in a Perceptual but Not a Memory Task in Methadone Maintenance Patients. *Scientific Reports*, 7(1). doi: 10.1038/s41598-017-06707-w
- Samaha, J., & Postle, B. R. (2017). Correlated individual differences suggest a common mechanism underlying metacognition in visual perception and visual short-term memory. *Proceedings of the Royal Society B: Biological Sciences*, 284(1867).
- Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: is one measure better than the other? *Consciousness and cognition*, 19(4), 1069–1078.
- Schacter, D. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(1), 39–54.
- Schnyer, D. M., Verfaellie, M., Alexander, M. P., LaFleche, G., Nicholls, L., & Kaszniak, A. W. (2004). A role for right medial prefrontal cortex in accurate feeling-of-knowing judgments: Evidence from patients with lesions to frontal cortex. *Neuropsychologia*, 42(7), 957–966.
- Schooler, J. W. (2002). Re-representing consciousness: Dissociations between experience and meta-consciousness. *Trends in cognitive sciences*, 6(8), 339–344.
- Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review*, 1(3), 357–375.

- Schwartz, B. L., Boduroglu, A., & Tekcan, A. (2016, December). Methodological concerns: the feeling-of-knowing task affects resolution. *Metacognition and Learning, 11*(3), 305–316.
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14*(4), 332–348.
- Schönbrodt, F. D., & Perugini, M. (2013, October). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612.
- Scott, R. B., Dienes, Z., Barrett, A. B., Bor, D., & Seth, A. K. (2014). Blind insight: metacognitive discrimination despite chance task performance. *Psychological science, 25*(12), 2199–2208.
- Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends in cognitive sciences, 12*(8), 314–321.
- Sharp, G. L., Cutler, B. L., & Penrod, S. D. (1988). Performance feedback improves the resolution of confidence judgments. *Organizational Behavior and Human Decision Processes, 42*(3), 271–283.
- Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in cognitive sciences, 18*(4), 186–193.
- Shea, N., & Frith, C. D. (2019). The global workspace needs metacognition. *Trends in cognitive sciences, 23*(7), 560–571.
- Sherman, Barrett, A. B., & Kanai, R. (2015). Inferences about consciousness using subjective reports of confidence. *Behavioral methods in consciousness research, 87*–106.
- Sherman, Seth, A., Barrett, A., & Kanai, R. (2015, September). Prior expectations facilitate metacognition for perceptual decision. *Consciousness and Cognition, 35*, 53–65.
- Sherman, Seth, A. K., & Barrett, A. B. (2018). Quantifying metacognitive thresholds using signal-detection theory. *bioRxiv, 361543*.
- Shimamura, A. P., & Squire, L. R. (1986). Memory and metamemory: a study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(3), 452–460.
- Siedlecka, M., Hobot, J., Skóra, Z., Paulewicz, B., Timmermans, B., & Wierzchoń, M. (2019). Motor response influences perceptual awareness judgements. *Consciousness and cognition, 75*, 102804.
- Siedlecka, M., Paulewicz, B., & Wierzchoń, M. (2016). But I Was So Sure! Metacognitive

- Judgments Are Less Accurate Given Prospectively than Retrospectively. *Frontiers in Psychology*, 7. doi: 10.3389/fpsyg.2016.00218
- Siedlecka, M., Skóra, Z., Paulewicz, B., Fijałkowska, S., Timmermans, B., & Wierchoń, M. (2019). Responses improve the accuracy of confidence judgements in memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(4), 712–723.
- Sikström, S., & Jönsson, F. (2005). A model for stochastic drift in memory strength to account for judgments of learning. *Psychological Review*, 112(4), 932–950.
- Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(4), 907–912. doi: 10.1037//0278-7393.27.4.907
- Smith, J. D., Beran, M. J., Couchman, J. J., & Coutinho, M. V. (2008). The comparative study of metacognition: Sharper paradigms, safer inferences. *Psychonomic bulletin & review*, 15(4), 679–691.
- Smith, J. D., Schull, J., Strote, J., McGee, K., Egnor, R., & Erb, L. (1995). The uncertain response in the bottlenosed dolphin (*tursiops truncatus*). *Journal of Experimental Psychology: General*, 124(4), 391–408.
- Snizek, J. A. (1992). Groups under uncertainty: An examination of confidence in group decision making. *Organizational Behavior and Human Decision Processes*, 52(1), 124–155.
- Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2), 553–558.
- Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(1), 204–221.
- Song, C., Kanai, R., Fleming, S. M., Weil, R. S., Schwarzkopf, D. S., & Rees, G. (2011). Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Consciousness and Cognition*, 20(4), 1787–1792. doi: 10.1016/j.concog.2010.12.011
- Souchay, C. (2007). Metamemory in alzheimer's disease. *Cortex*, 43(7), 987–1003.
- Souchay, C., Bacon, E., & Danion, J.-M. (2006). Metamemory in schizophrenia: An exploration of the feeling-of-knowing state. *Journal of Clinical and Experimental Neuropsychology*, 28(5), 828–840.
- Souchay, C., & Isingrini, M. (2012). Are feeling-of-knowing and judgment-of-learning

- different? evidence from older adults. *Acta Psychologica*, 139(3), 458–464.
- Souchay, C., Isingrini, M., & Espagnet, L. (2000). Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology*, 14(2), 299–309.
- Souchay, C., Isingrini, M., & Gil, R. (2002). Alzheimer's disease and feeling-of-knowing in episodic memory. *Neuropsychologia*, 40(13), 2386–2396.
- Souchay, C., Moulin, C. J., Clarys, D., Taconnat, L., & Isingrini, M. (2007). Diminished episodic memory awareness in older adults: Evidence from feeling-of-knowing and recollection. *Consciousness and Cognition*, 16(4), 769–784. doi: 10.1016/j.concog.2006.11.002
- Souchay, C., & Moulin, C. J. A. (2013). Eliciting the implicit: Metacognition in Alzheimer's disease. *Cognitive Neuroscience*, 4(3-4), 203–204. doi: 10.1080/17588928.2013.853657
- Spehn, M. K., & Reder, L. M. (2000). The unconscious feeling of knowing: A commentary on koriat's paper. *Consciousness and Cognition*, 9, 187–192.
- Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Psychology Press.
- Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A. E., Paliwal, S., Gard, T., ... Petzschner, F. H. (2016, November). Allostatic Self-efficacy: A Metacognitive Theory of Dyshomeostasis-Induced Fatigue and Depression. *Frontiers in Human Neuroscience*, 10. doi: 10.3389/fnhum.2016.00550
- Su, N., Li, T., Zheng, J., Hu, X., Fan, T., & Luo, L. (2018). How font size affects judgments of learning: Simultaneous mediating effect of item-specific beliefs about fluency and moderating effect of beliefs about font size and memory. *PloS one*, 13(7), e0200888.
- Sunderland, A., Harris, J. E., & Baddeley, A. D. (1983). Do laboratory tests predict everyday memory? a neuropsychological study. *Journal of verbal learning and verbal behavior*, 22(3), 341–357.
- Susser, J. A., & Mulligan, N. W. (2015, August). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review*, 22(4), 1014–1019. doi: 10.3758/s13423-014-0768-1
- Susser, J. A., Panitz, J., Buchin, Z., & Mulligan, N. W. (2017, August). The motoric fluency effect on metamemory. *Journal of Memory and Language*, 95, 116–123. doi: 10.1016/j.jml.2017.03.002
- Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: a meta-analysis. *Psychological bulletin*, 121(3), 371–394.

- Thompson, & Morsanyi, K. (2012). Analytic thinking: do you feel like it? *Mind & Society*, 11(1), 93–105.
- Thompson, Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013, August). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, 128(2), 237–251.
- Tulving, E. (1985). How many memory systems are there? *American psychologist*, 40(4), 385–398.
- Tunney, R. J., & Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Memory & cognition*, 31(7), 1060–1071.
- Underwood, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of experimental psychology*, 71(5), 673–679.
- Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & cognition*, 43(4), 647–658.
- Unkelbach, C., & Greifeneder, R. (2013). A general model of fluency effects in judgment and decision making. In *The experience of thinking* (pp. 21–42). Psychology Press.
- Vaccaro, A. G., & Fleming, S. M. (2018). Thinking about thinking: a coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain and neuroscience advances*, 2, 1–14.
- Van den Berg, R., Yoo, A. H., & Ma, W. J. (2017). Fechner's law in metacognition: A quantitative model of visual working memory confidence. *Psychological review*, 124(2), 197–214.
- Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006, April). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning*, 1(1), 3–14.
- Vickers, D. (1979). Uncertainty, choice, and the marginal efficiencies. *Journal of Post Keynesian Economics*, 2(2), 240–254.
- Vickers, D., & Packer, J. (1982). Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta psychologica*, 50(2), 179–197.
- Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... others (2018). Bayesian inference for psychology. part i: Theoretical advantages and practical ramifications. *Psychonomic bulletin & review*, 25(1), 35–57.
- Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., ... Blakemore, S.-J. (2013). The development of metacognitive ability in adolescence.

- Consciousness and cognition*, 22(1), 264–271.
- Weiskrantz, L., Warrington, E. K., Sanders, M., & Marshall, J. (1974). Visual capacity in the hemianopic field following a restricted occipital ablation. *Brain*, 97(4), 709–728.
- Whitmarsh, S., Oostenveld, R., Almeida, R., & Lundqvist, D. (2017, February). Metacognition of attention during tactile discrimination. *NeuroImage*, 147, 121–129.
- Whittlesea, W. L. D., B. W. (2001). The discrepancy-attribution hypothesis: I. the heuristic basis of feelings and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1), 3–13.
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological review*, 114(1), 152–176.
- Wojcik, D. Z., Moulin, C. J. A., & Souchay, C. (2013). Metamemory in children with autism: Exploring “feeling-of-knowing” in episodic and semantic memory. *Neuropsychology*, 27(1), 19–27.
- Wokke, M. E., Achoui, D., & Cleeremans, A. (2019). Action information contributes to metacognitive decision-making. *bioRxiv*, 657957.
- Yang, C., Huang, T. S.-T., & Shanks, D. R. (2018, April). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language*, 99, 99–110.
- Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1310–1321.
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1341–1354.
- 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.
- Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & cognition*, 41(2), 229–241.
- Zajonc, R. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9(2), 1–27.
- Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, 15(1),

41-44.